Real-time 3D Eyelids Tracking From Semantic Edges

Quan Wen, Feng Xu, Ming Lu, Jun-Hai Yong
Tsinghua University

Background

Facial capture and animation is crucial in many applications

Face capture in computer games

Face animation in movies

Background

Facial tracking focus less on the eyes

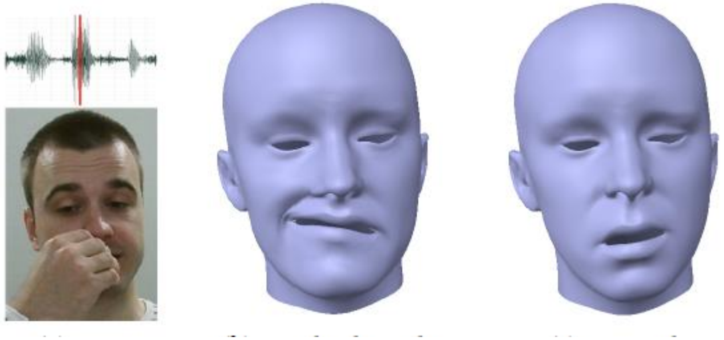
[Bouaziz et al. 2013]

[Cao et al. 2014]

[Li et al. 2013]

[Cao et al. 2015]

[Hsieh et al. 2015]

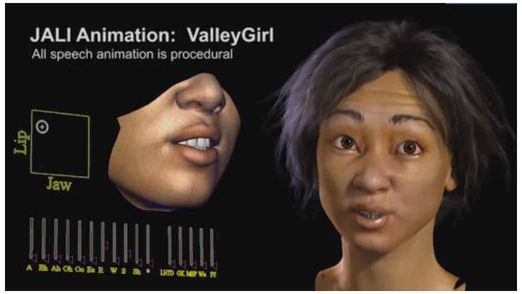


[Liu et al. 2015]

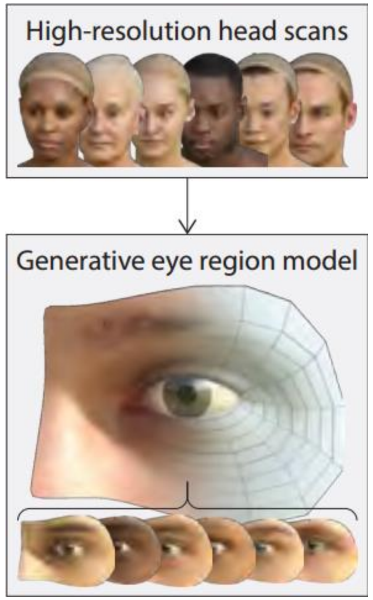
Background

Facial organs tracking

[Bérard et al. 2016]

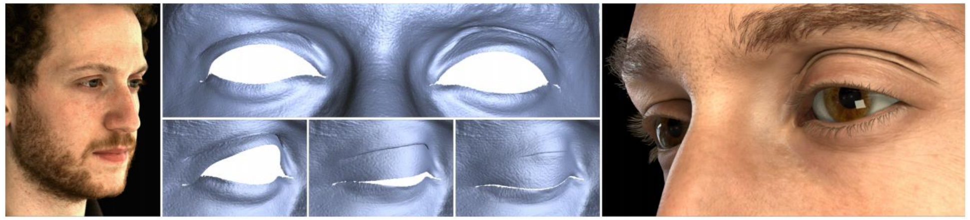


[Edwards et al. 2016]



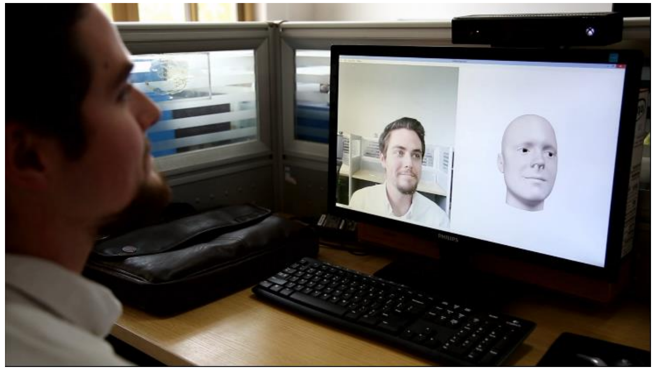
[Wood et al. 2016]

[Wang et al. 2016]



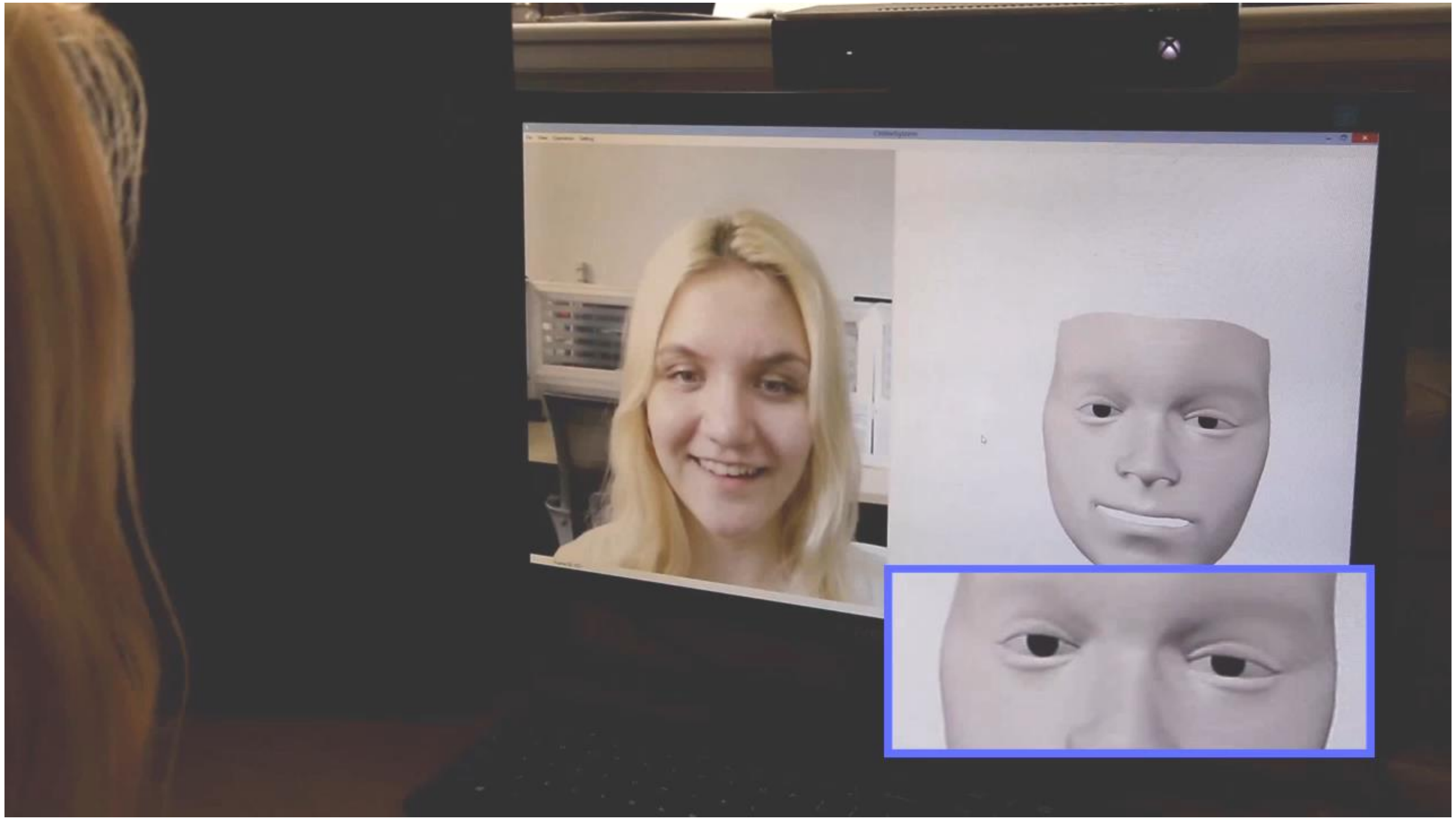
[Bermano et al. 2015]

[Wu et al. 2016]



[Wen et al. 2016]

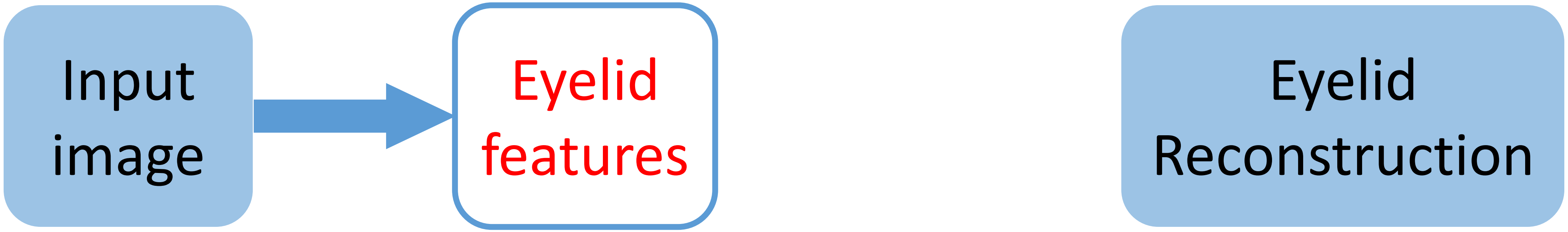
Our Work



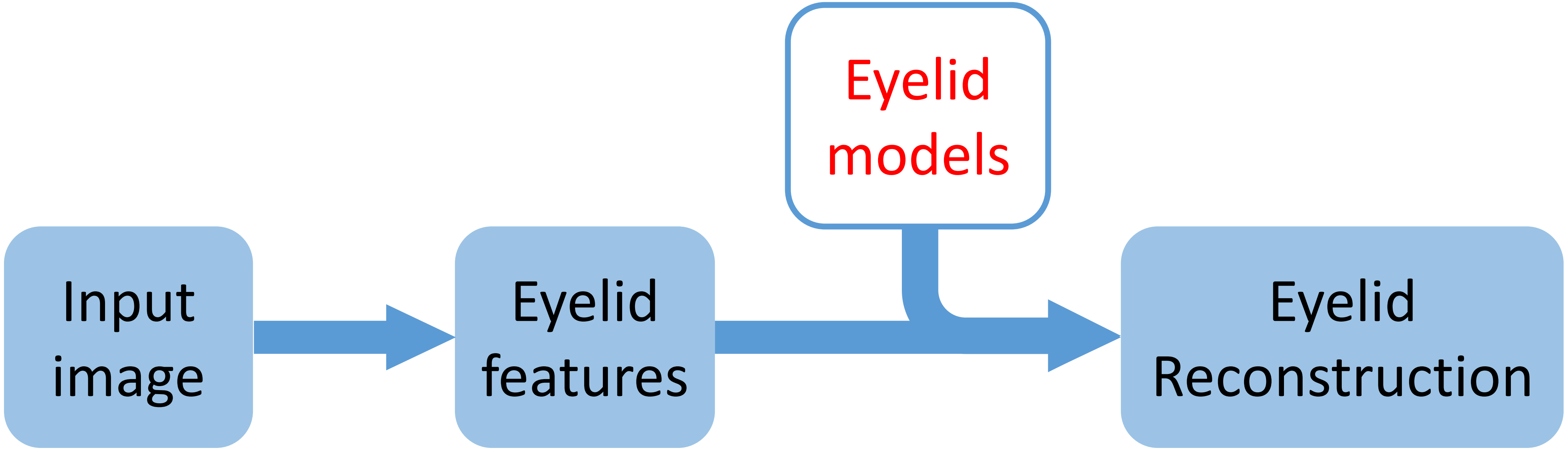
A real-time 3D eyelids tracking system

Overview

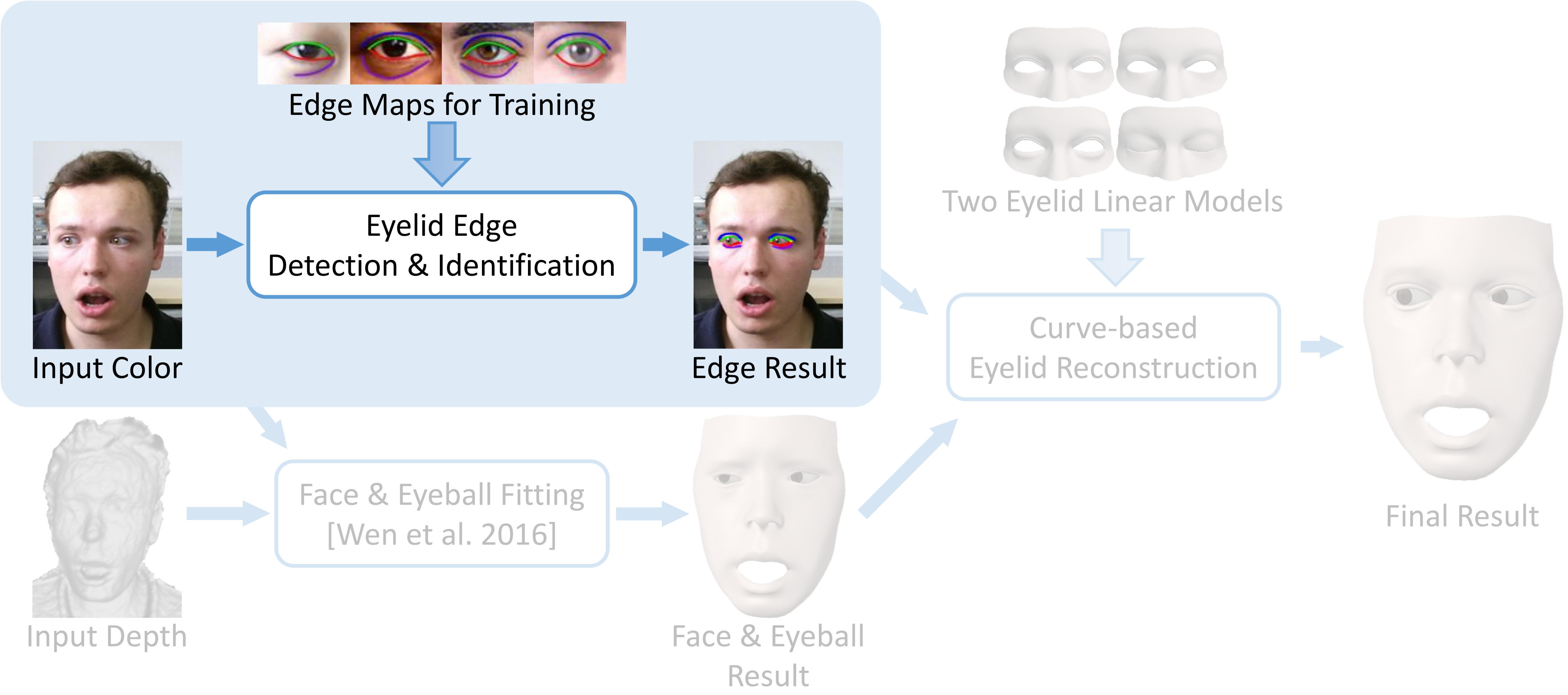
Overview



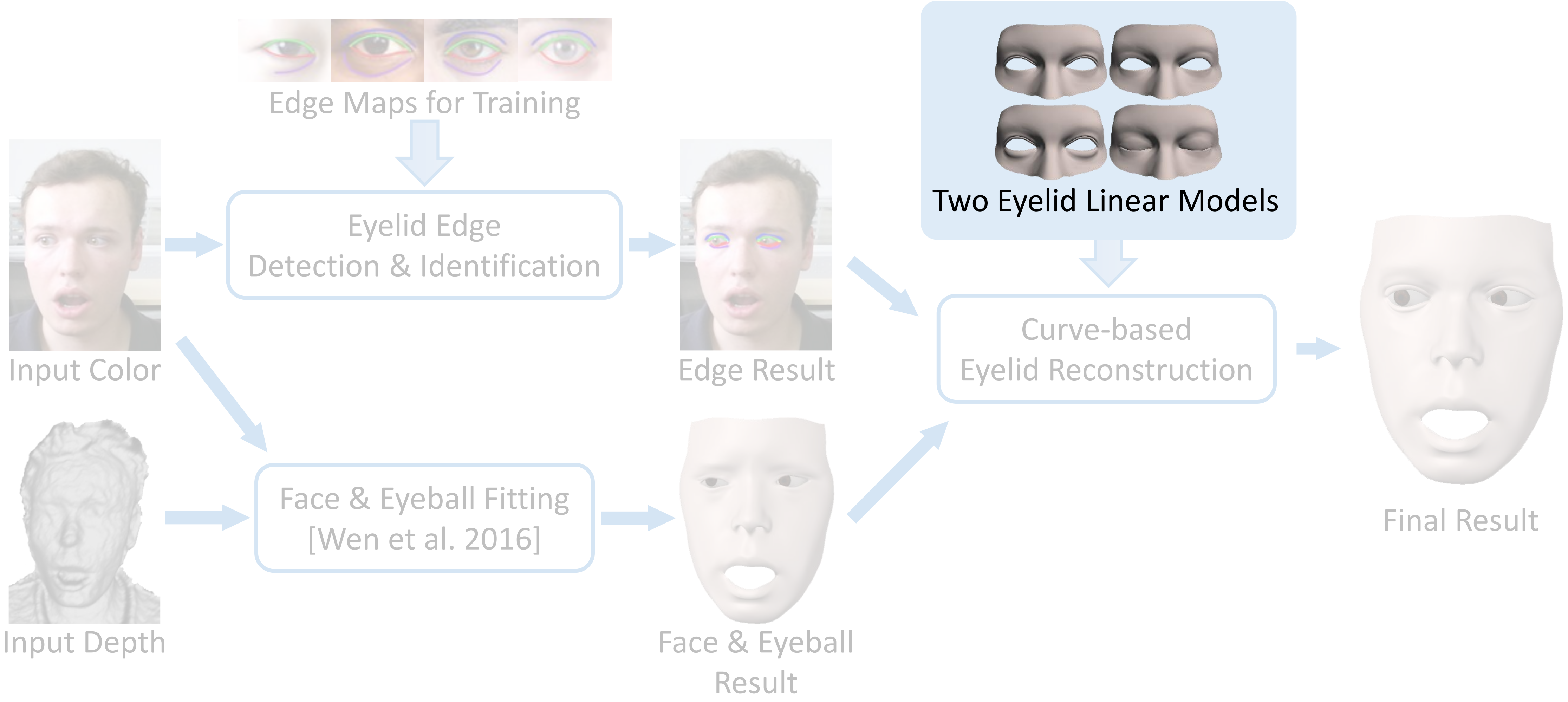
Overview


Eyelid
models
Input
image
Eyelid
features
Eyelid
Reconstruction

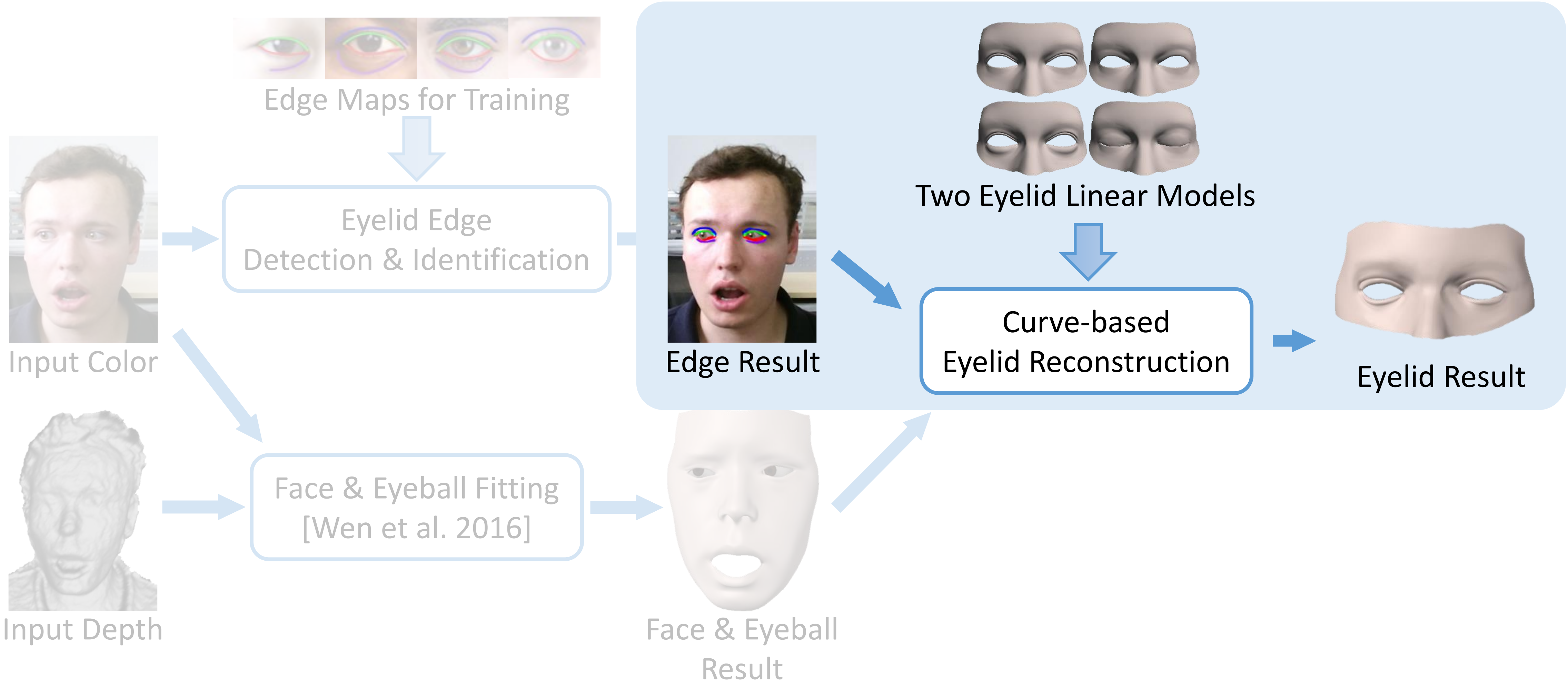
Overview


Edge Maps for Training
Input Color
Eyelid Edge
Detection & Identification
Edge Result
Two Eyelid Linear Models
Curve-based
Eyelid Reconstruction
Final Result
Input Depth
Face & Eyeball Fitting
[Wen et al. 2016]
Face & Eyeball
Result

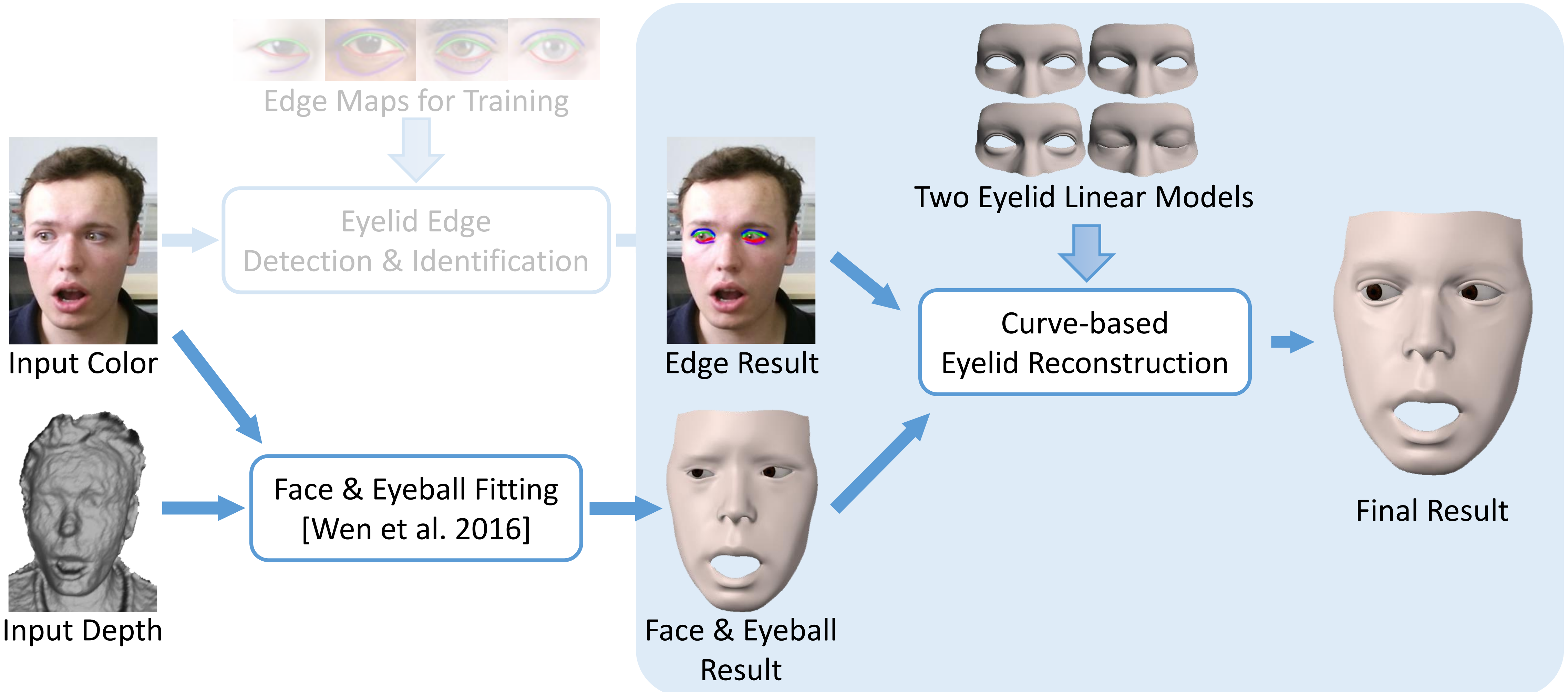
Overview


Edge Maps for Training
Input Color
Eyelid Edge
Detection & Identification
Edge Result
Two Eyelid Linear Models
Curve-based
Eyelid Reconstruction
Final Result
Input Depth
Face & Eyeball Fitting
[Wen et al. 2016]
Face & Eyeball
Result

Overview

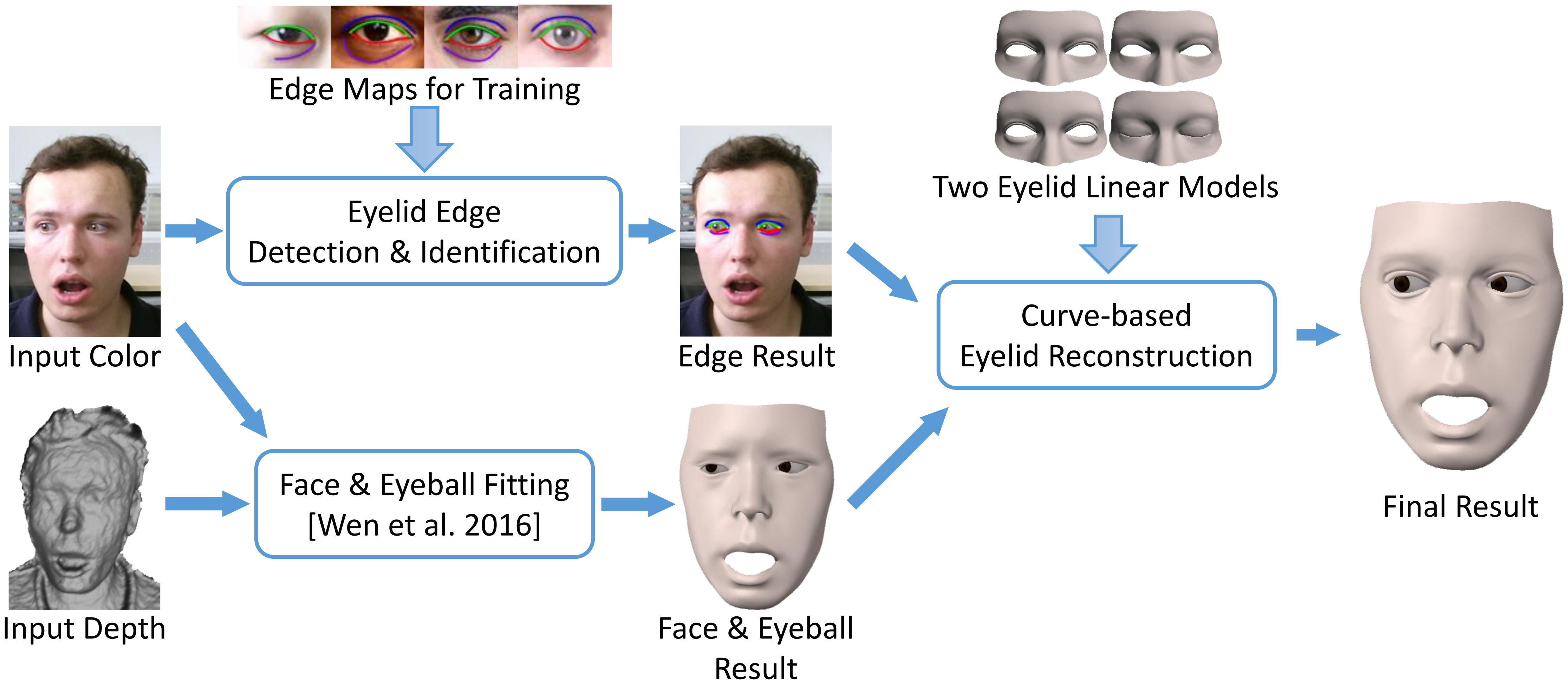

Edge Maps for Training
Input Color
Eyelid Edge
Detection & Identification
Input Depth
Face & Eyeball Fitting
[Wen et al. 2016]
Face & Eyeball
Result
Edge Result
Two Eyelid Linear Models
Curve-based
Eyelid Reconstruction
Eyelid Result

Overview



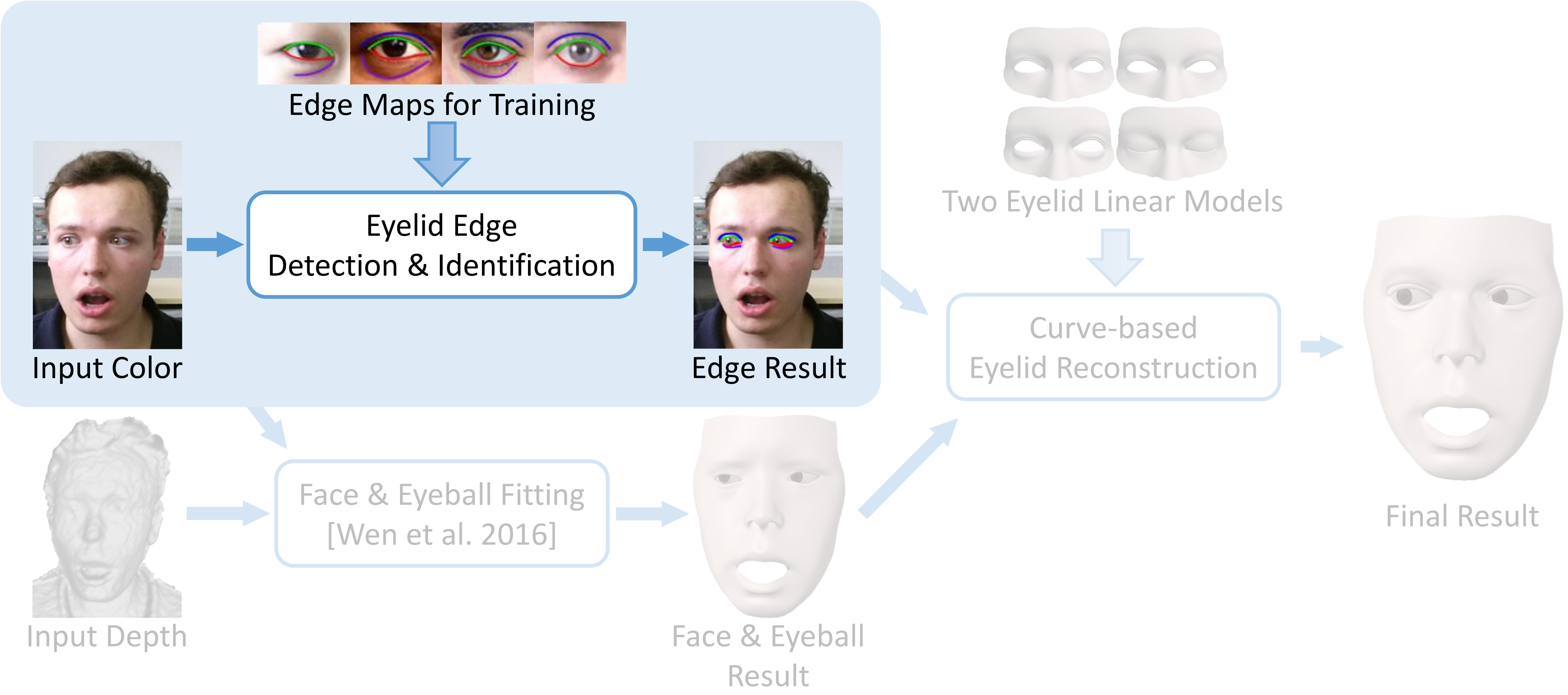
Edge Maps for Training
Input Color
Eyelid Edge
Detection & Identification
Edge Result
Two Eyelid Linear Models
Curve-based
Eyelid Reconstruction
Final Result
Input Depth
Face & Eyeball Fitting
[Wen et al. 2016]
Face & Eyeball
Result

Overview

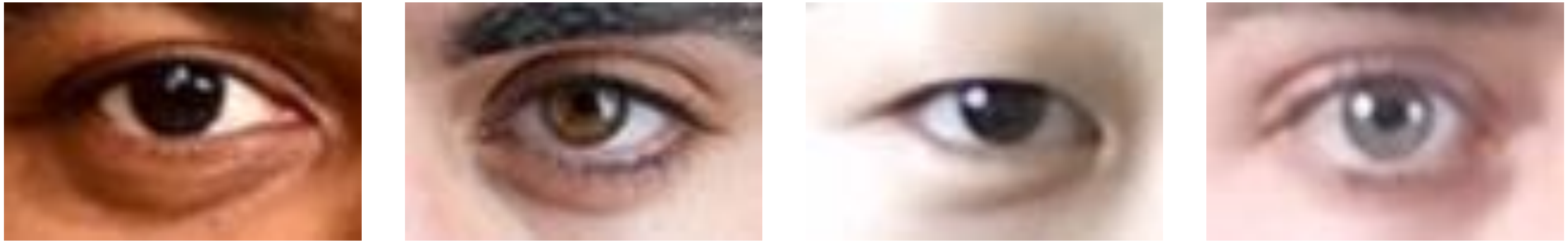


Edge Maps for Training
Input Color
Eyelid Edge
Detection & Identification
Edge Result
Two Eyelid Linear Models
Curve-based
Eyelid Reconstruction
Final Result
Input Depth
Face & Eyeball Fitting
[Wen et al. 2016]
Face & Eyeball
Result

Eyelid Edge Detection and Identification

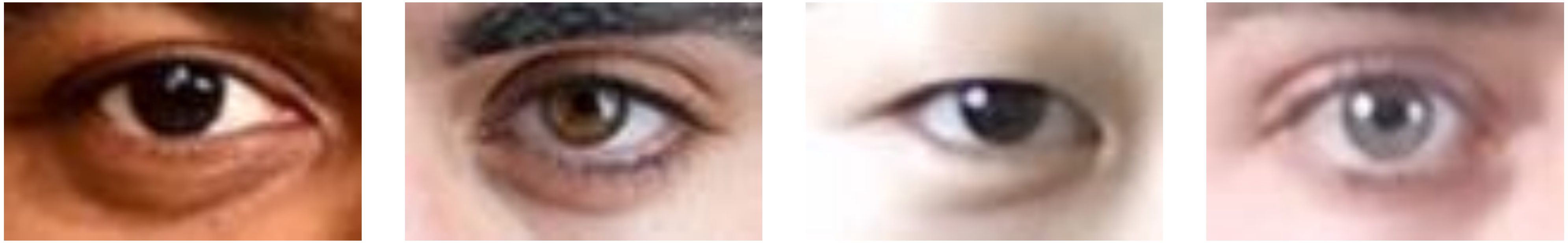


Semantic Eyelid Edges



Main features of the eyes:
double-fold, top eyelid, bottom eyelid, bulge

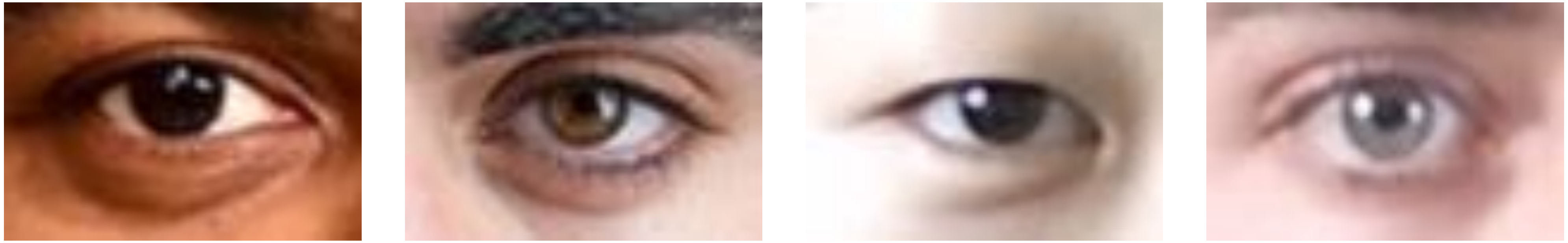
Semantic Eyelid Edges



Main features of the eyes:

double-fold, top eyelid, bottom eyelid, bulge

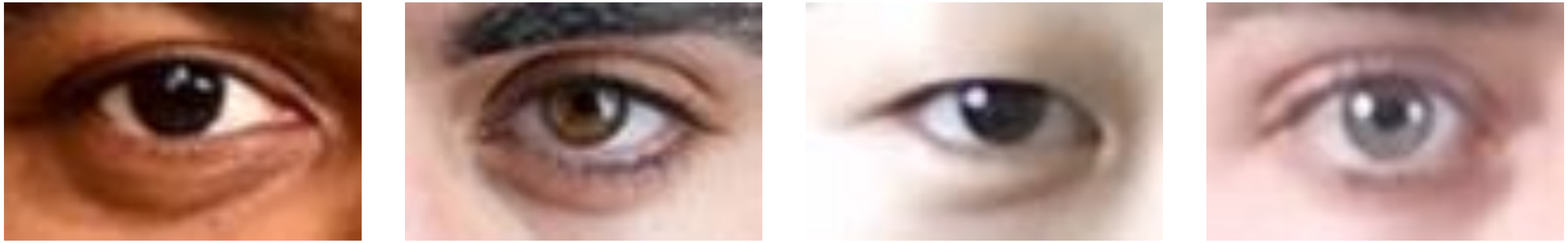
Semantic Eyelid Edges



Main features of the eyes:

double-fold, top eyelid, bottom eyelid, bulge

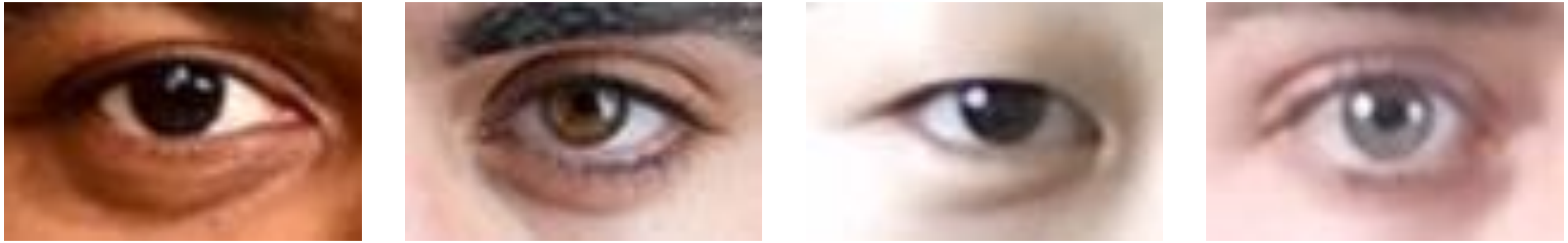
Semantic Eyelid Edges



Main features of the eyes:

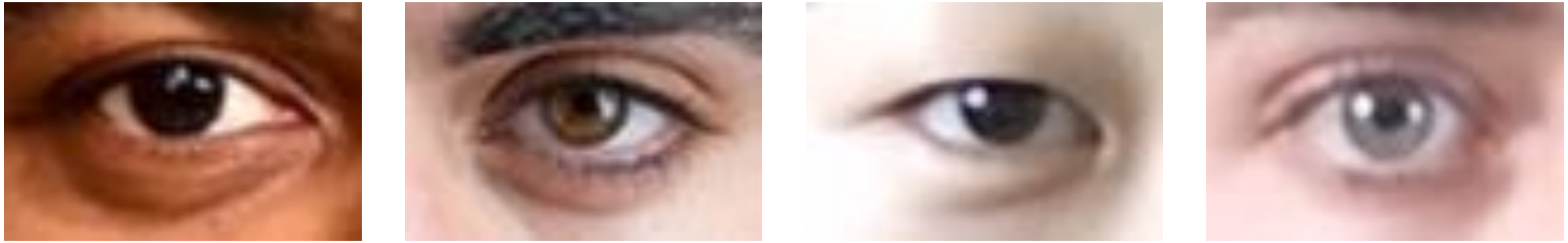
double-fold, top eyelid, bottom eyelid, bulge

Semantic Eyelid Edges



Main features of the eyes:
double-fold, top eyelid, bottom eyelid, bulge

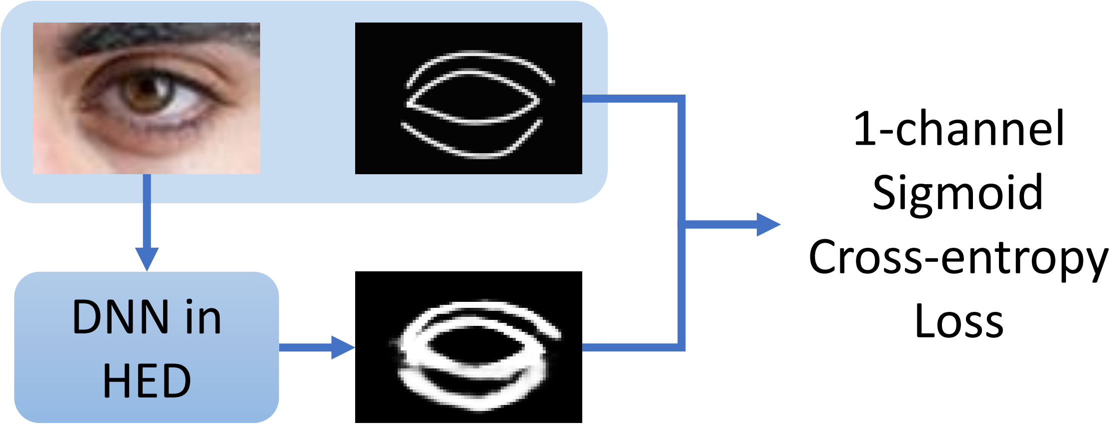
Semantic Eyelid Edges



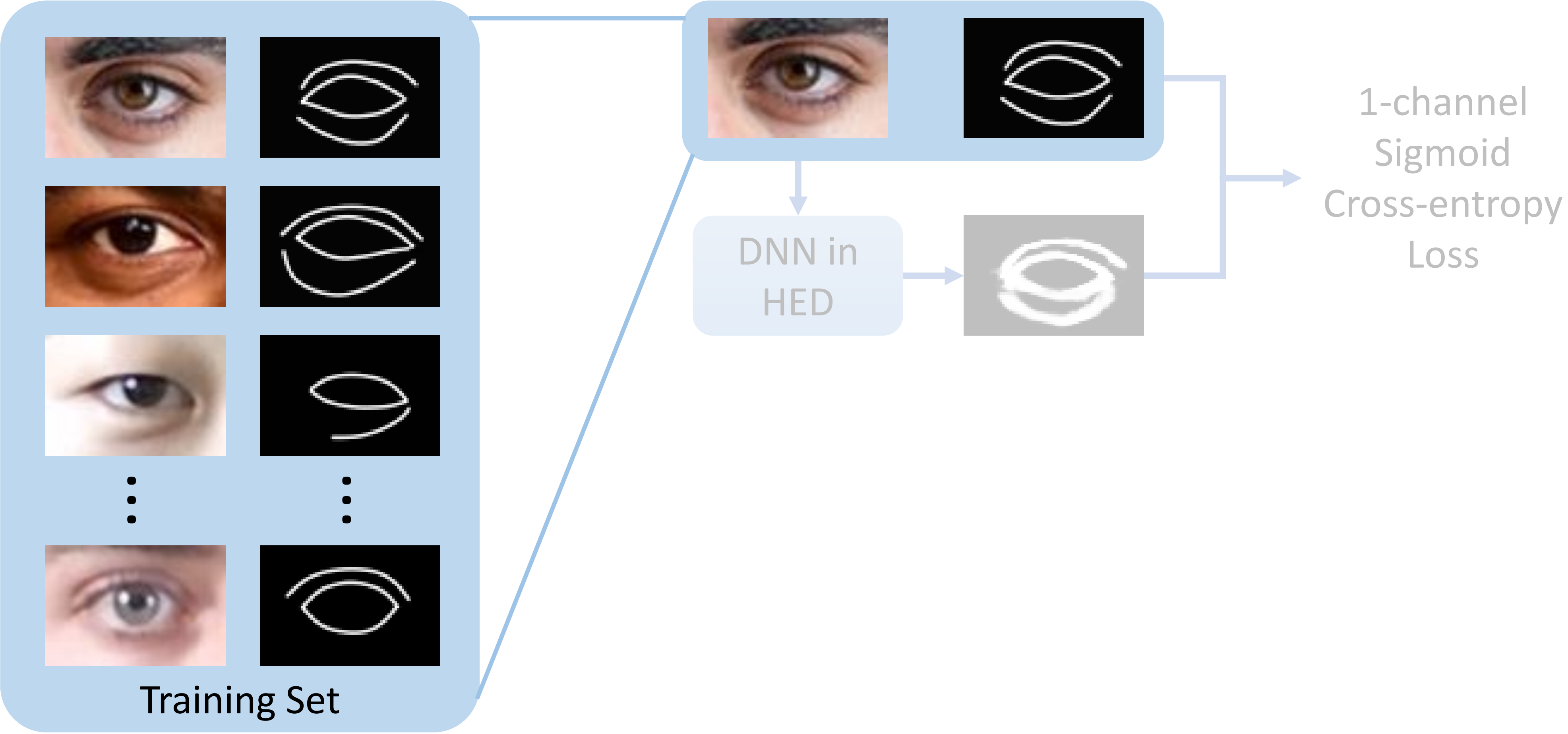
Main features of the eyes:
double-fold, top eyelid, bottom eyelid, bulge

Network

HED
[Xie and Tu 2015]

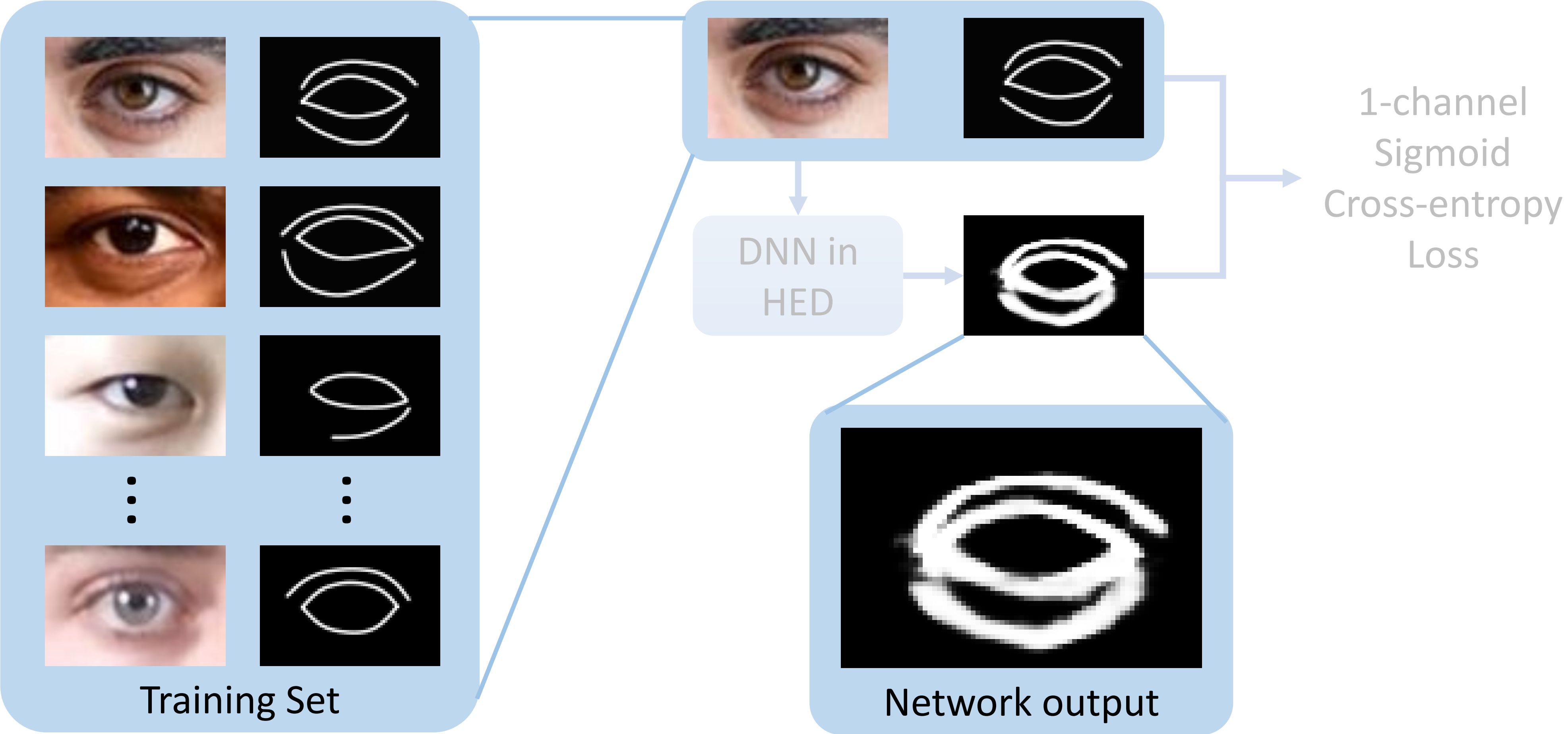


Network



Training Set
DNN in HED
1-channel Sigmoid Cross-entropy Loss

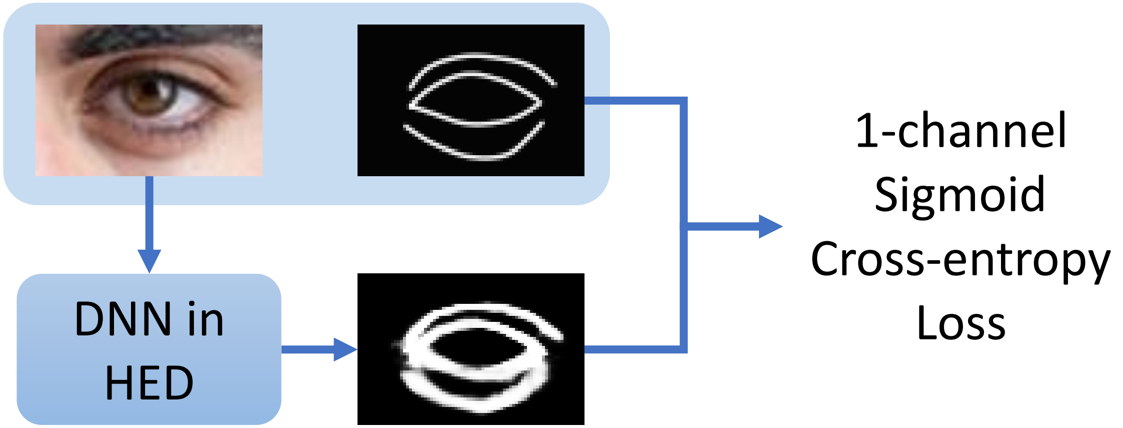
Network



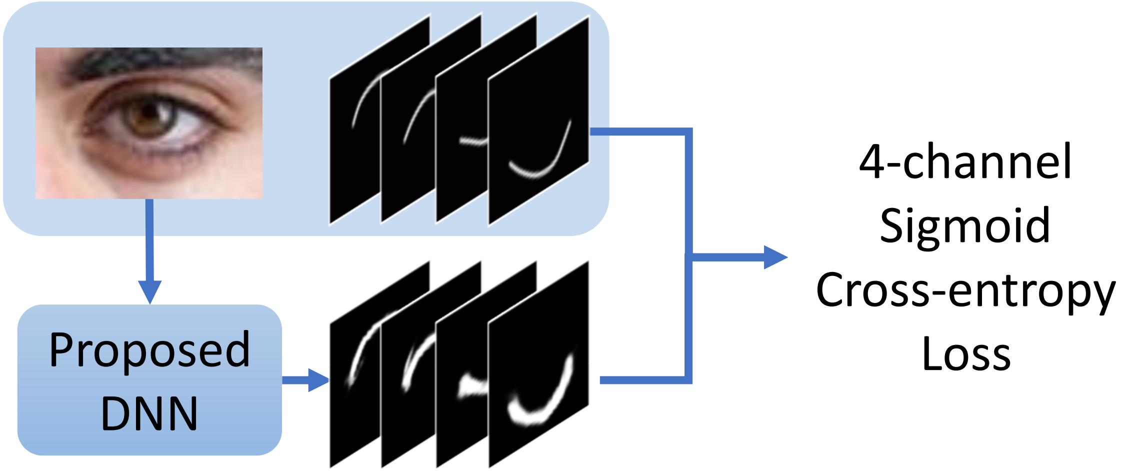
Training Set
DNN in HED
1-channel Sigmoid Cross-entropy Loss
Network output

Network

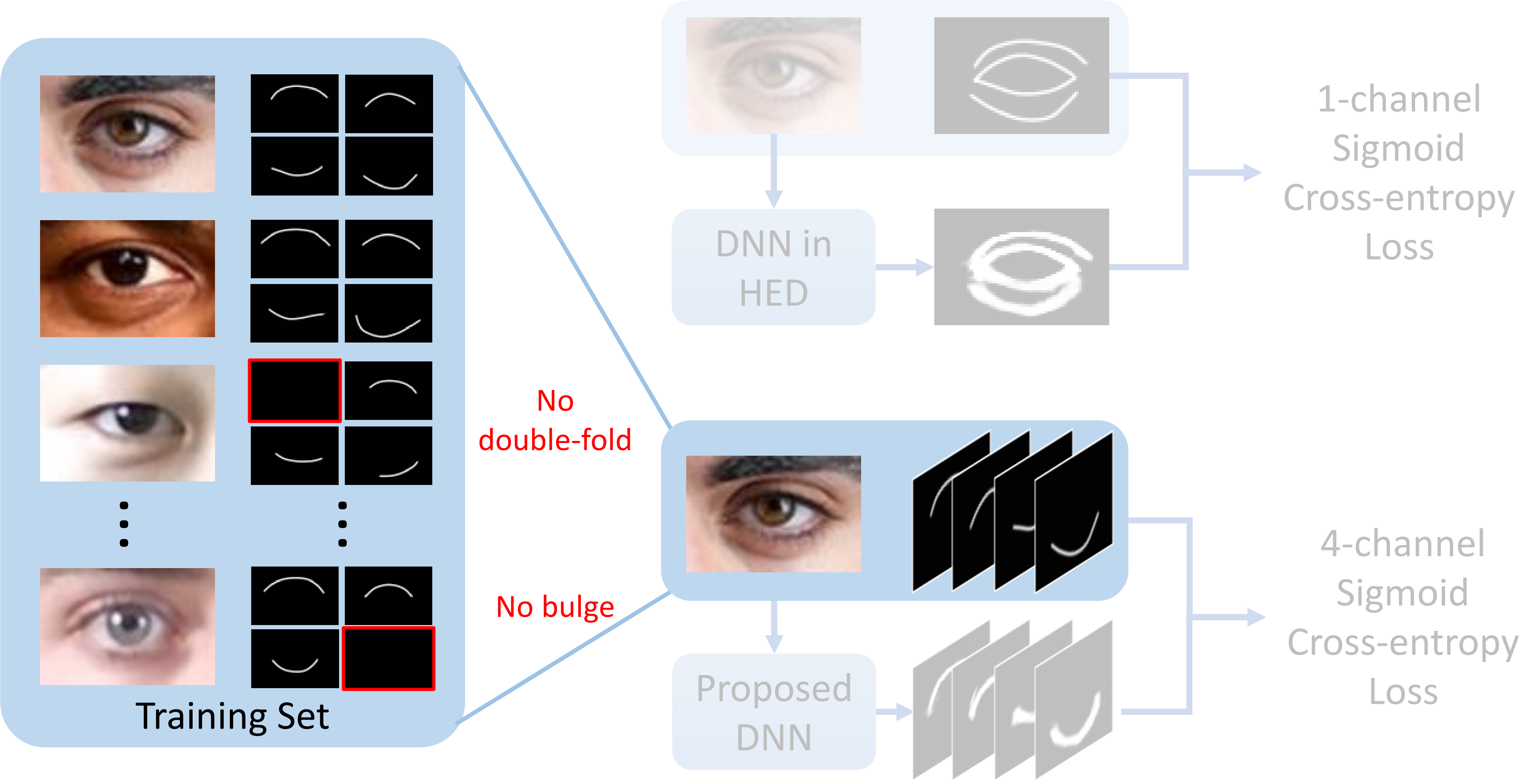
HED
[Xie and Tu 2015]



Proposed eyelid edge detection and identification

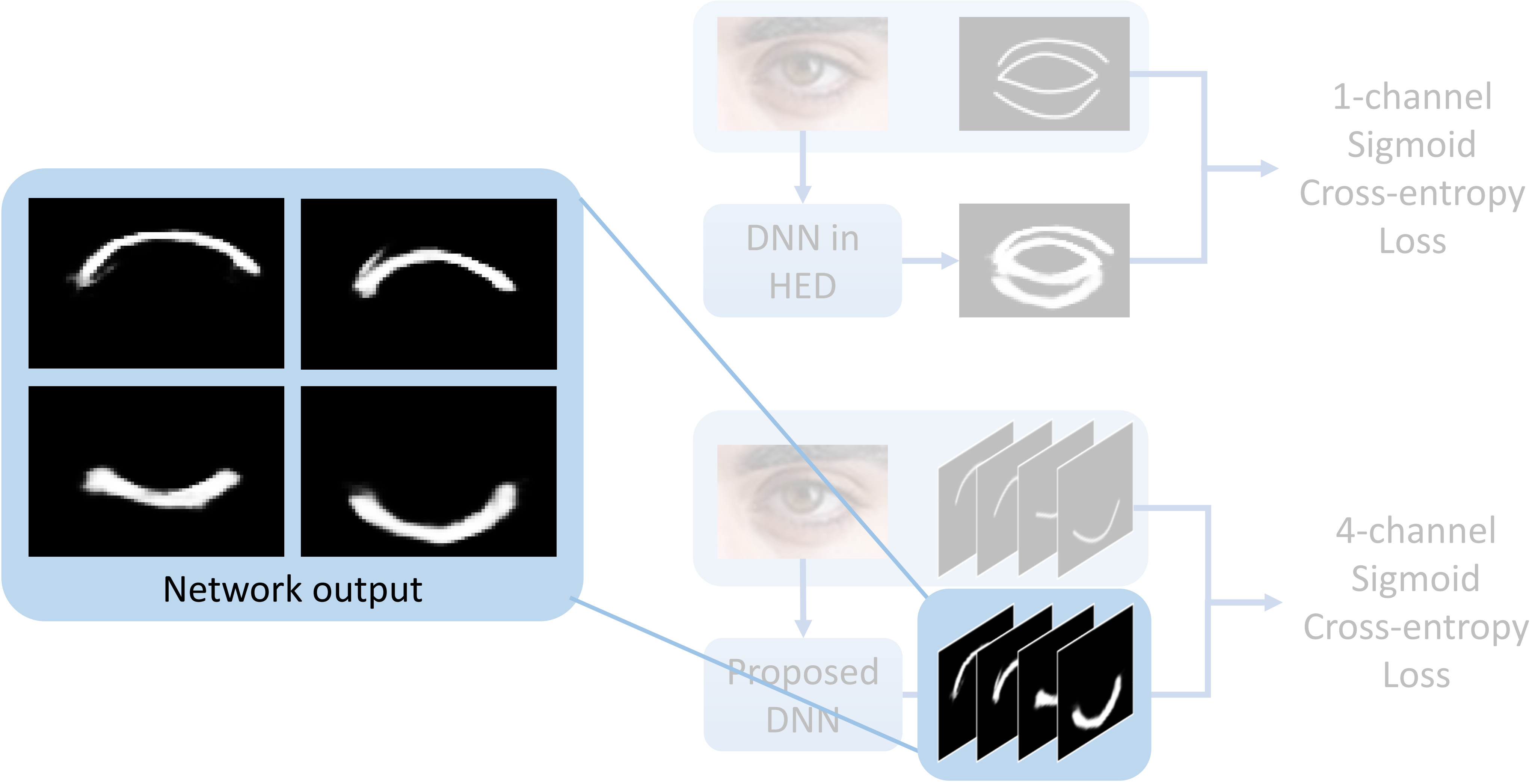


Network



No
double-fold
No bulge
Training Set
DNN in
HED
1-channel
Sigmoid
Cross-entropy
Loss
Proposed
DNN
4-channel
Sigmoid
Cross-entropy
Loss

Network

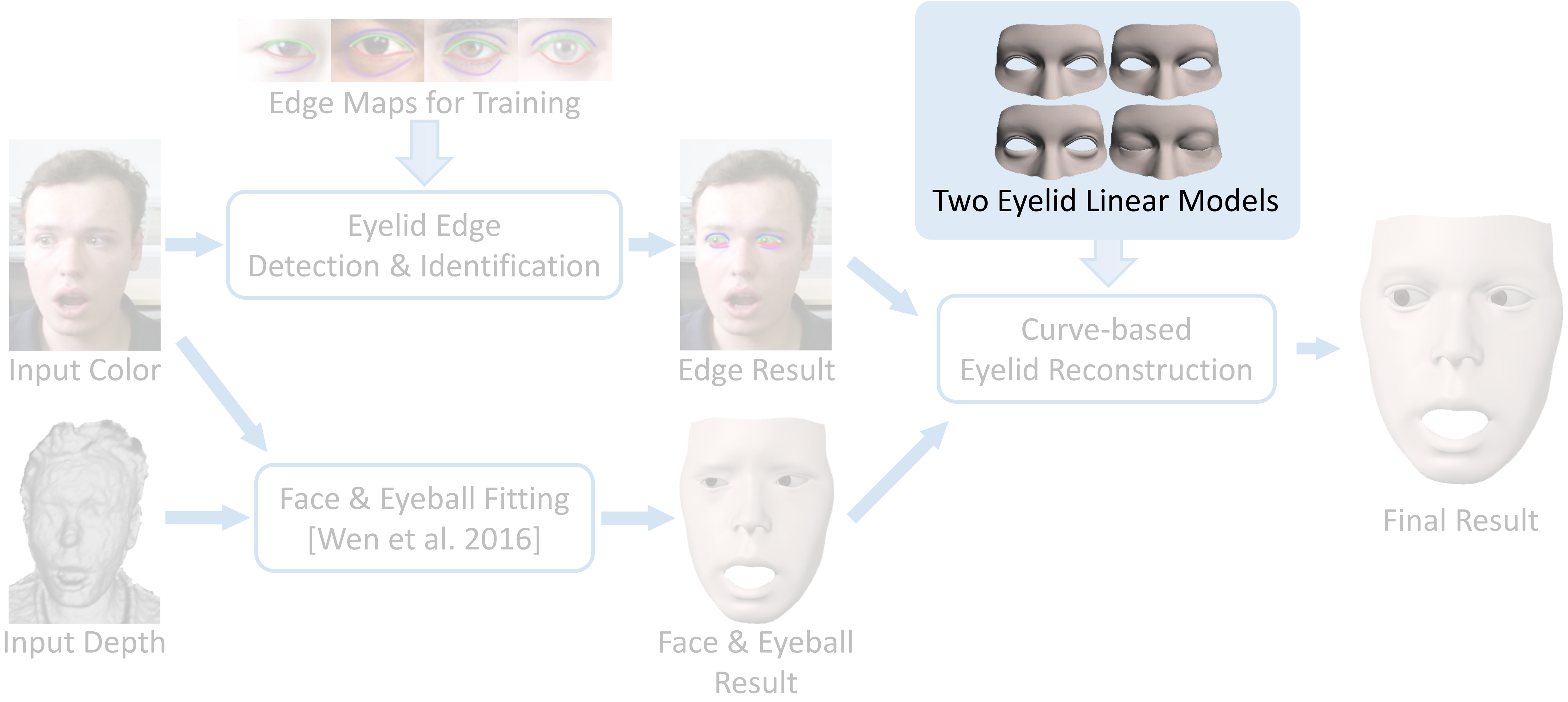


Network output
DNN in HED
1-channel Sigmoid Cross-entropy Loss
Proposed DNN
4-channel Sigmoid Cross-entropy Loss

Eyelid Edge Detection and Identification

Results

Eyelid Linear Models



Shape Linear Rig

Eyelid shape categories

Position

Contour shape

Double-fold

Bulge

Shape Linear Rig

Linear rig B^{id}

$$B^{id} = \{b_k^{id} | k = 0, \dots, N^{id} - 1\}, N^{id} = 29$$

b_k^{id} models in B^{id}

N^{id} number of b_k^{id}

b_0^{id}
(basic)

b_{11}^{id}
(contour: downturned)

b_{21}^{id}
(double-fold: single)

b_{23}^{id}
(bulge: parallel)

Shape Linear Rig

Synthesized shape model of a specific user

$$\color{red}{E_N} = b_0^{id} + \sum_{k=1}^{N^{id}-1} w_k^{id}(b_k^{id} - b_0^{id})$$

b_0^{id}
(basic)

b_{11}^{id}
(contour: downturned)

b_{21}^{id}
(double-fold: single)

b_{23}^{id}
(bulge: parallel)

Shape Linear Rig

Synthesized shape model of a specific user

$$E_N = b_0^{id} + \sum_{k=1}^{N^{id}-1} w_k^{id}(b_k^{id} - b_0^{id})$$

b_0^{id} basic model in B^{id}

b_0^{id}
(basic)

b_{11}^{id}
(contour: downturned)

b_{21}^{id}
(double-fold: single)

b_{23}^{id}
(bulge: parallel)

Shape Linear Rig

Synthesized shape model of a specific user

$$E_N = b_0^{id} + \sum_{k=1}^{N^{id}-1} w_k^{id}(b_k^{id} - b_0^{id})$$

b_0^{id} basic model in B^{id}

b_k^{id} shape models in B^{id}

b_0^{id}
(basic)

b_{11}^{id}
(contour: downturned)

b_{21}^{id}
(double-fold: single)

b_{23}^{id}
(bulge: parallel)

Shape Linear Rig

Synthesized shape model of a specific user

$$E_N = b_0^{id} + \sum_{k=1}^{N^{id}-1} w_k^{id}(b_k^{id} - b_0^{id})$$

b_0^{id} basic model in B^{id}

b_k^{id} shape models in B^{id}

w_k^{id} weight of b_k^{id}

b_0^{id}
(basic)

b_{11}^{id}
(contour: downturned)

b_{21}^{id}
(double-fold: single)

b_{23}^{id}
(bulge: parallel)

Shape Linear Rig

Synthesized shape model of a specific user

E_N

User-specific shape model

$=$

b_0^{id} (basic)

$+ \sum_{k=1}^{N^{id}-1} w_k^{id} ($

b_k^{id} (shape)

$-$

b_0^{id} (basic)

$)$

Pose Linear Rig

Generic linear rig B^{exp}

$$B^{exp} = \{b_k^{exp} | k = 0, \dots, N^{exp} - 1\}, N^{exp} = 23$$

b_k^{exp} models in B^{exp}

N^{exp} number of b_k^{exp}

b_0^{exp}
(basic)

b_1^{exp}
(close)

b_3^{exp}
(inner close)

b_5^{exp}
(outer close)

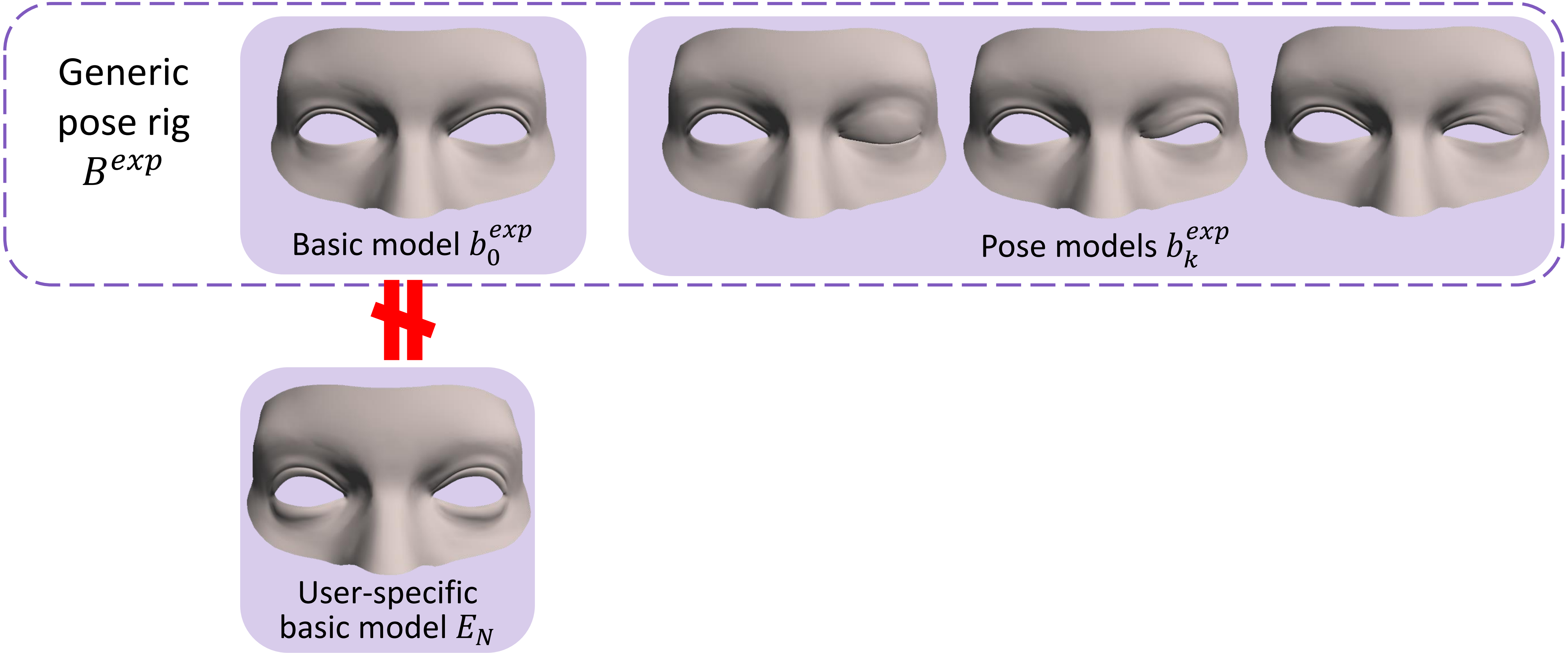
Pose Linear Rig

Generic pose rig B^{exp}

Basic model b_0^{exp}

Pose models b_k^{exp}

Pose Linear Rig



Pose Linear Rig

Generic pose rig B^{exp}

Basic model b_0^{exp}

Pose models b_k^{exp}

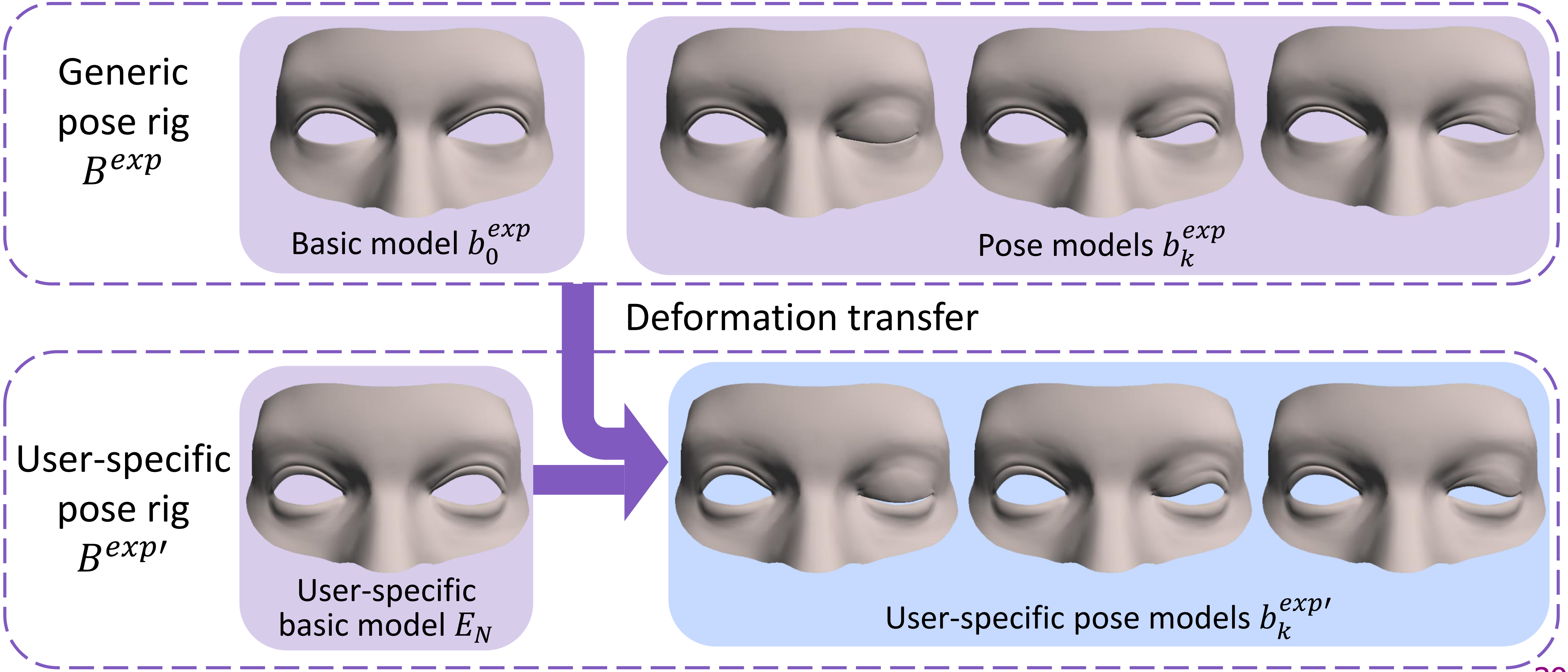
User-specific pose rig $B^{exp'}$

User-specific basic model E_N

?

User-specific pose models $b_k^{exp'}$

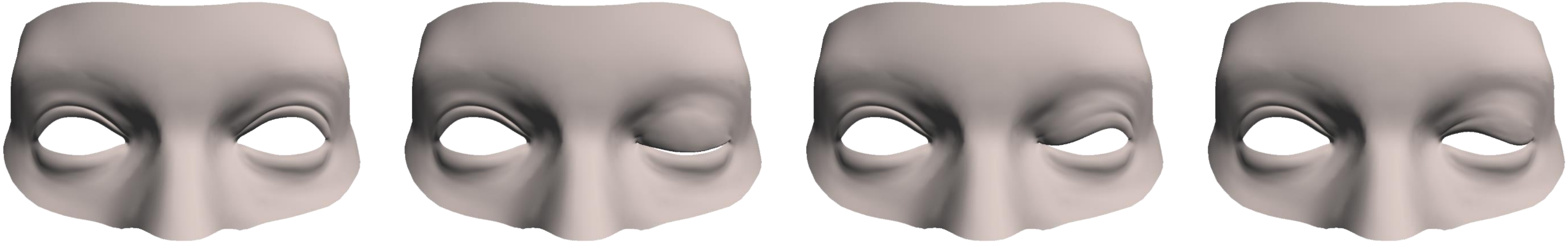
Pose Linear Rig



Pose Linear Rig

User-specific eyelid model in tracking

$$E_P = b_0^{exp'} + \sum_{k=1}^{N^{exp}-1} w_k^{exp}(b_k^{exp'} - b_0^{exp'})$$



$b_0^{exp'}$ (basic) | $b_1^{exp'}$ (close) | $b_3^{exp'}$ (inner close) | $b_5^{exp'}$ (outer close)

Pose Linear Rig

User-specific eyelid model in tracking

$$E_P = b_0^{exp'} + \sum_{k=1}^{N^{exp}-1} w_k^{exp} (b_k^{exp'} - b_0^{exp'})$$

$b_0^{exp'}$ basic model in $B^{exp'}$

$b_0^{exp'}$
(basic)

$b_1^{exp'}$
(close)

$b_3^{exp'}$
(inner close)

$b_5^{exp'}$
(outer close)

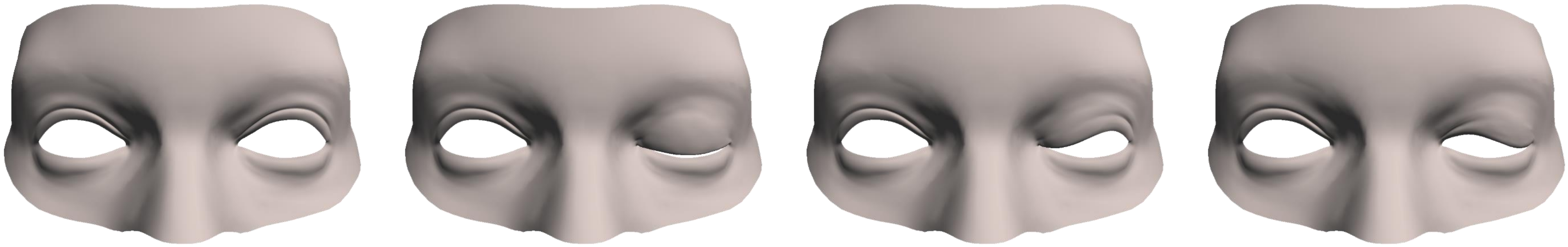
Pose Linear Rig

User-specific eyelid model in tracking

$$E_P = b_0^{exp'} + \sum_{k=1}^{N^{exp}-1} w_k^{exp} (b_k^{exp'} - b_0^{exp'})$$

$b_0^{exp'}$ basic model in $B^{exp'}$

$b_k^{exp'}$ pose models in $B^{exp'}$



$b_0^{exp'}$ (basic) $b_1^{exp'}$ (close) $b_3^{exp'}$ (inner close) $b_5^{exp'}$ (outer close)

Pose Linear Rig

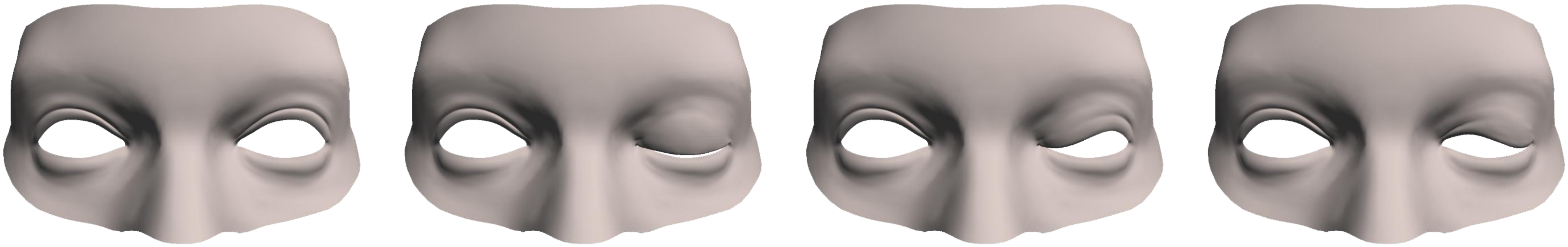
User-specific eyelid model in tracking

$$E_P = b_0^{exp'} + \sum_{k=1}^{N^{exp}-1} \textcolor{red}{w_k^{exp}} (b_k^{exp'} - b_0^{exp'})$$

$b_0^{exp'}$ basic model in $B^{exp'}$

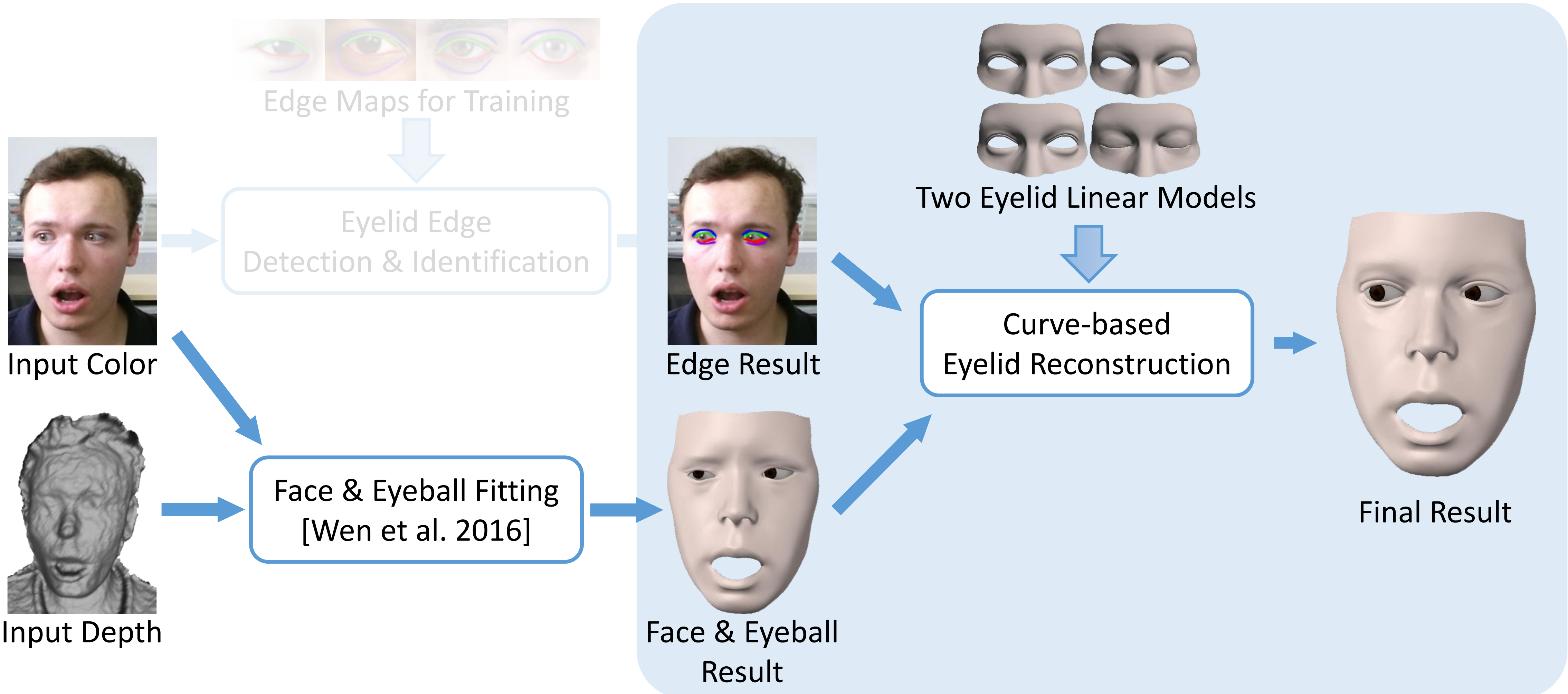
$b_k^{exp'}$ pose models in $B^{exp'}$

$\textcolor{red}{w_k^{exp}}$ weight of $b_k^{exp'}$



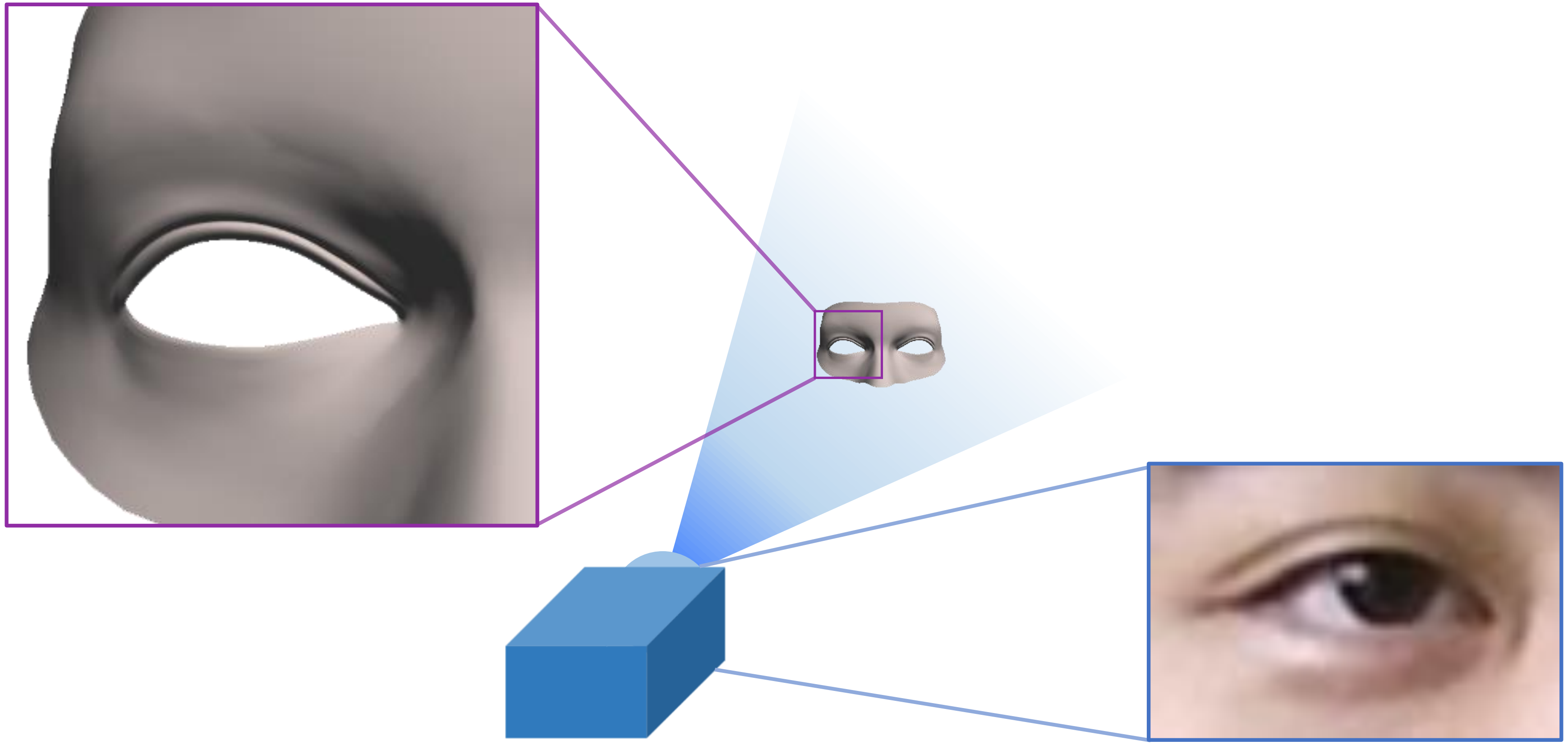
$b_0^{exp'}$ (basic) | $b_1^{exp'}$ (close) | $b_3^{exp'}$ (inner close) | $b_5^{exp'}$ (outer close)

Curve-based Eyelid Reconstruction



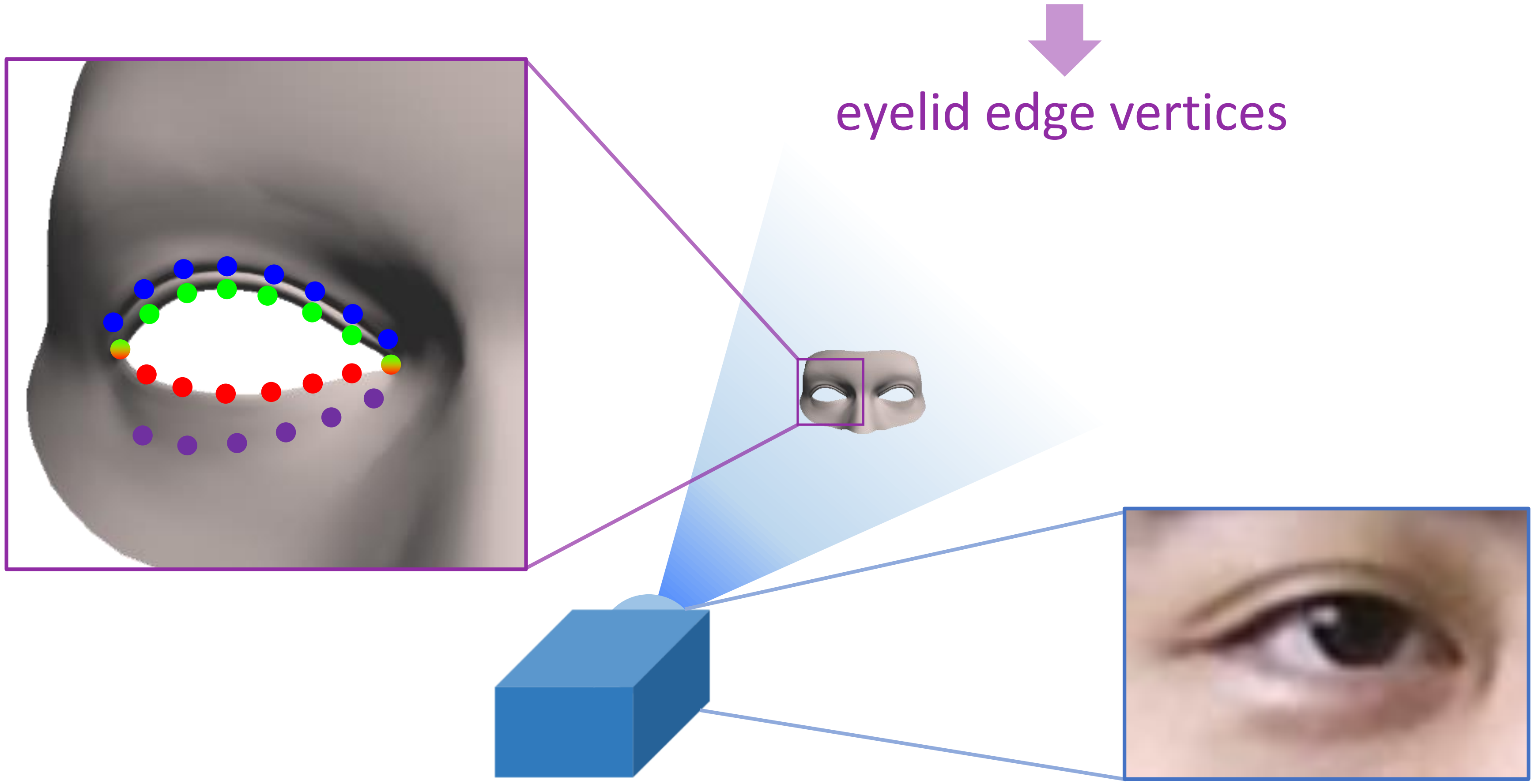
Curve-based Eyelid Reconstruction

Minimize the inconsistency between the projected eyelid model and the real image



Curve-based Eyelid Reconstruction

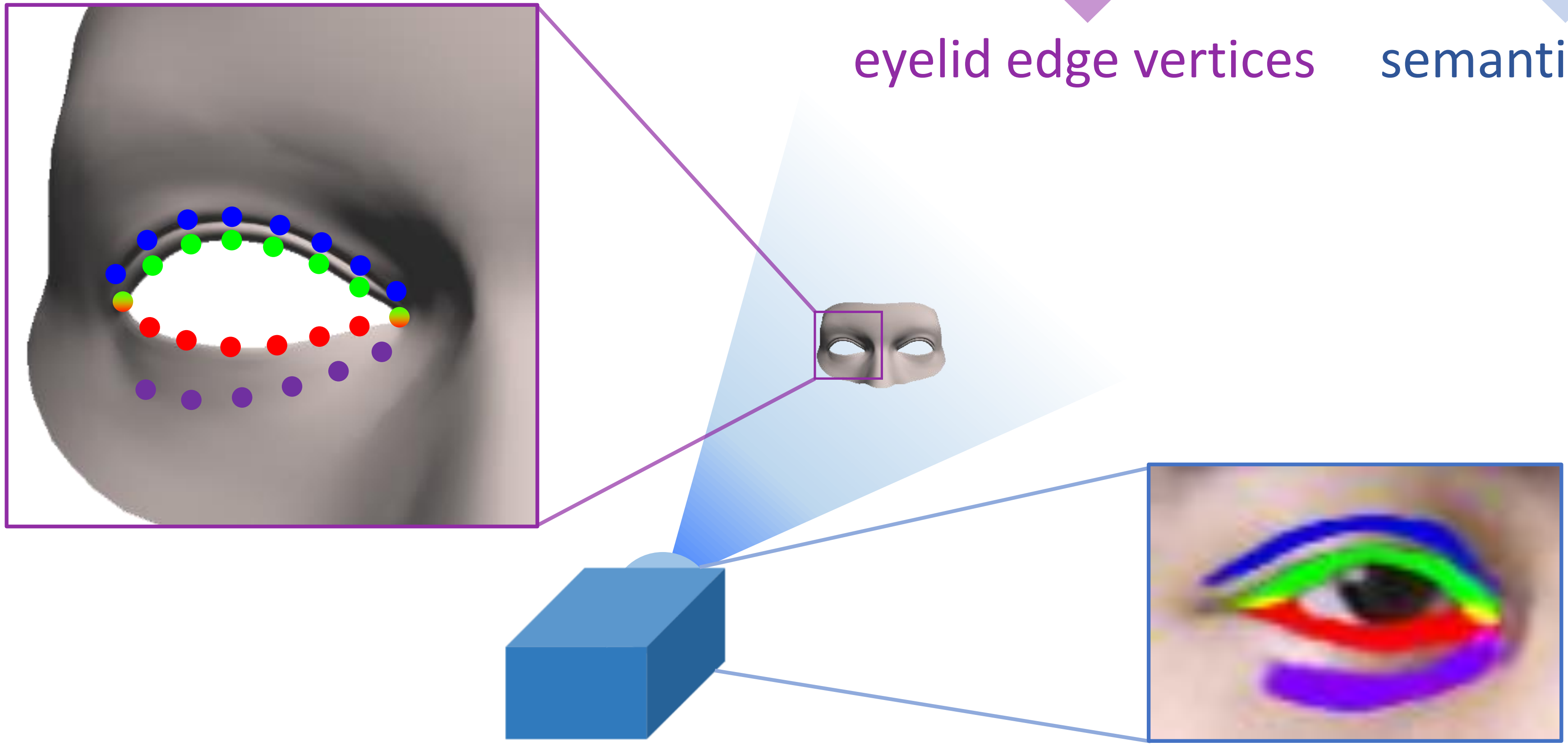
Minimize the inconsistency between the projected eyelid model and the real image



Curve-based Eyelid Reconstruction

Minimize the inconsistency between the projected eyelid model and the real image

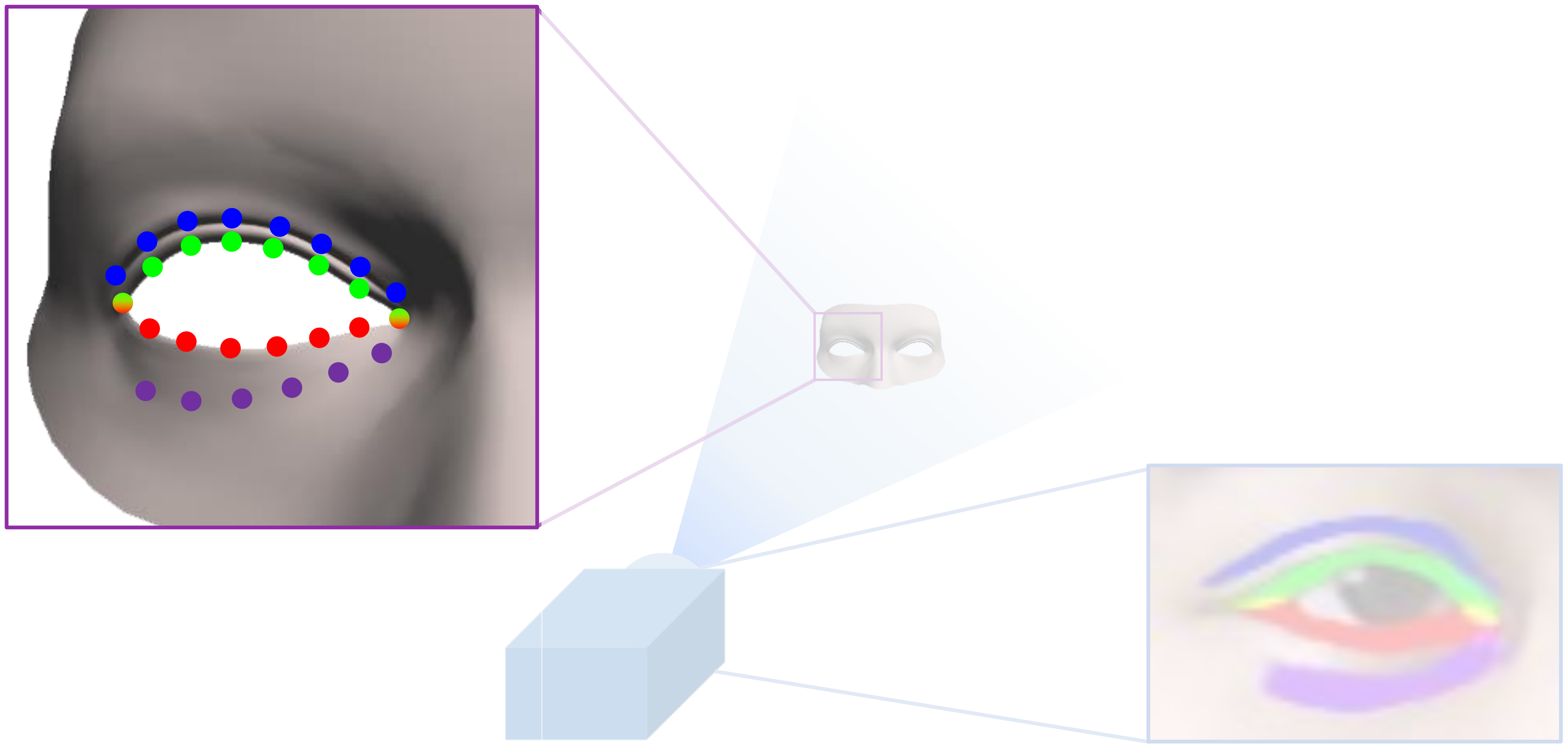
eyelid edge vertices   semantic edges



Correspondence obtaining

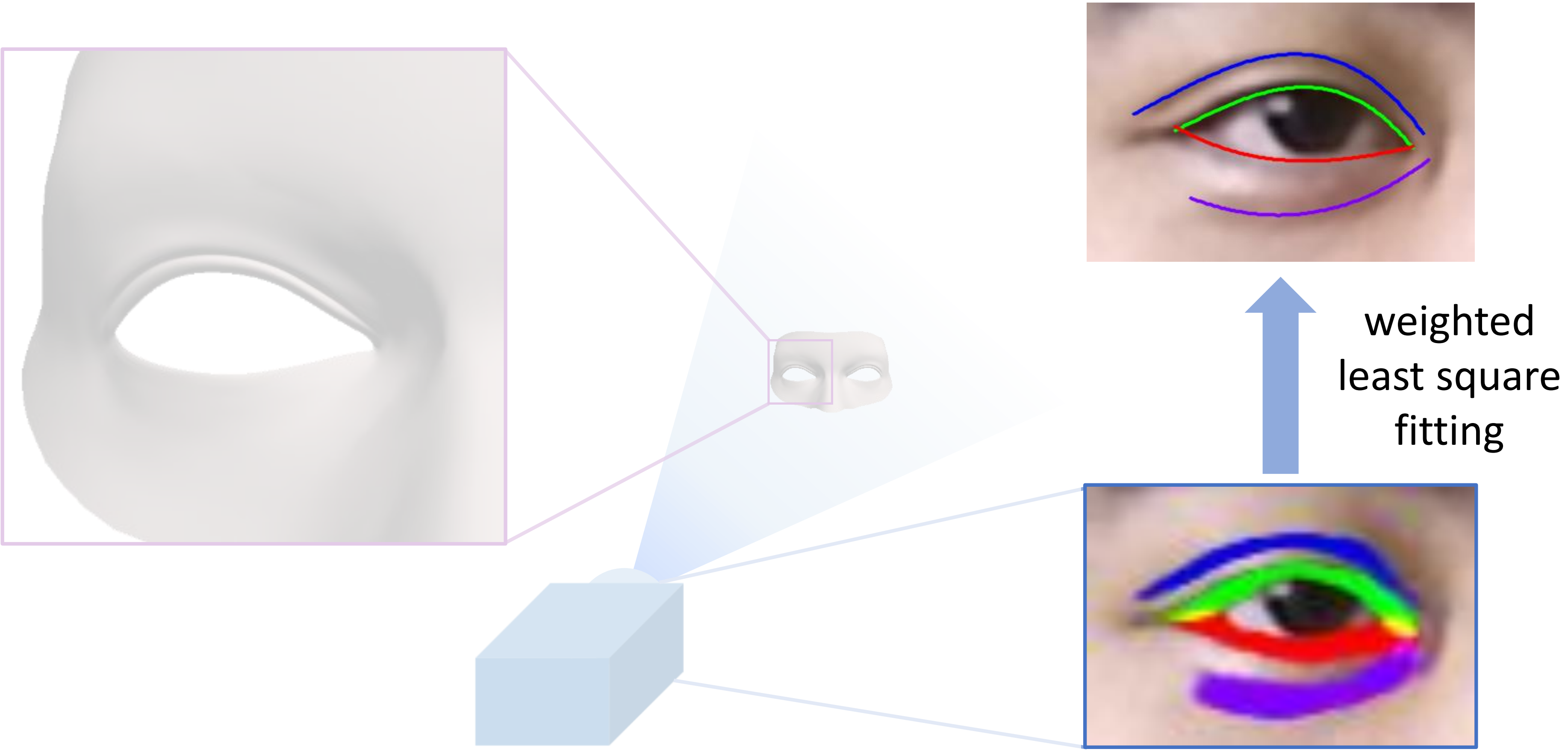
Label 3D edge vertices on the eyelid model as 3D landmarks

3D landmarks
double-fold
top eyelid
bottom eyelid
bulge



Curve Fitting

Fit four polynomials according to the semantic edge map



Correspondence obtaining

Obtain 2D landmarks according to relative curve length

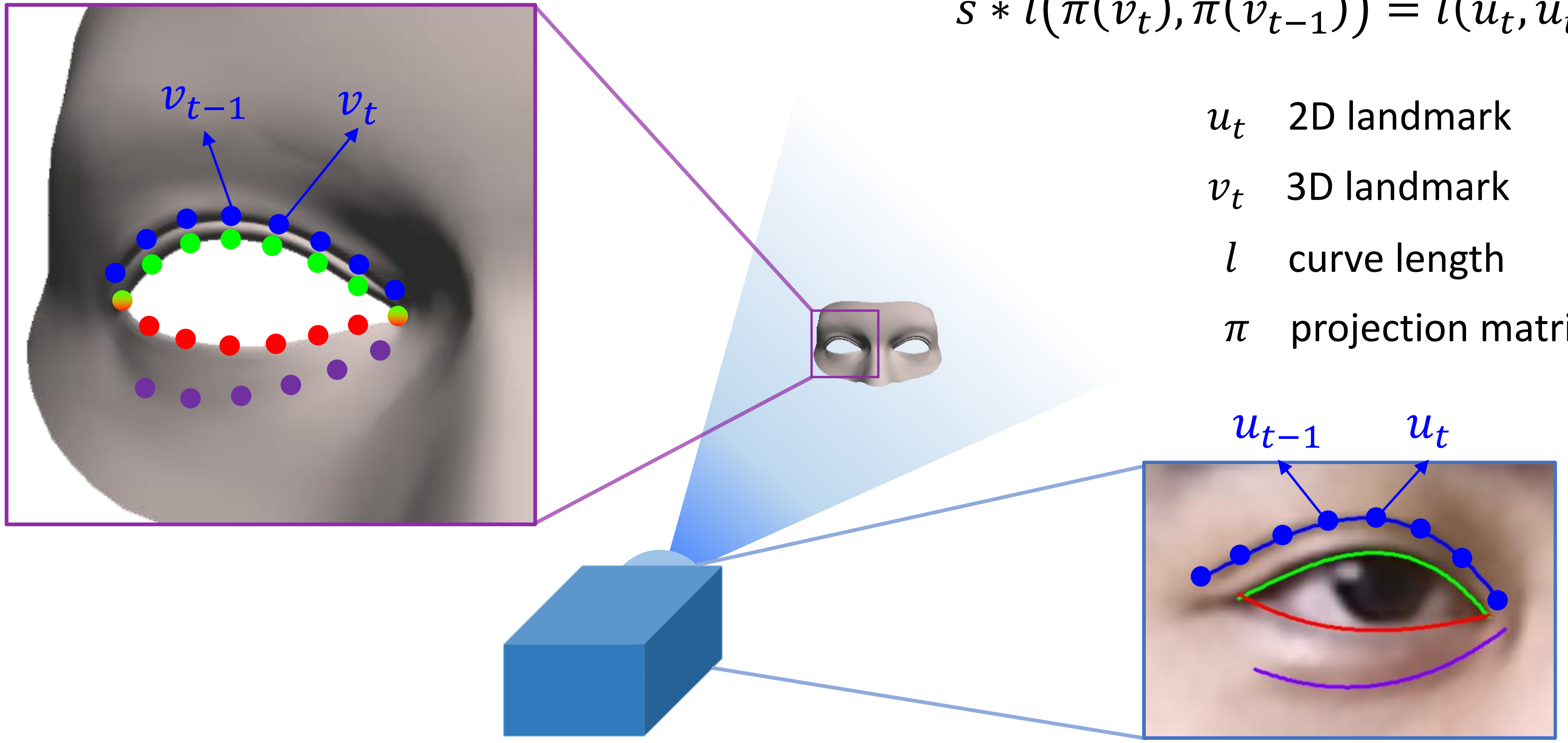
$$s * l\left(\pi(v_t), \pi(v_{t-1})\right) = l(u_t, u_{t-1})$$

u_t 2D landmark

v_t 3D landmark

l curve length

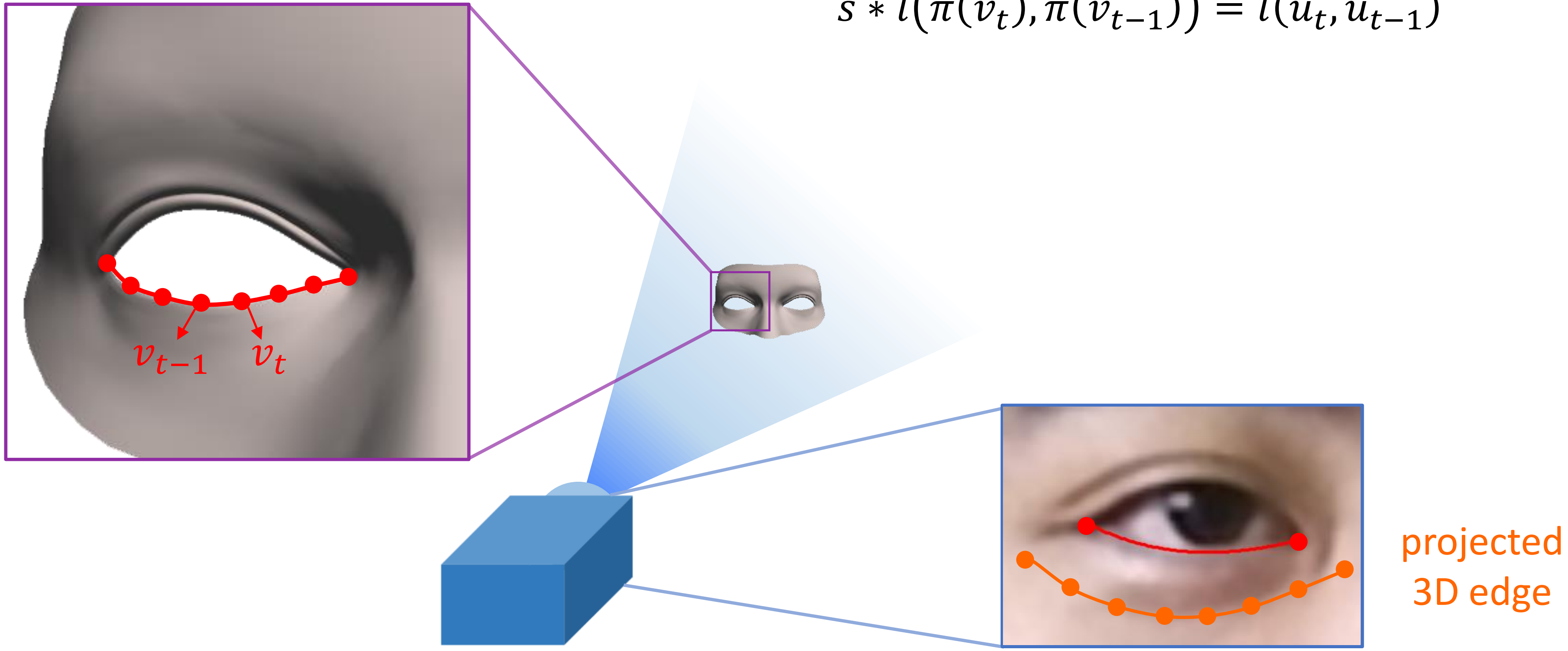
π projection matrix



Correspondence obtaining

Obtain 2D landmarks according to relative curve length

$$s * l(\pi(v_t), \pi(v_{t-1})) = l(u_t, u_{t-1})$$

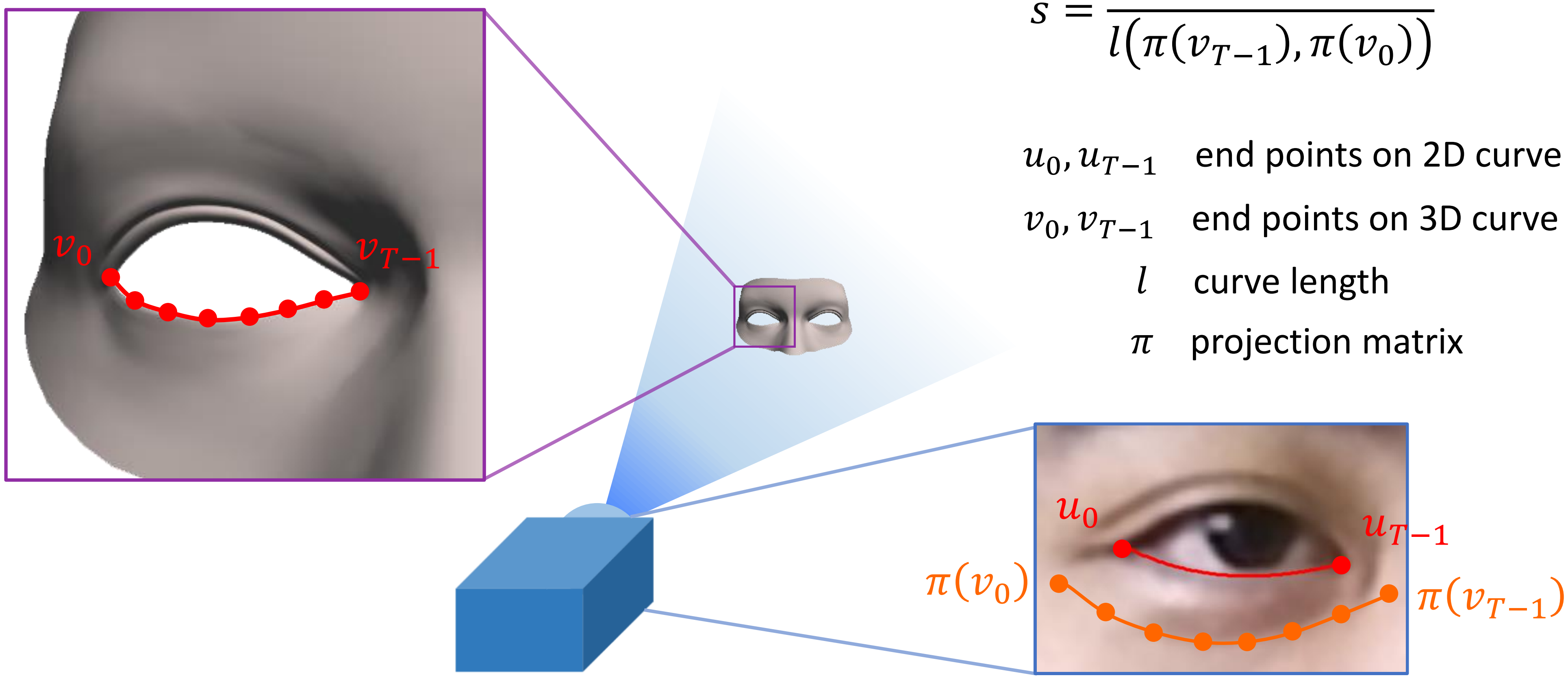


Correspondence obtaining

Curve length ratio s

$$s = \frac{l(u_{T-1}, u_0)}{l\big(\pi(v_{T-1}), \pi(v_0)\big)}$$

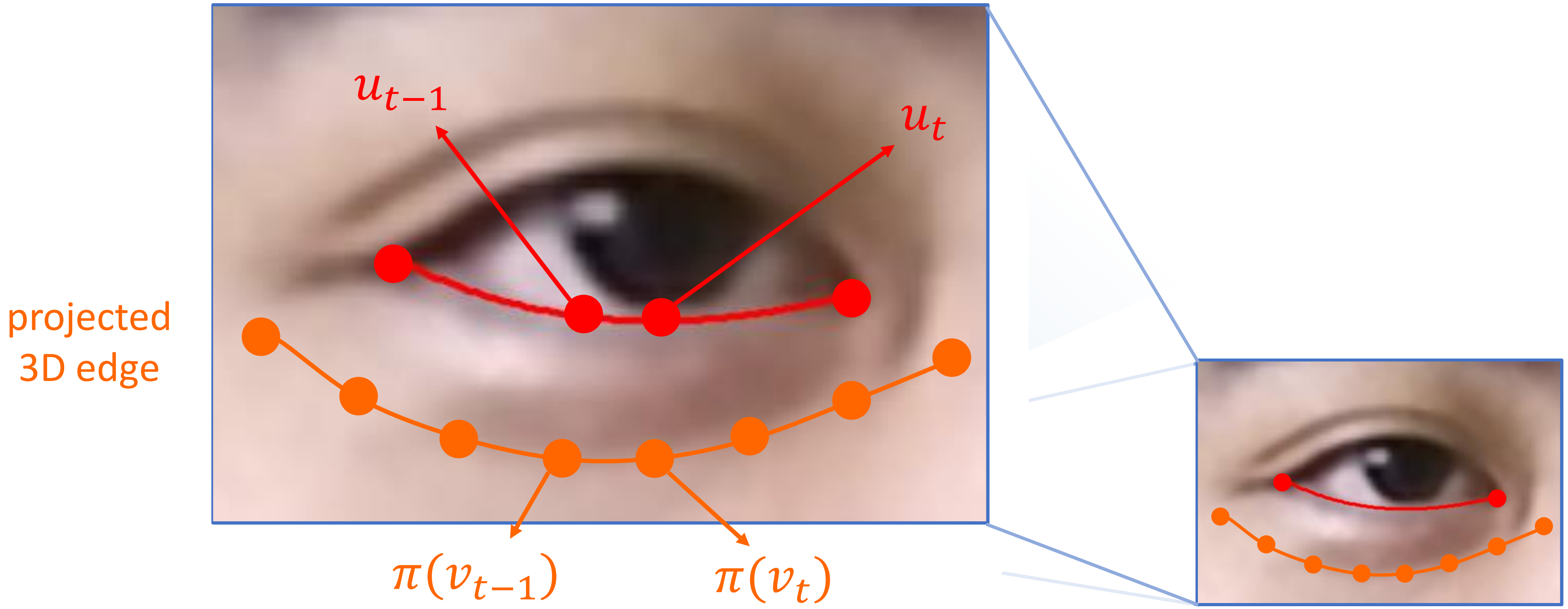
u_0, u_{T-1}	end points on 2D curve
v_0, v_{T-1}	end points on 3D curve
l	curve length
π	projection matrix



Correspondence obtaining

Obtain 2D landmarks according to relative curve length

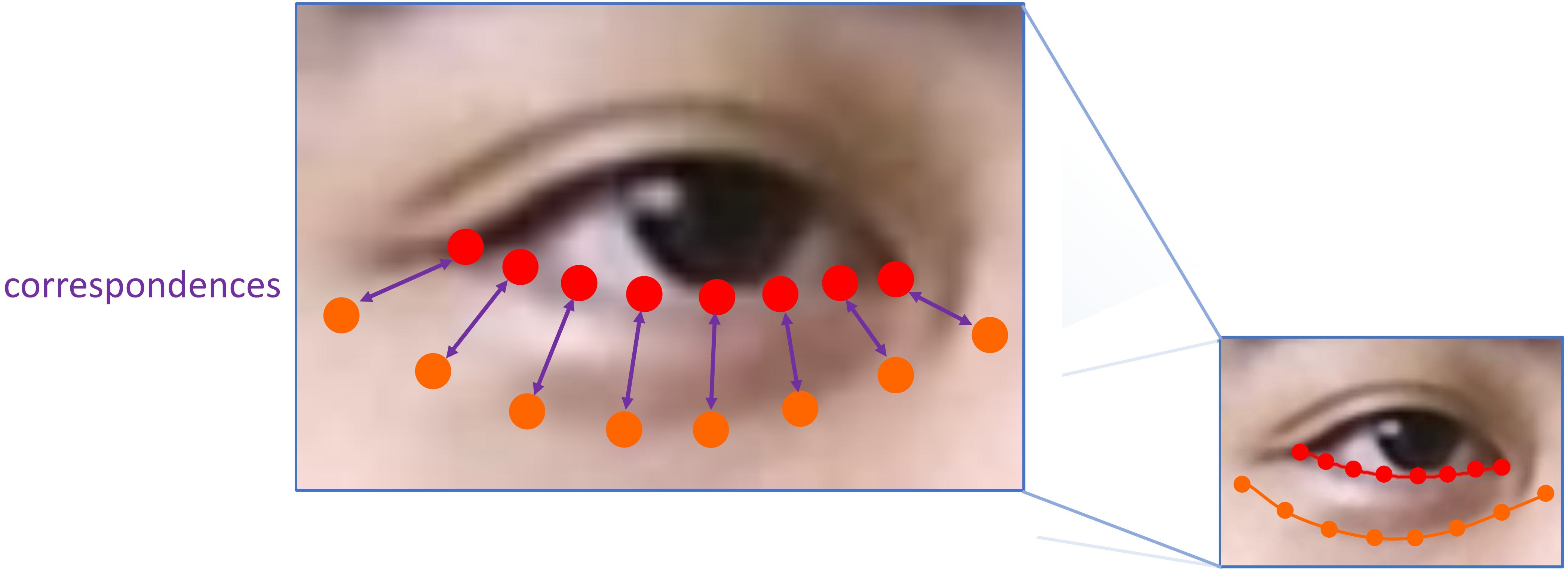
$$s * l\big(\pi(v_t), \pi(v_{t-1})\big) = l(u_t, u_{t-1})$$



Correspondence obtaining

Obtain 2D landmarks according to relative curve length

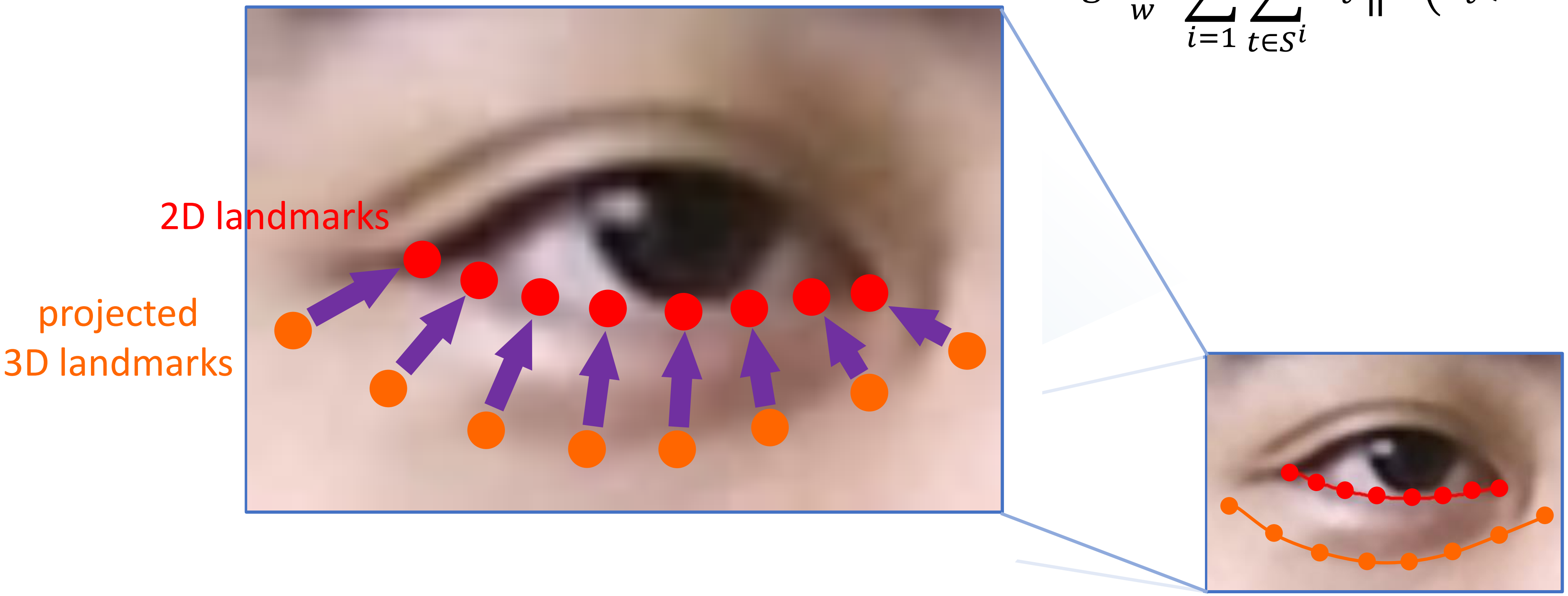
$$s * l\big(\pi(v_t), \pi(v_{t-1})\big) = l(u_t, u_{t-1})$$



Eyelid Reconstruction

Minimize the distances between the projected 3D eyelid landmarks and the 2D eyelid landmarks

$$\arg\min_{w} \sum_{i=1}^{4} \sum_{t \in S^i} \alpha_t^i \left\| \pi\left(v_t^i(w,B)\right) - u_t^i \right\|_2^2$$

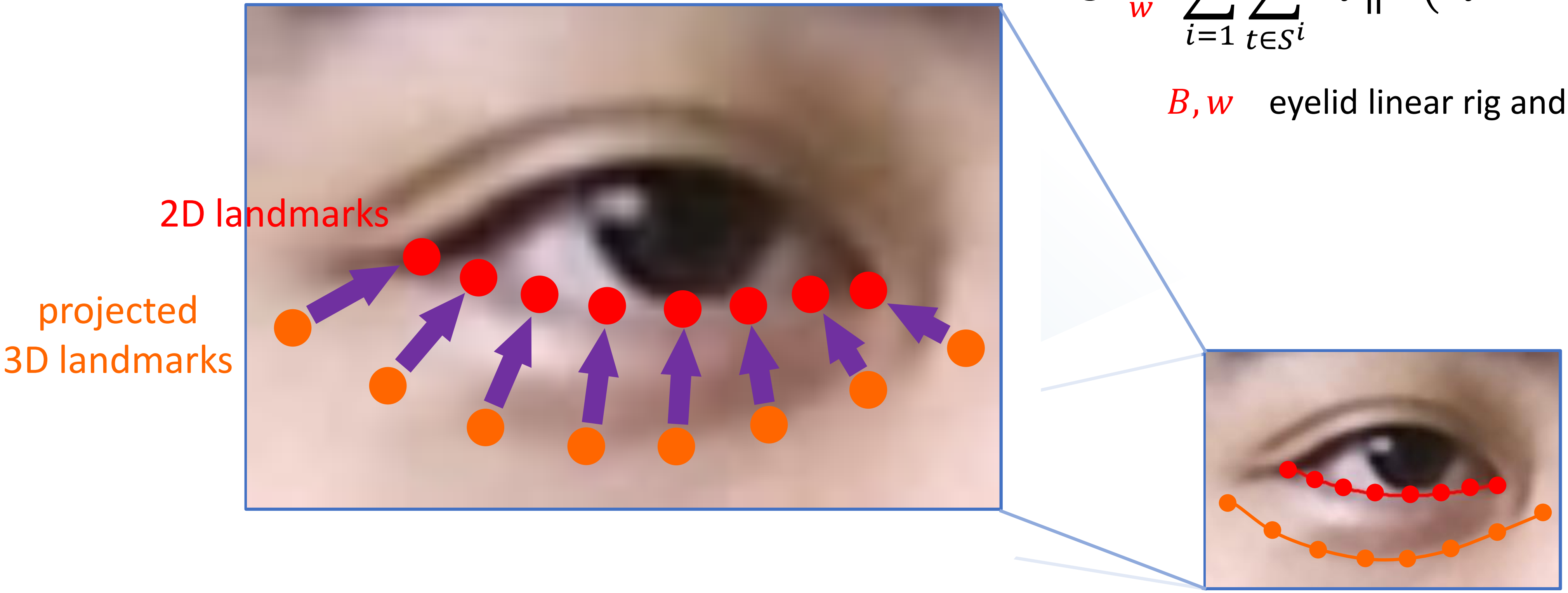


Eyelid Reconstruction

Minimize the distances between the projected 3D eyelid landmarks and the 2D eyelid landmarks

$$\arg\min_{\color{red}{w}} \sum_{i=1}^{4} \sum_{t \in S^i} \alpha_t^i \left\| \pi\left(v_t^i(\color{red}{w, B})\right) - u_t^i \right\|_2^2$$

$\color{red}{B, w}$ eyelid linear rig and weights



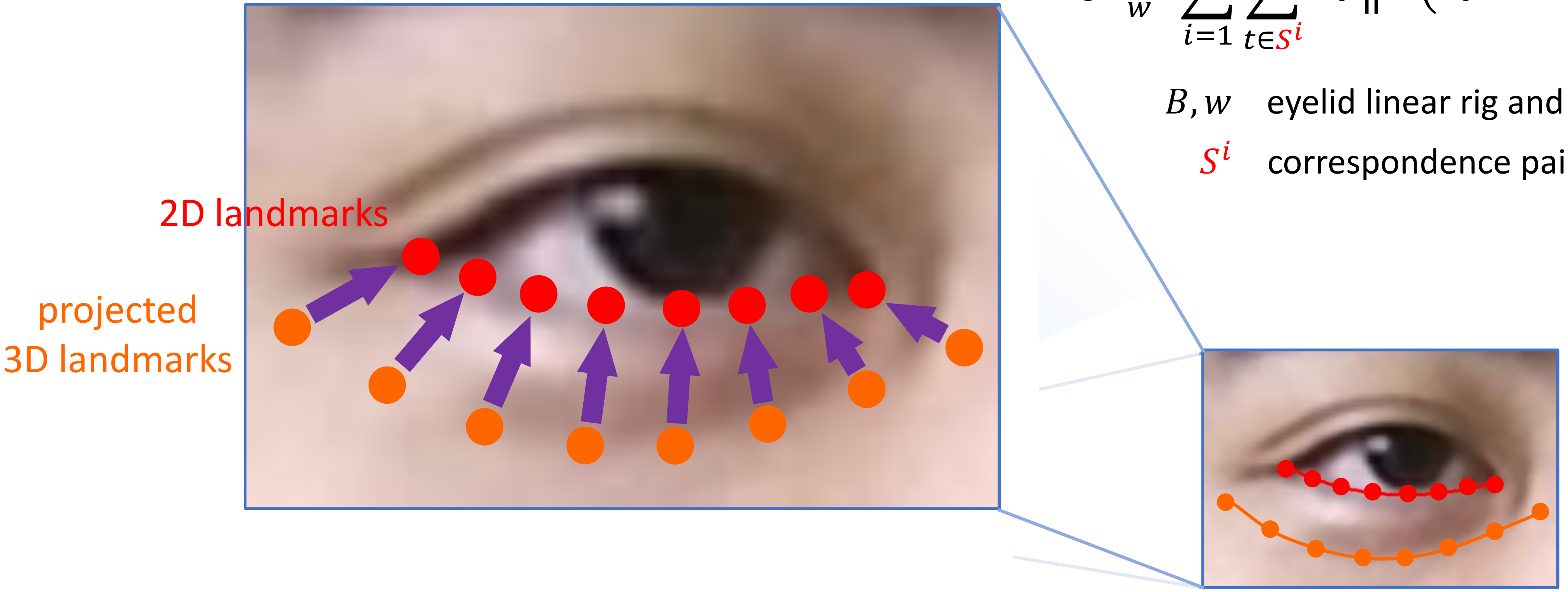
Eyelid Reconstruction

Minimize the distances between the projected 3D eyelid landmarks and the 2D eyelid landmarks

$$\arg\min_{w} \sum_{i=1}^{4} \sum_{t \in S^i} \alpha_t^i \left\| \pi\left(v_t^i(w, B)\right) - u_t^i \right\|_2^2$$

B, w eyelid linear rig and weights

S^i correspondence pairs of edge i



Eyelid Reconstruction

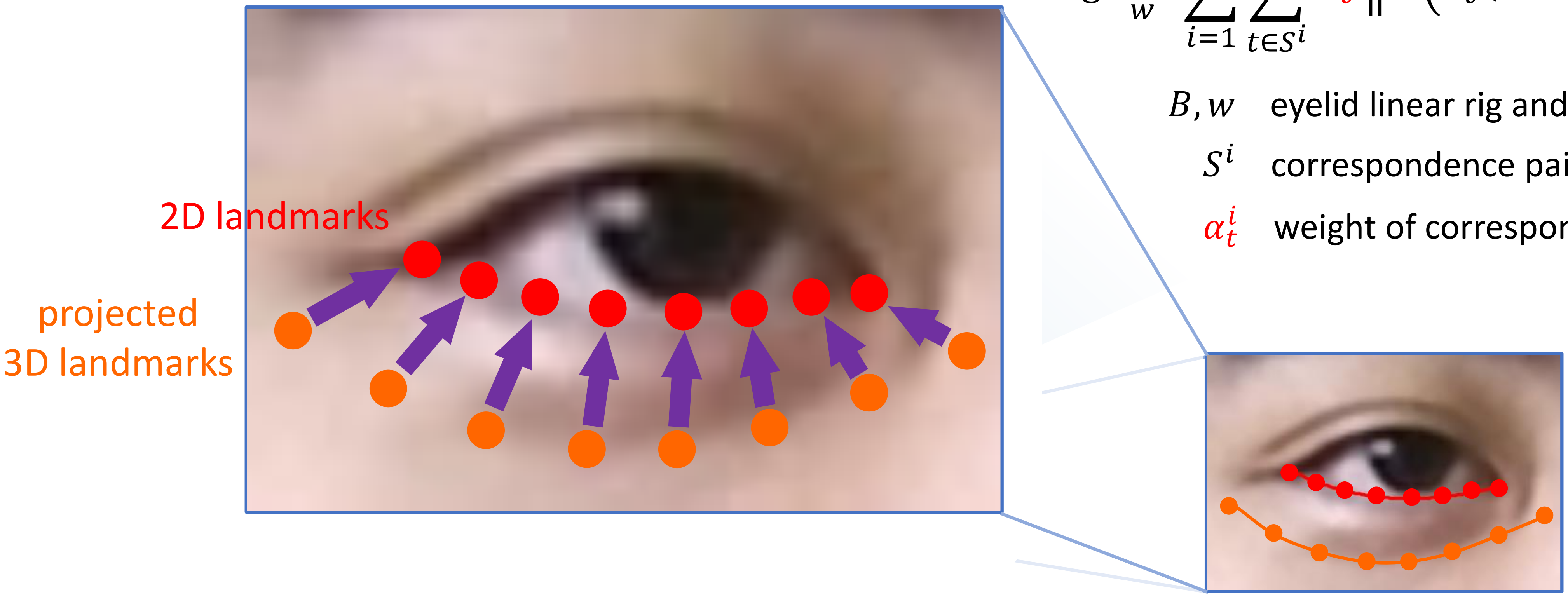
Minimize the distances between the projected 3D eyelid landmarks and the 2D eyelid landmarks

$$\arg\min_{w} \sum_{i=1}^{4} \sum_{t \in S^i} \alpha_t^i \left\| \pi\left(v_t^i(w, B)\right) - u_t^i \right\|_2^2$$

B, w eyelid linear rig and weights

S^i correspondence pairs of edge i

α_t^i weight of correspondence



Eyelid Reconstruction

Solve for the optimal weights w_{opt}

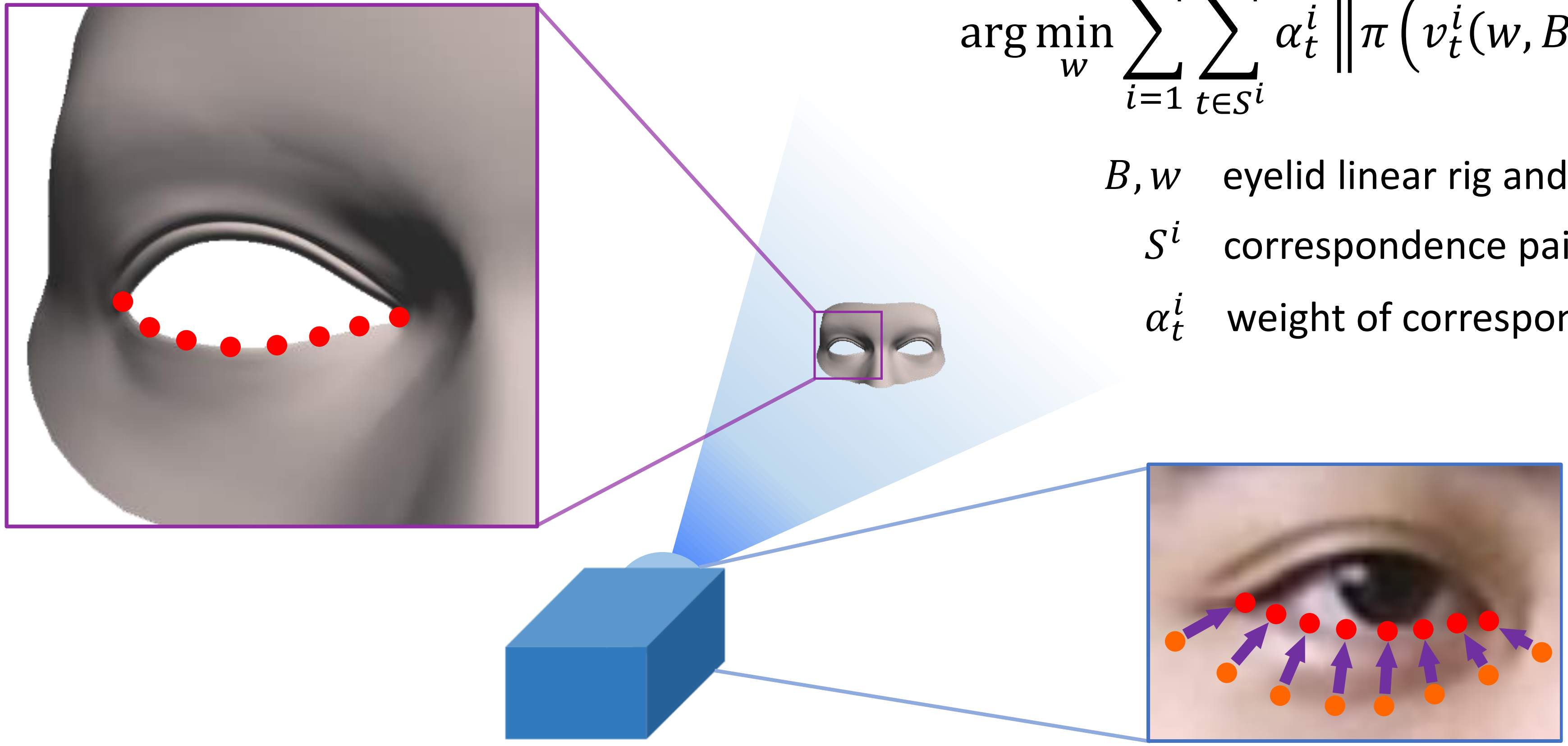
$$\arg\min_{w} \sum_{i=1}^{4} \sum_{t \in S^i} \alpha_t^i \left\| \pi\left(v_t^i(w, B)\right) - u_t^i \right\|_2^2$$

B, w eyelid linear rig and weights

S^i correspondence pairs of edge i

α_t^i weight of correspondence

initial weights
w_0



Eyelid Reconstruction

Solve for the optimal weights w_{opt}

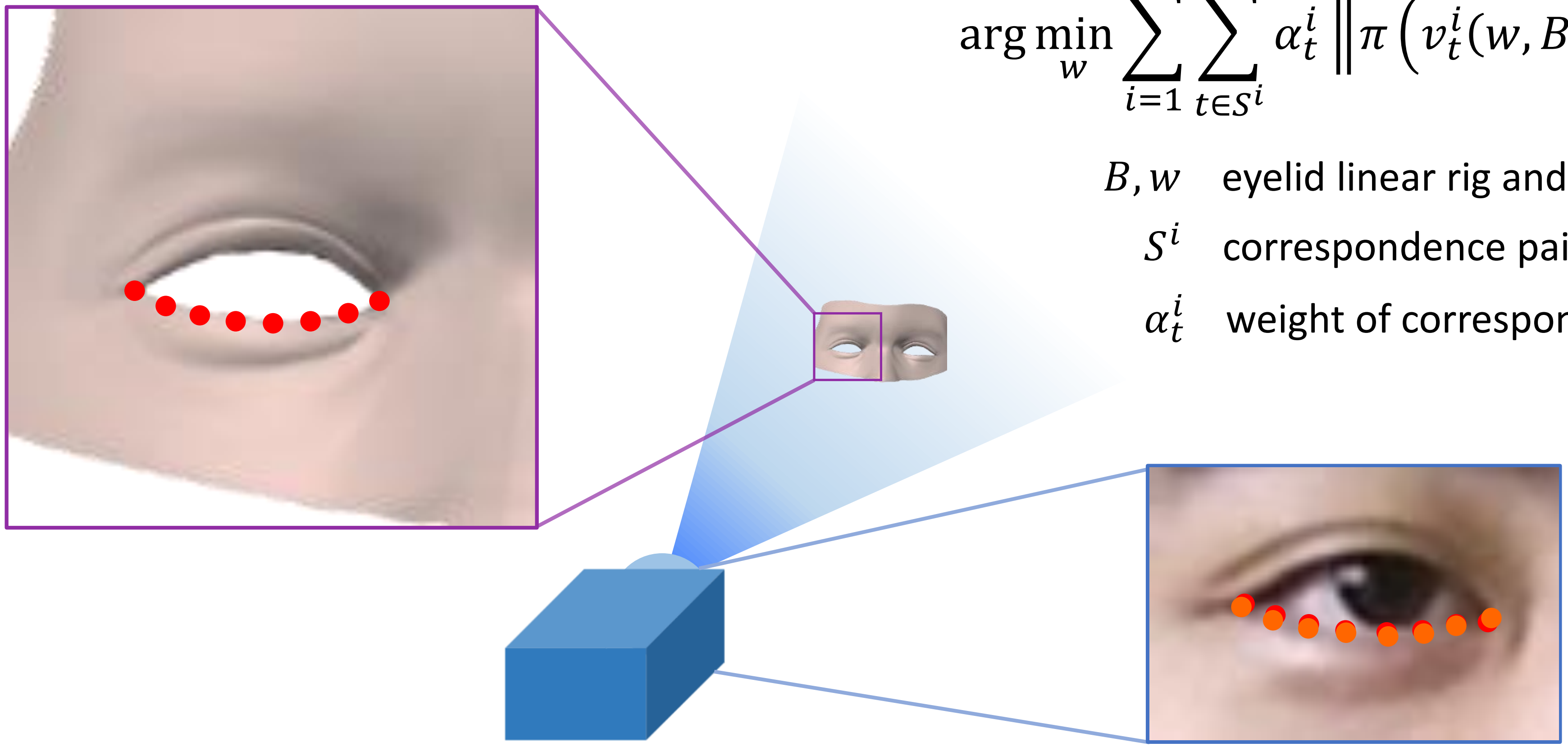
$$\arg\min_{w} \sum_{i=1}^{4} \sum_{t \in S^i} \alpha_t^i \left\| \pi\left(v_t^i(w, B)\right) - u_t^i \right\|_2^2$$

B, w eyelid linear rig and weights

S^i correspondence pairs of edge i

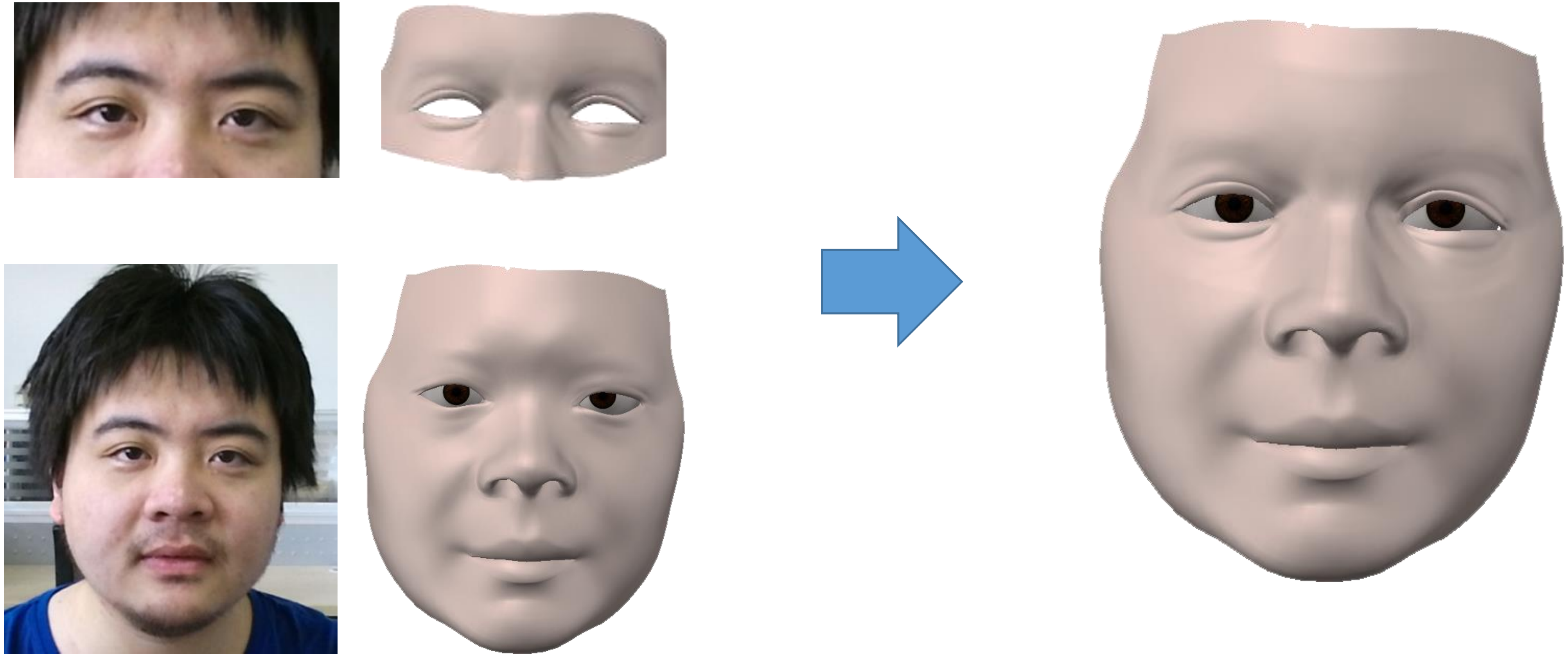
α_t^i weight of correspondence

optimal weights w_{opt}



Eyelid Reconstruction

Integrate into a face and eyeball fitting result [Wen et al. 2016]



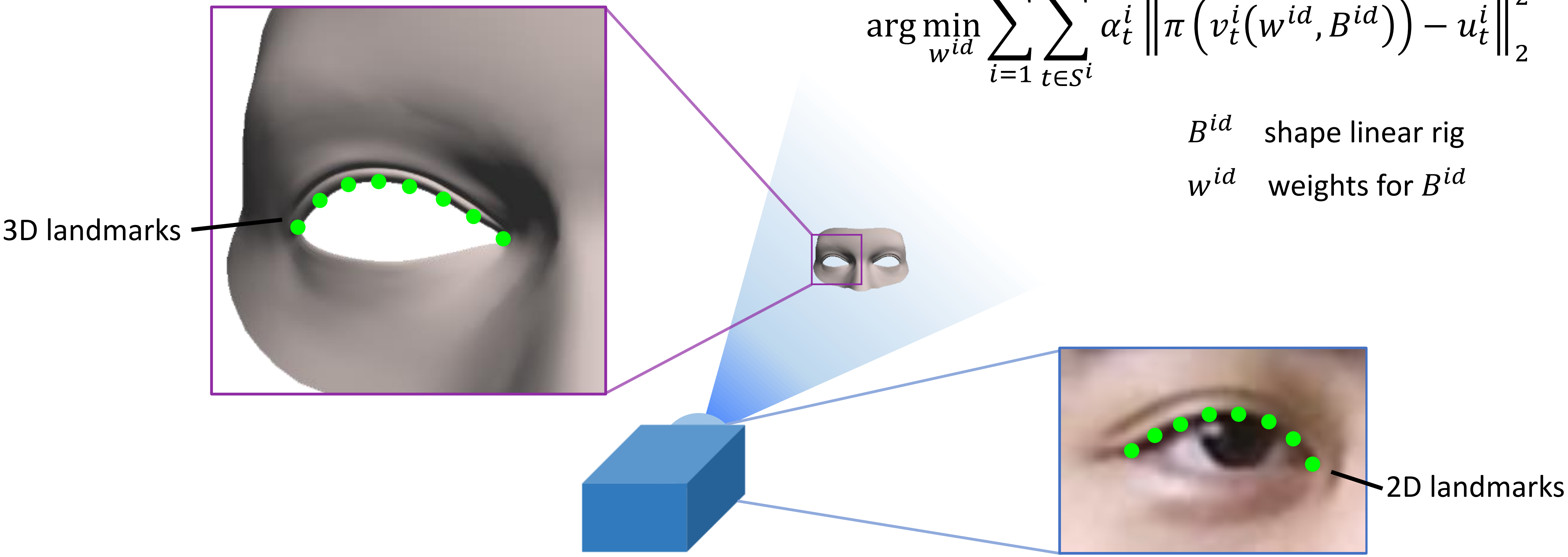
Shape Reconstruction

Solve for the optimal w^{id}

$$\arg\min_{w^{id}} \sum_{i=1}^{4} \sum_{t \in S^i} \alpha_t^i \left\| \pi\left(v_t^i(w^{id}, B^{id})\right) - u_t^i \right\|_2^2$$

B^{id} shape linear rig

w^{id} weights for B^{id}



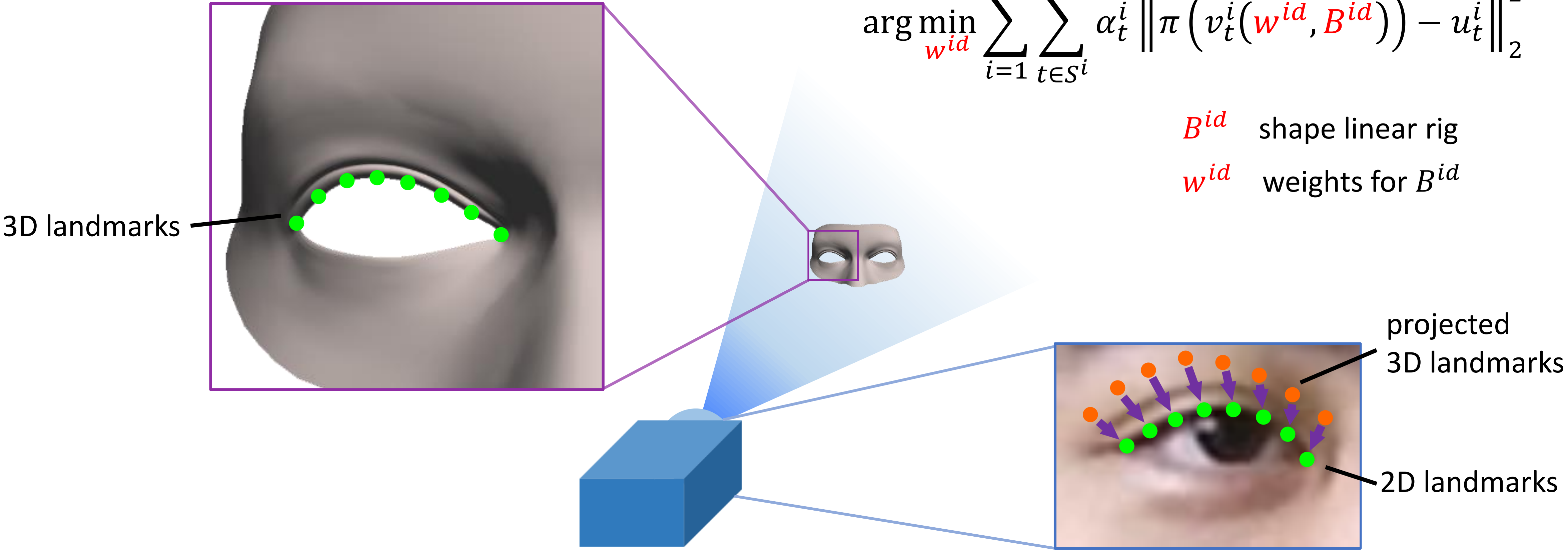
Shape Reconstruction

Solve for the optimal w^{id}

$$\arg\min_{w^{id}} \sum_{i=1}^{4} \sum_{t \in S^i} \alpha_t^i \left\| \pi\left(v_t^i(w^{id}, B^{id})\right) - u_t^i \right\|_2^2$$

B^{id} shape linear rig

w^{id} weights for B^{id}



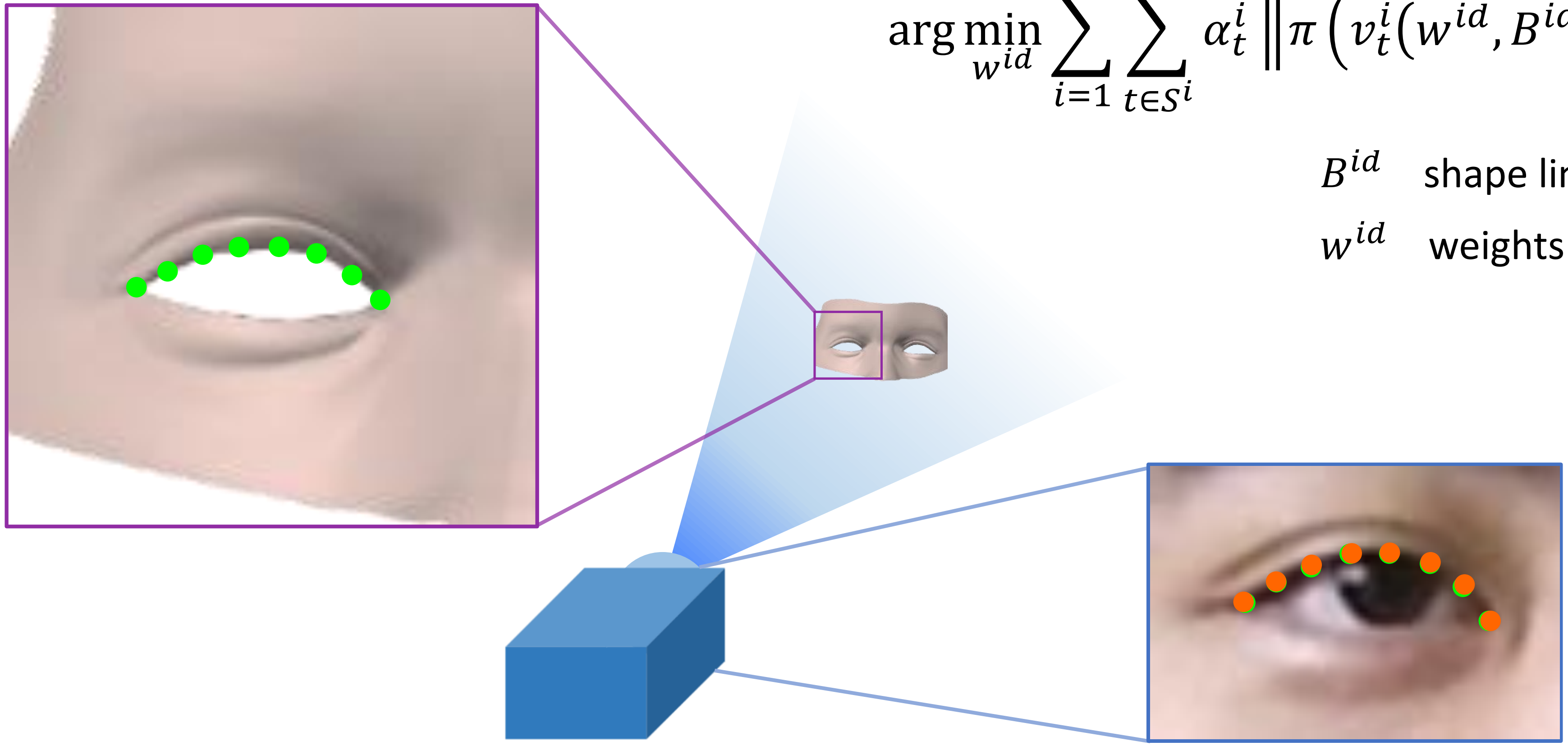
Shape Reconstruction

Solve for the optimal w^{id}

$$\arg\min_{w^{id}} \sum_{i=1}^{4} \sum_{t \in S^i} \alpha_t^i \left\| \pi\left(v_t^i(w^{id}, B^{id})\right) - u_t^i \right\|_2^2$$

B^{id} shape linear rig

w^{id} weights for B^{id}



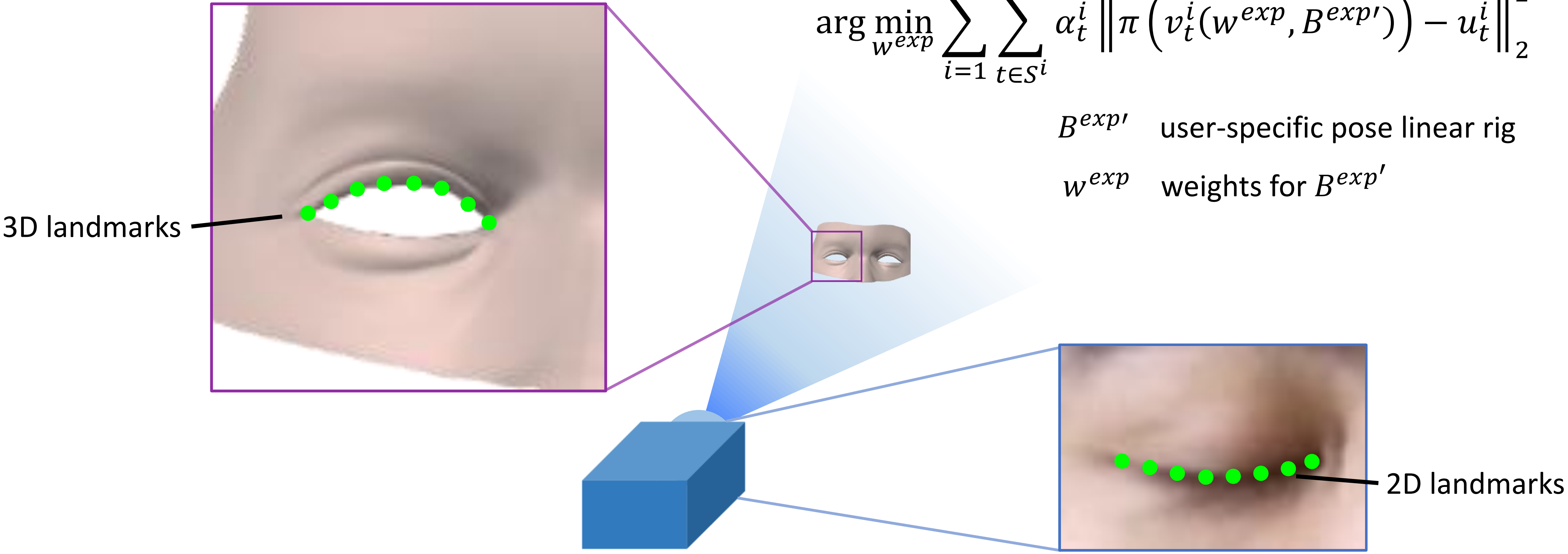
Pose Reconstruction

Solve for the optimal w^{exp}

$$\arg\min_{w^{exp}} \sum_{i=1}^{4} \sum_{t \in S^i} \alpha_t^i \left\| \pi\left(v_t^i(w^{exp}, B^{exp\prime})\right) - u_t^i \right\|_2^2$$

$B^{exp\prime}$ user-specific pose linear rig

w^{exp} weights for $B^{exp\prime}$



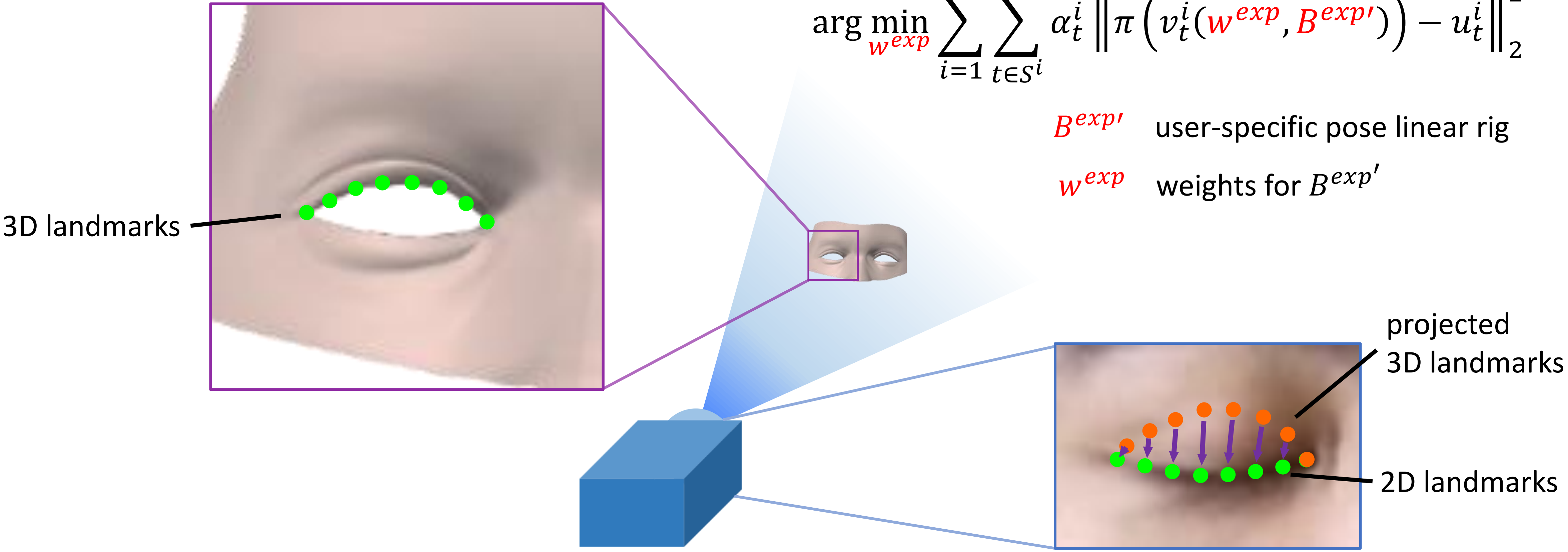
Pose Reconstruction

Solve for the optimal w^{exp}

$$\arg\min_{w^{exp}} \sum_{i=1}^{4} \sum_{t \in S^i} \alpha_t^i \left\| \pi\left(v_t^i(w^{exp}, B^{exp\prime}) \right) - u_t^i \right\|_2^2$$

$B^{exp\prime}$ user-specific pose linear rig

w^{exp} weights for $B^{exp\prime}$



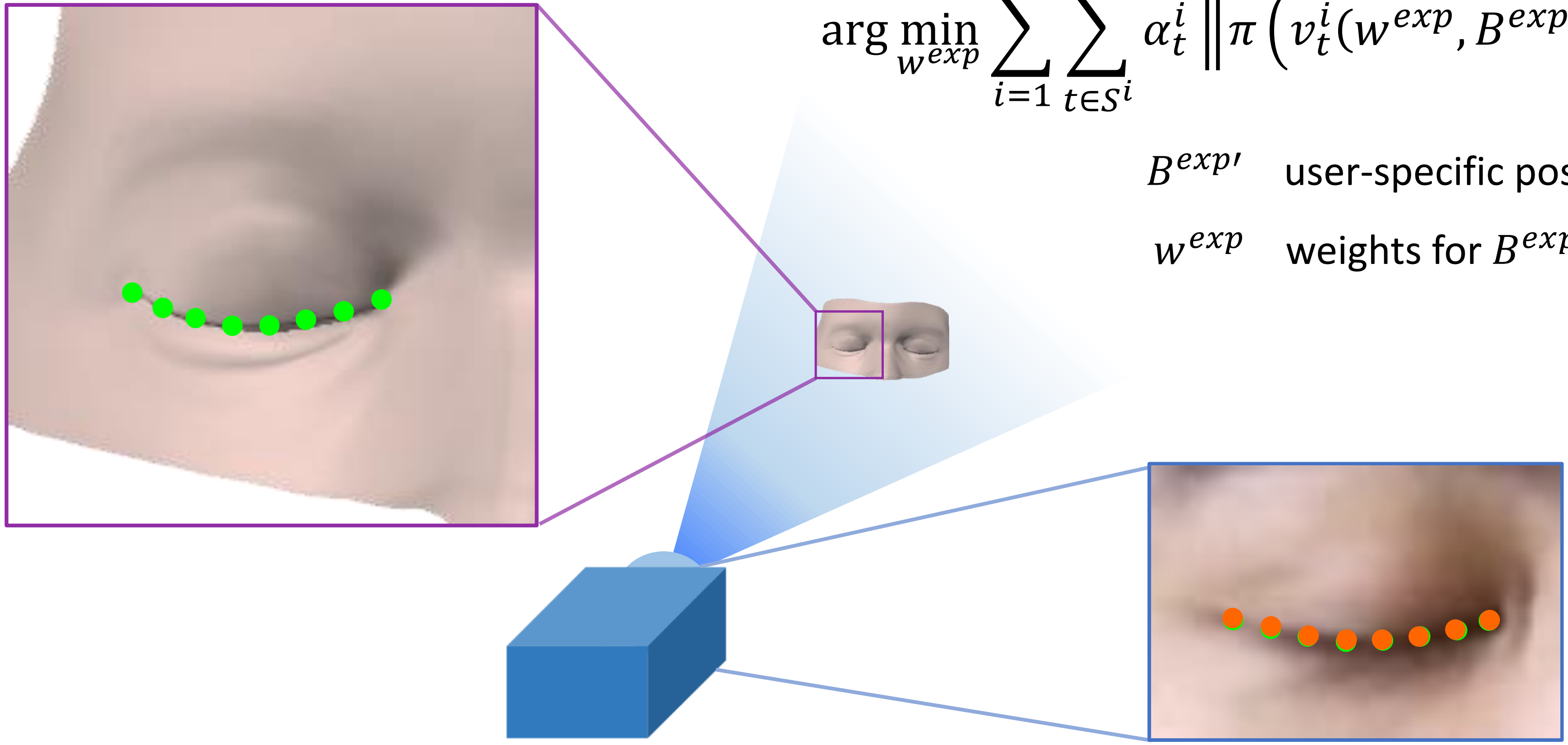
Pose Reconstruction

Solve for the optimal w^{exp}

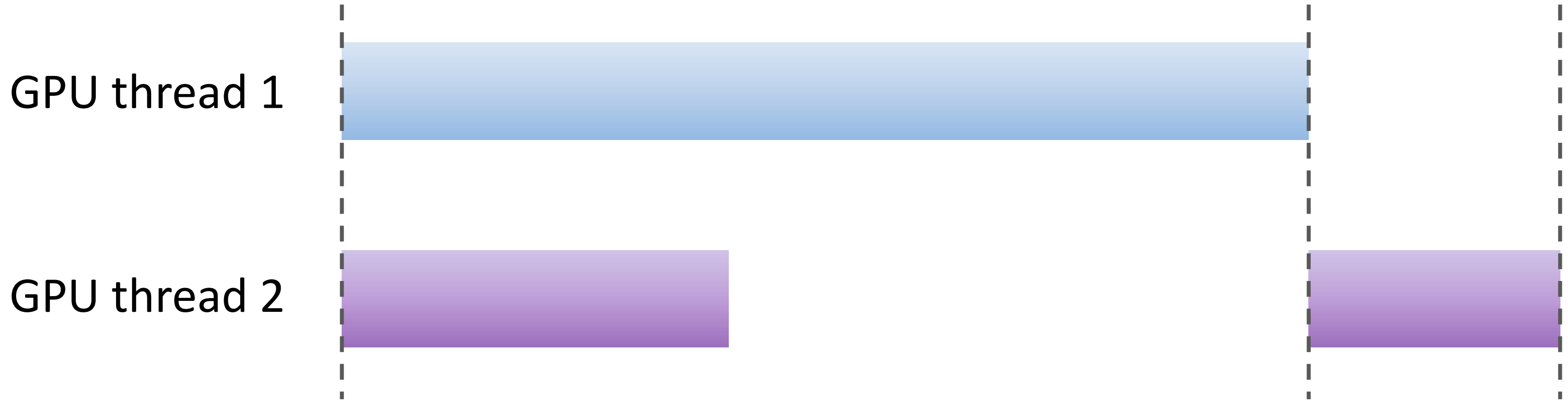
$$\arg\min_{w^{exp}} \sum_{i=1}^{4} \sum_{t \in S^i} \alpha_t^i \left\| \pi\left(v_t^i(w^{exp}, B^{exp\prime})\right) - u_t^i \right\|_2^2$$

$B^{exp\prime}$ user-specific pose linear rig

w^{exp} weights for $B^{exp\prime}$

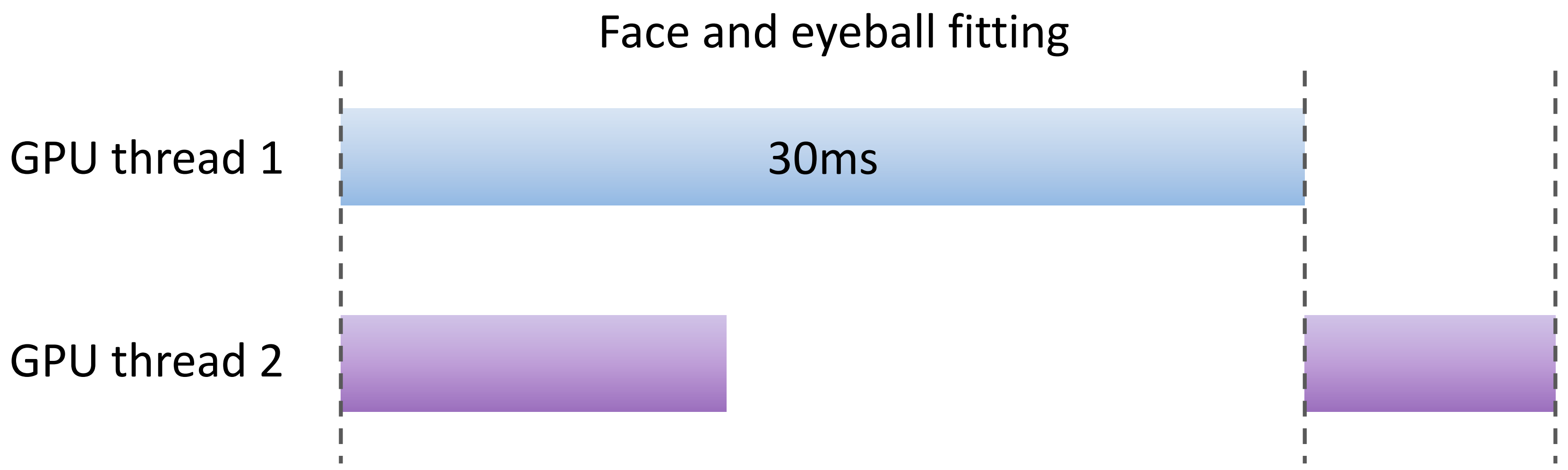


Performance



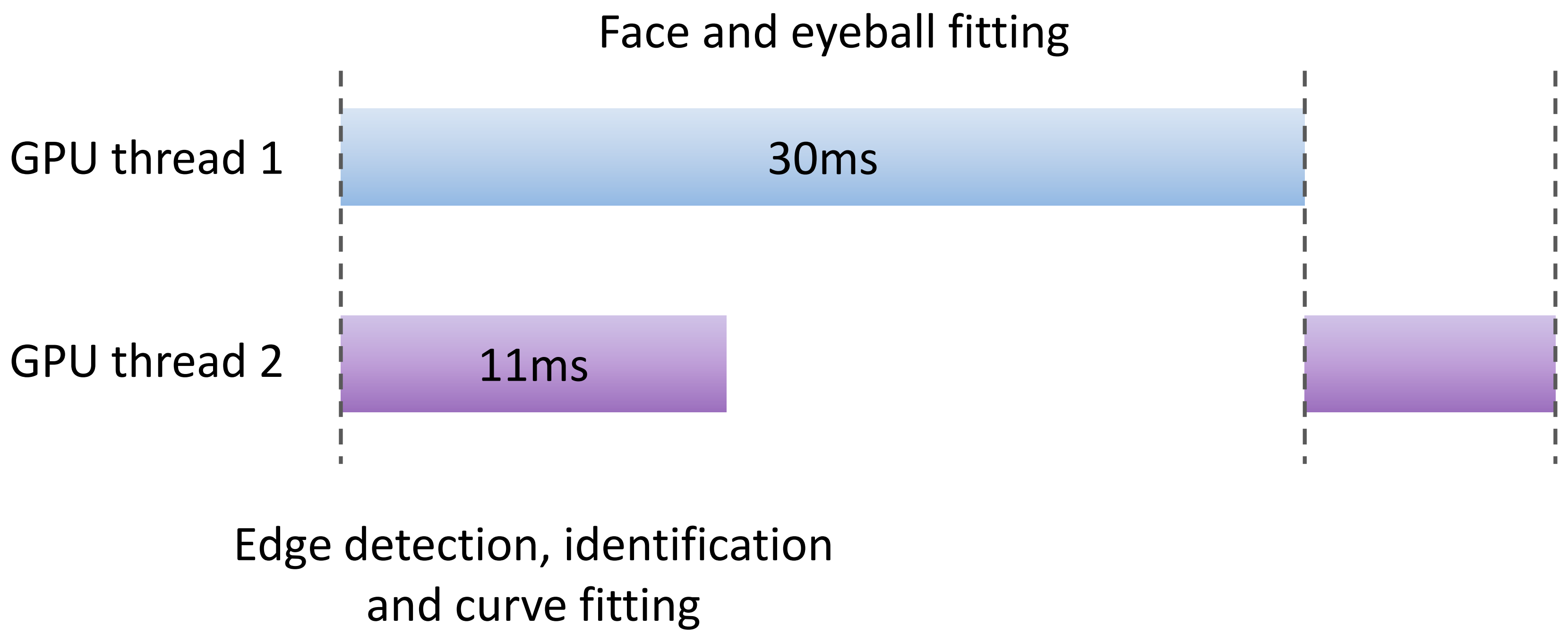
* Test Environment: 3.6GHz eight-core CPU, 16G RAM, NVIDIA Geforce GTX 980

Performance



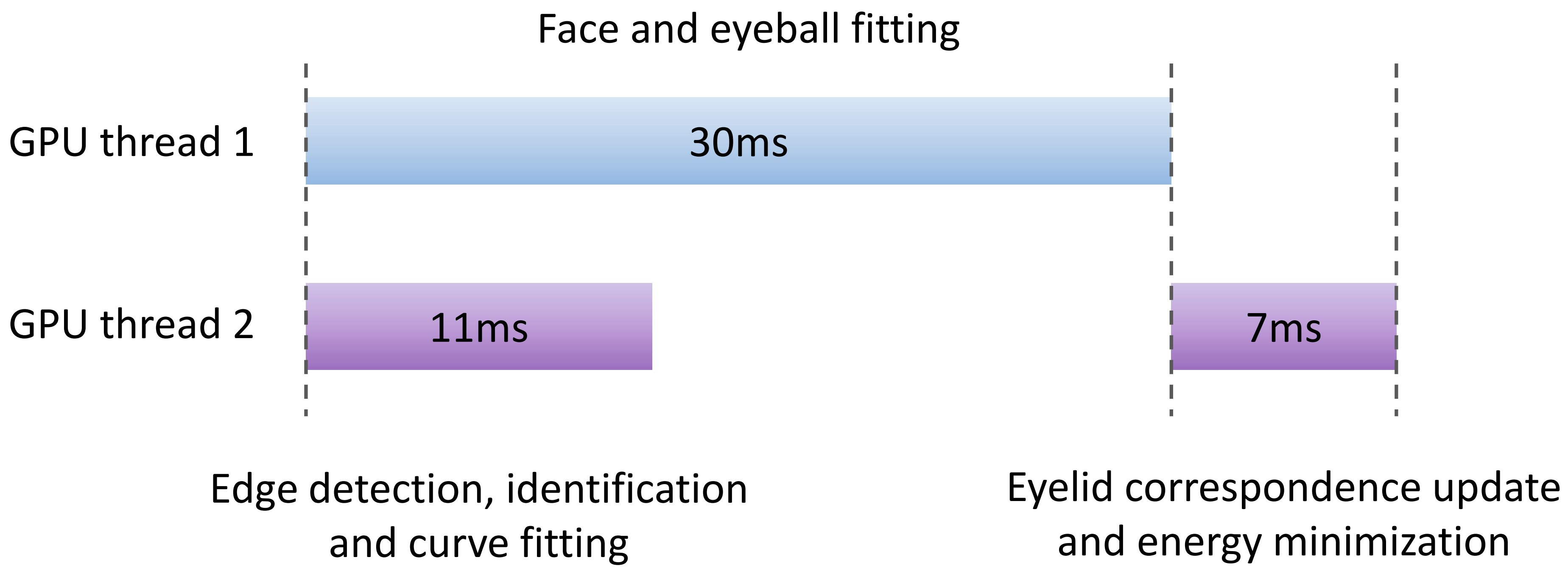
* Test Environment: 3.6GHz eight-core CPU, 16G RAM, NVIDIA Geforce GTX 980

Performance



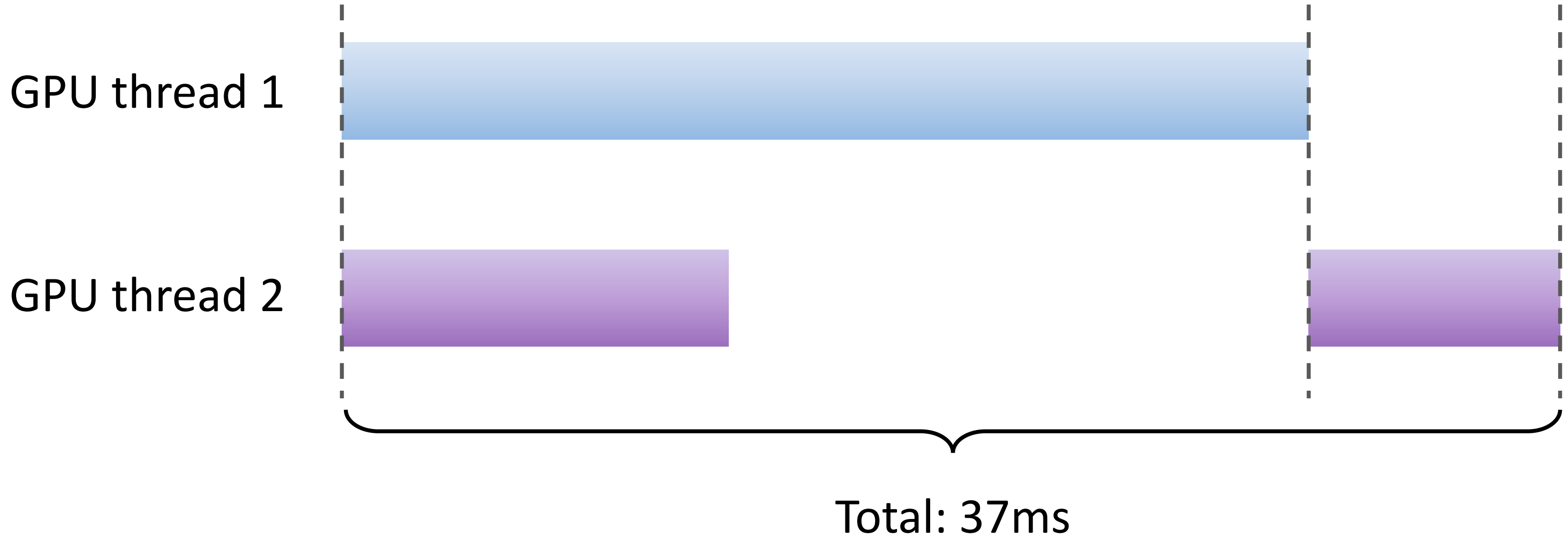
* Test Environment: 3.6GHz eight-core CPU, 16G RAM, NVIDIA Geforce GTX 980

Performance



* Test Environment: 3.6GHz eight-core CPU, 16G RAM, NVIDIA Geforce GTX 980

Performance



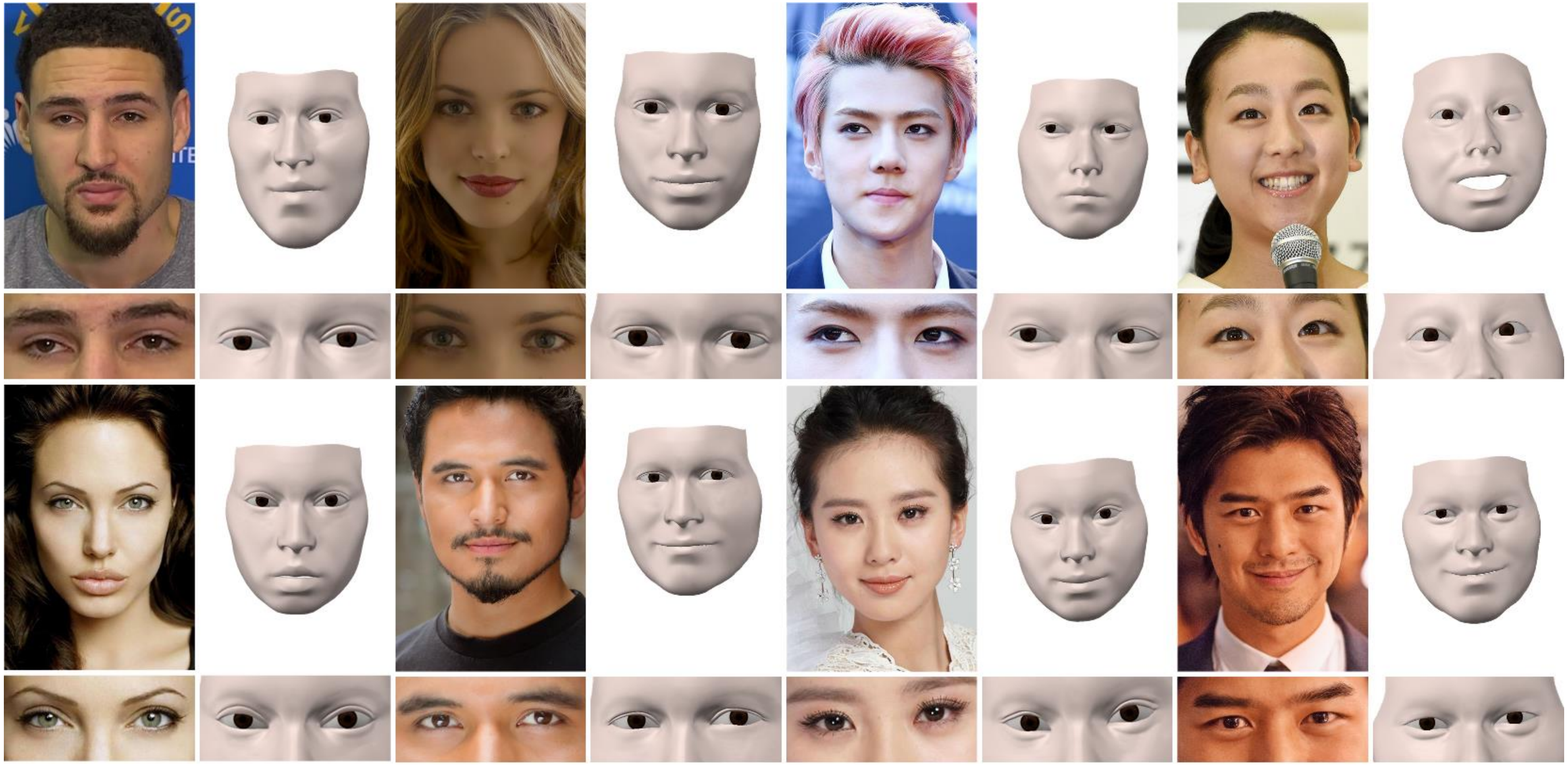
* Test Environment: 3.6GHz eight-core CPU, 16G RAM, NVIDIA Geforce GTX 980

Comparisons

Color Frame　　Our Result　　[Wen et al. 2016]

Live Results

Results of Internet Images



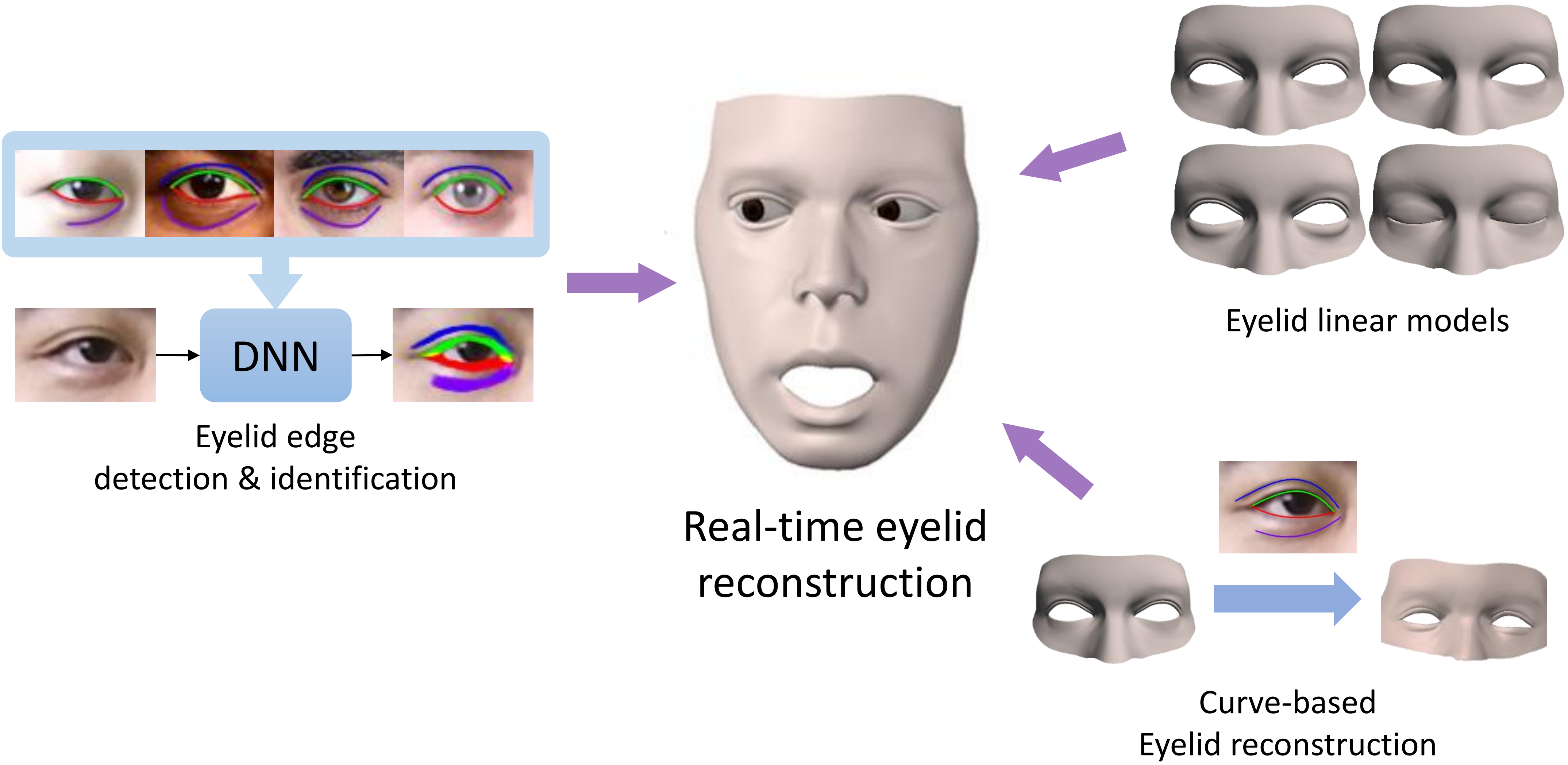
Limitations

Eyelid tracking in challenging lighting conditions

Depth requirement for the tracking system

More shape variations and wrinkle details

Conclusion


DNN
Eyelid edge
detection & identification
Real-time eyelid
reconstruction
Eyelid linear models
Curve-based
Eyelid reconstruction

Thank you

http://feng-xu.com/projects/Realtime3DEyelids/
(Training set and eyelid model are available now)