北京交通大學

TransFill: Reference-guided Image Inpainting by Merging Multiple Color and Spatial Transformations

Yuqian Zhou 1 , Connelly Barnes 2 , Eli Shechtman 2 , Sohrab Amirghodsi 2
1 IFP, UIUC, 2 Adobe Research

CVPR 2021

数字媒体信息处理研究中心
MePro Center of Digital Media Information Processing

Kang Liao

Motivation

Photo 1

Photo 2

😕 **It's hard to capture all perfect faces in one shot**

Motivation

Photo 1

Photo 2

🤔 **Replace the naughty girl from other frames**

Motivation

Photo 1

Photo 2

Everybody looks good in one shot

Reference-guided Image Inpainting

Challenges

Reference-guided image inpainting is challenging due to different views, different lights, different scales, etc.

Related Technologies

Image Inpainting

Video Inpainting

Image Alignment

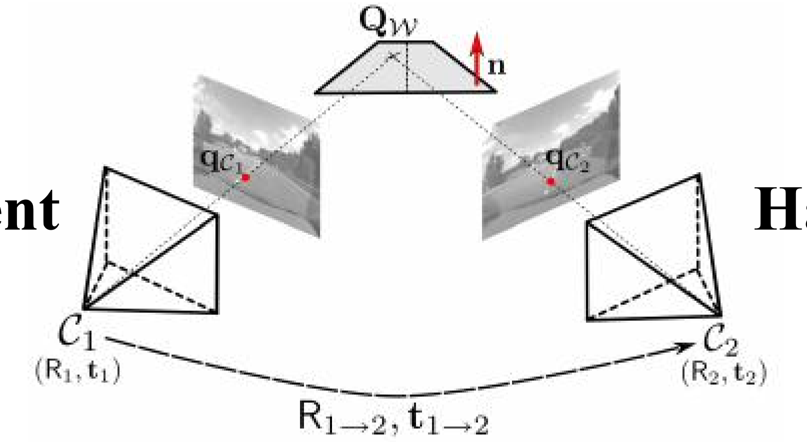
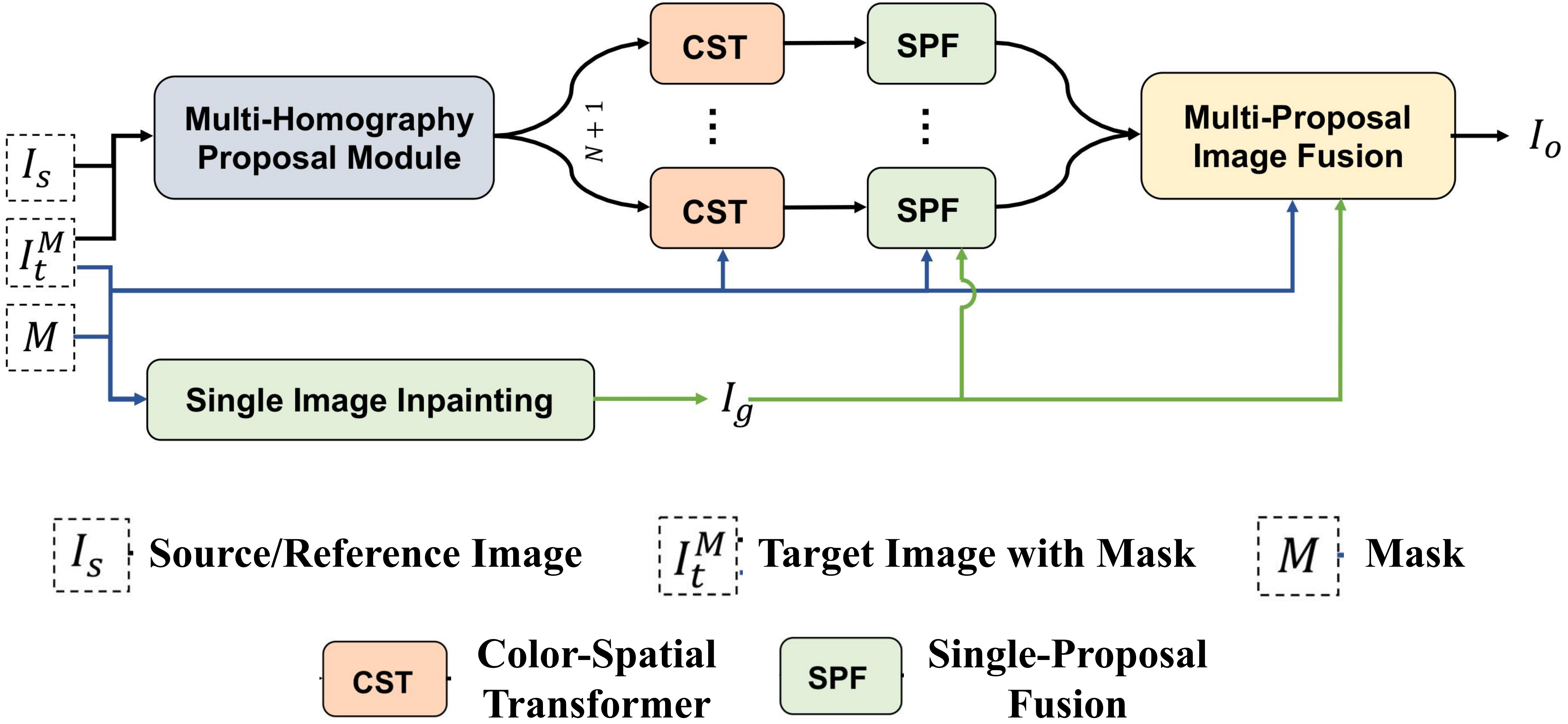
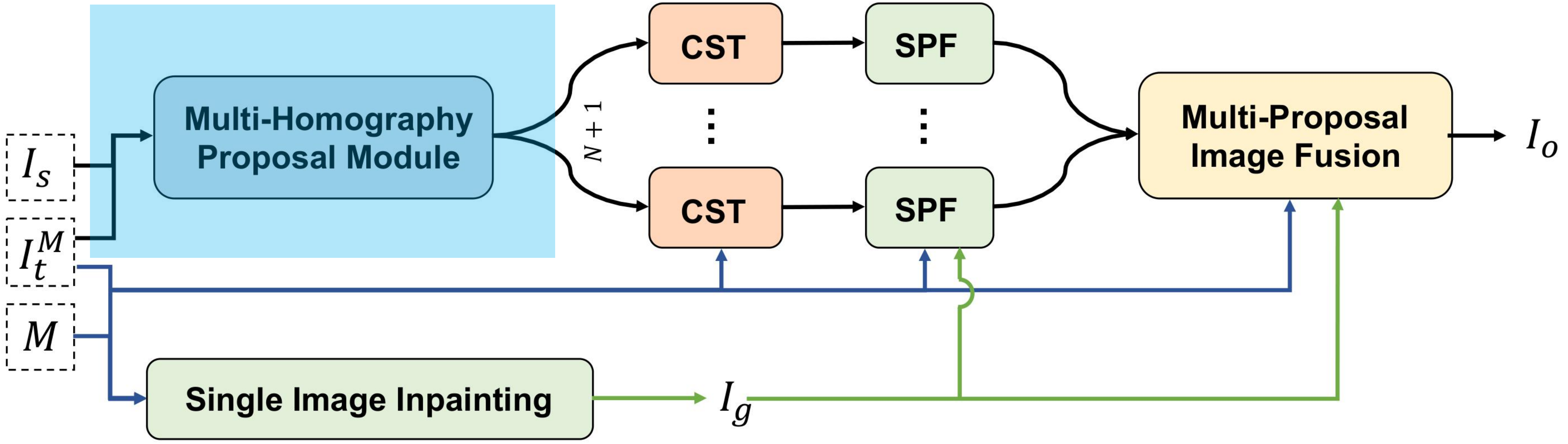


Image Harmonization

Proposed Framework



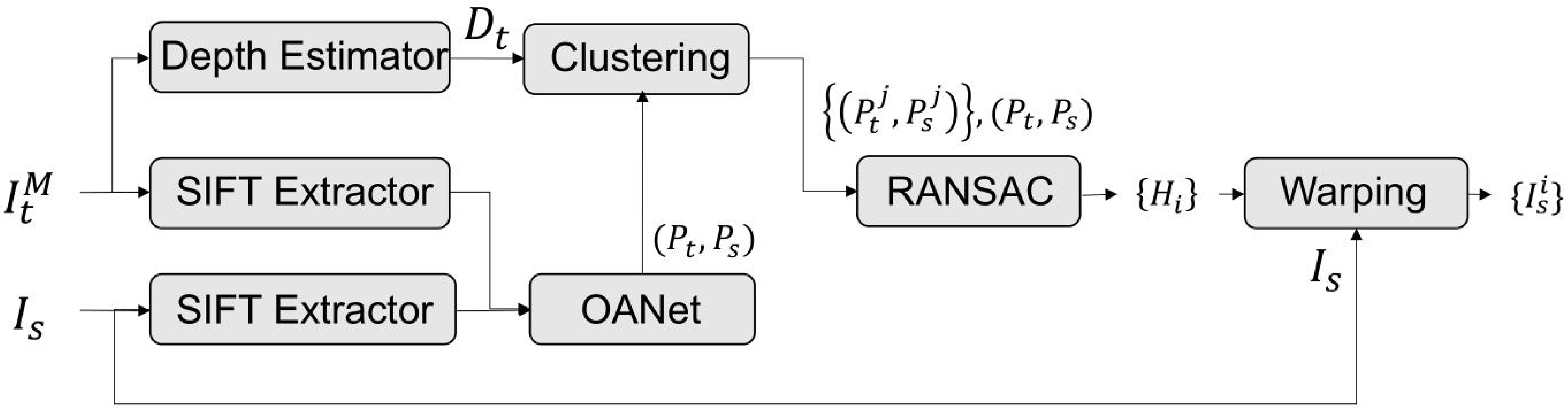
Multi-homography Proposal Module



I_s Source/Reference Image I_t^M Target Image with Mask M Mask

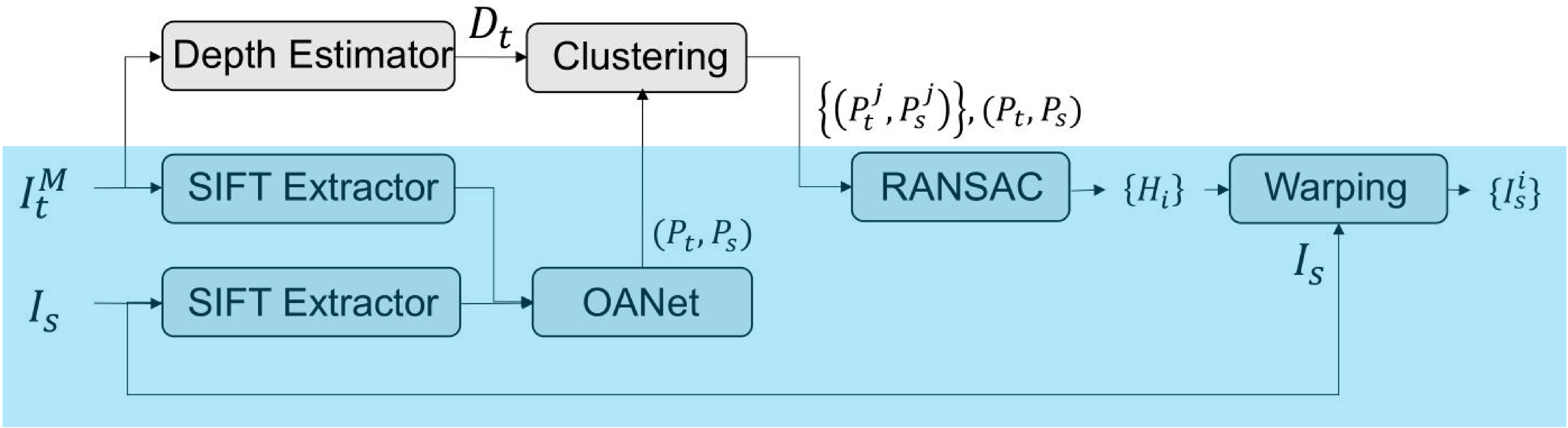
CST Color-Spatial Transformer SPF Single-Proposal Fusion

Multi-homography Proposal Module

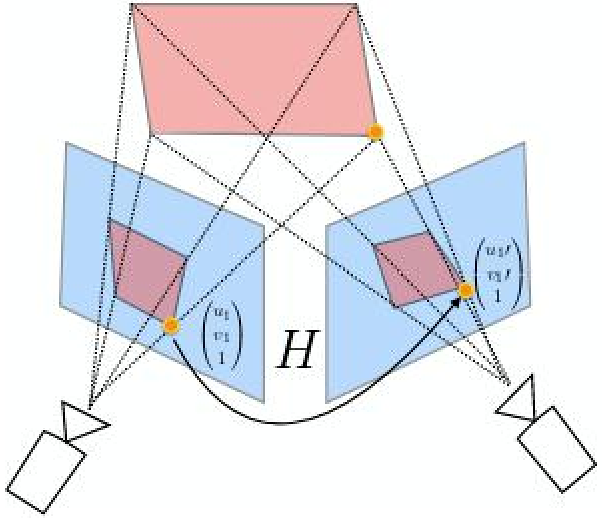


Multi-homography Proposal Module

First, Let's go to a simple case: one homography



Homography Transformation



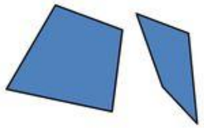
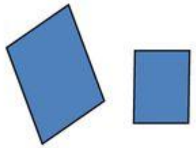
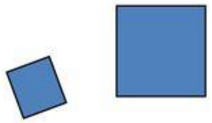
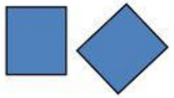
$$H_{4point} = \begin{pmatrix} \Delta u_1 & \Delta v_1 \\ \Delta u_2 & \Delta v_2 \\ \Delta u_3 & \Delta v_3 \\ \Delta u_4 & \Delta v_4 \end{pmatrix}$$

$\updownarrow$ 1-to-1 mapping

$$H_{matrix} = \begin{pmatrix} H_{11} & H_{12} & H_{13} \\ H_{21} & H_{22} & H_{23} \\ H_{31} & H_{32} & H_{33} \end{pmatrix}$$

Homography Matrix

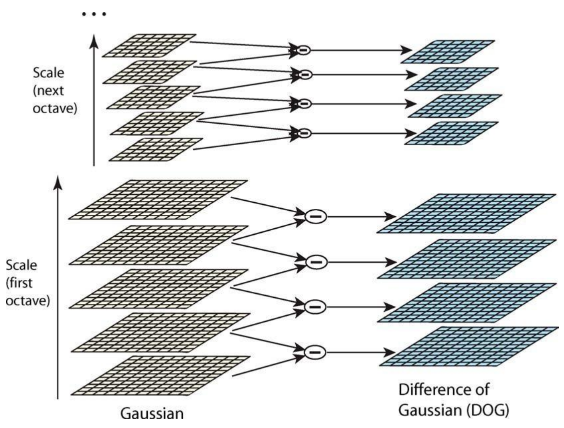
Homography is most general, encompasses other transformations

Projective 8 dof	$\begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix}$		Views of a plane from different viewpoints, any view of a scene from the same viewpoint.
Affine 6 dof	$\begin{bmatrix} a_{11} & a_{12} & t_x \\ a_{21} & a_{22} & t_y \\ 0 & 0 & 1 \end{bmatrix}$		Images of a "far away" object under any rotation
Similarity 4 dof	$\begin{bmatrix} sr_{11} & sr_{12} & t_x \\ sr_{21} & sr_{22} & t_y \\ 0 & 0 & 1 \end{bmatrix}$		Camera looking at an assembly line w/ zoom.
Euclidean 3 dof	$\begin{bmatrix} r_{11} & r_{12} & t_x \\ r_{21} & r_{22} & t_y \\ 0 & 0 & 1 \end{bmatrix}$		Camera looking at an assembly line.

Computer Vision, Robert Pless

Different Transformations

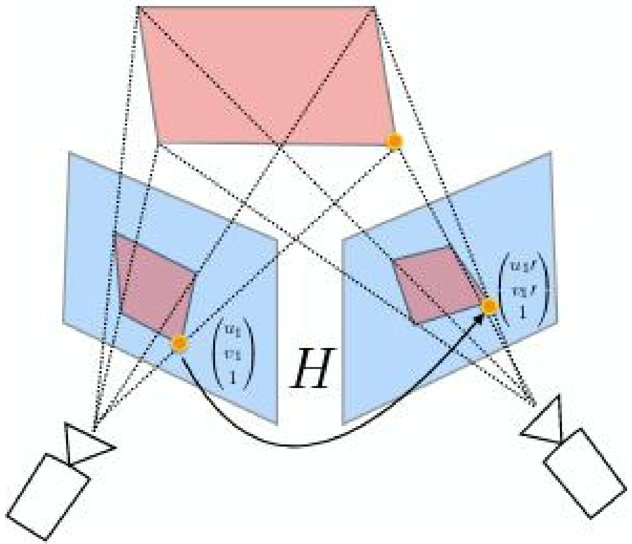
Homography Estimation



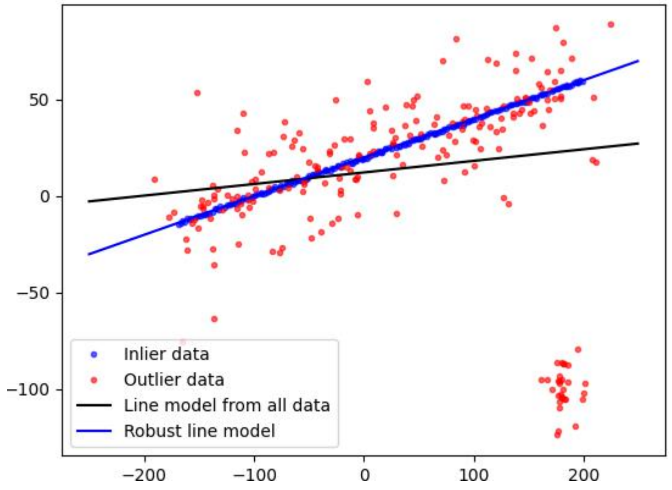
SIFT Detector

Feature Points

Points Matching



Homography



RANSAC: Control Outlier

Image Alignment

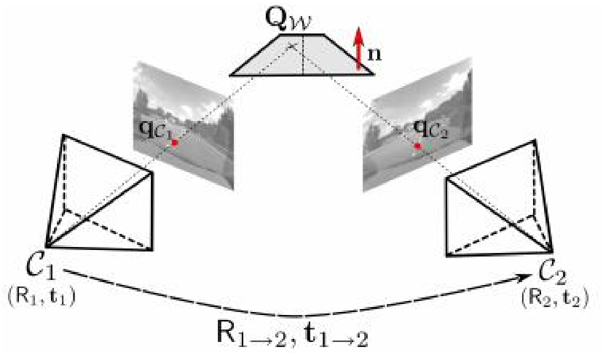


Image Sequences with Different Views

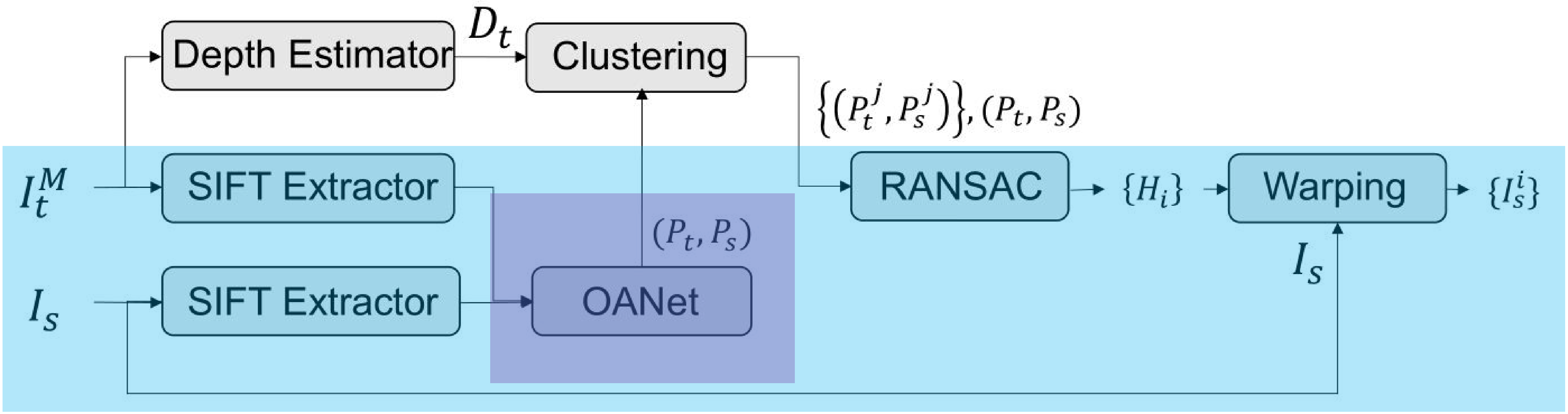
Homography

Warp

View Normalization

Multi-homography Proposal Module

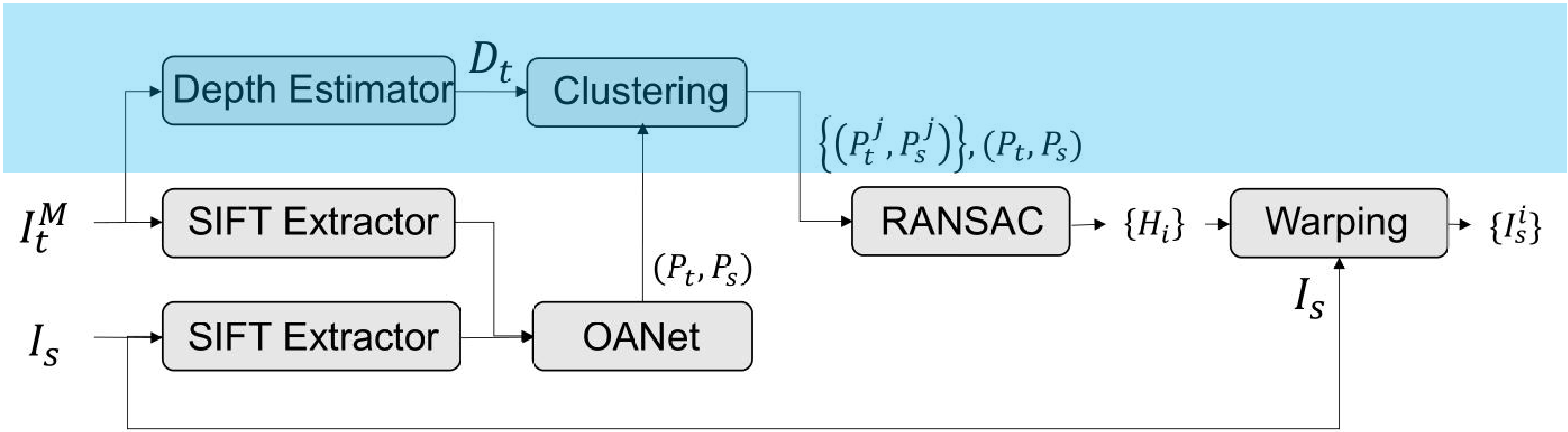
First, Let's go to a simple case: one homography



feed all the extracted feature points and their descriptors into
a pre trained OANet for outlier rejection

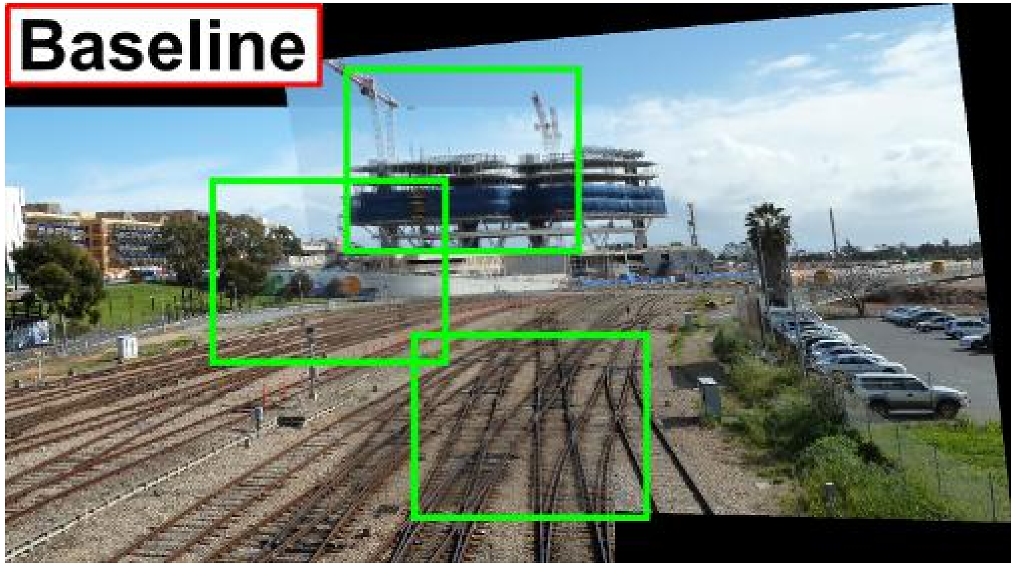
Multi-homography Proposal Module

Then, Why multi-homography



Multi-homography Proposal Module

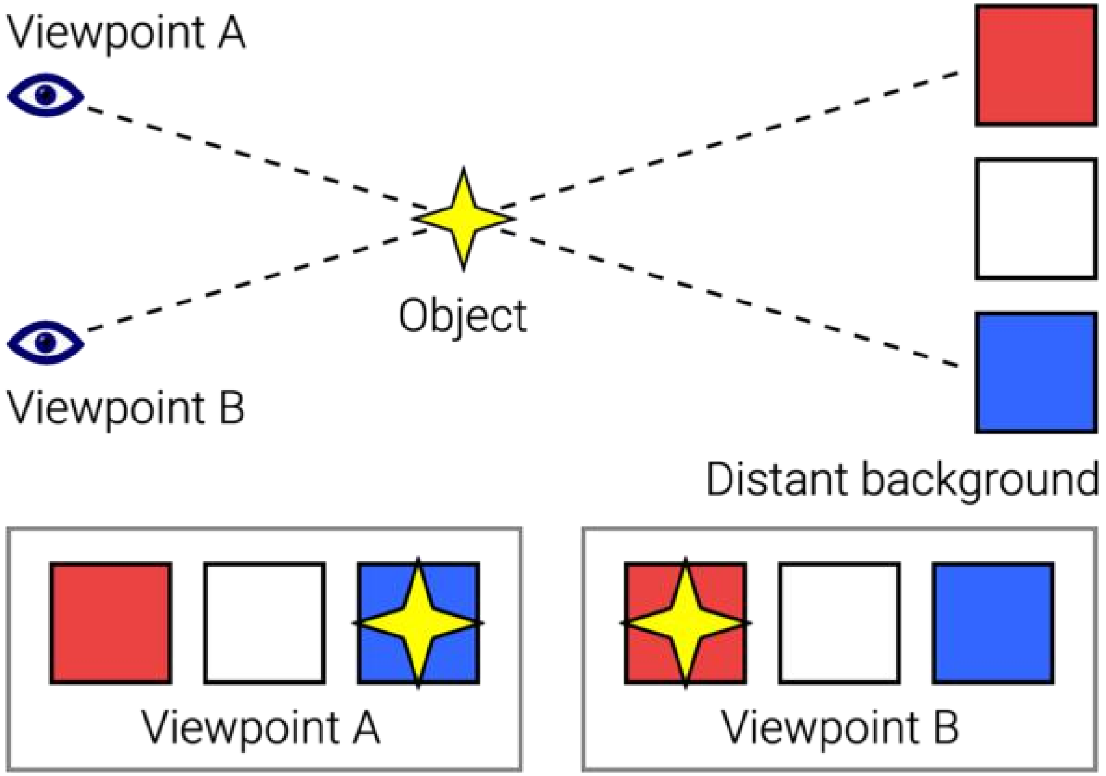
Then, Why multi-homography



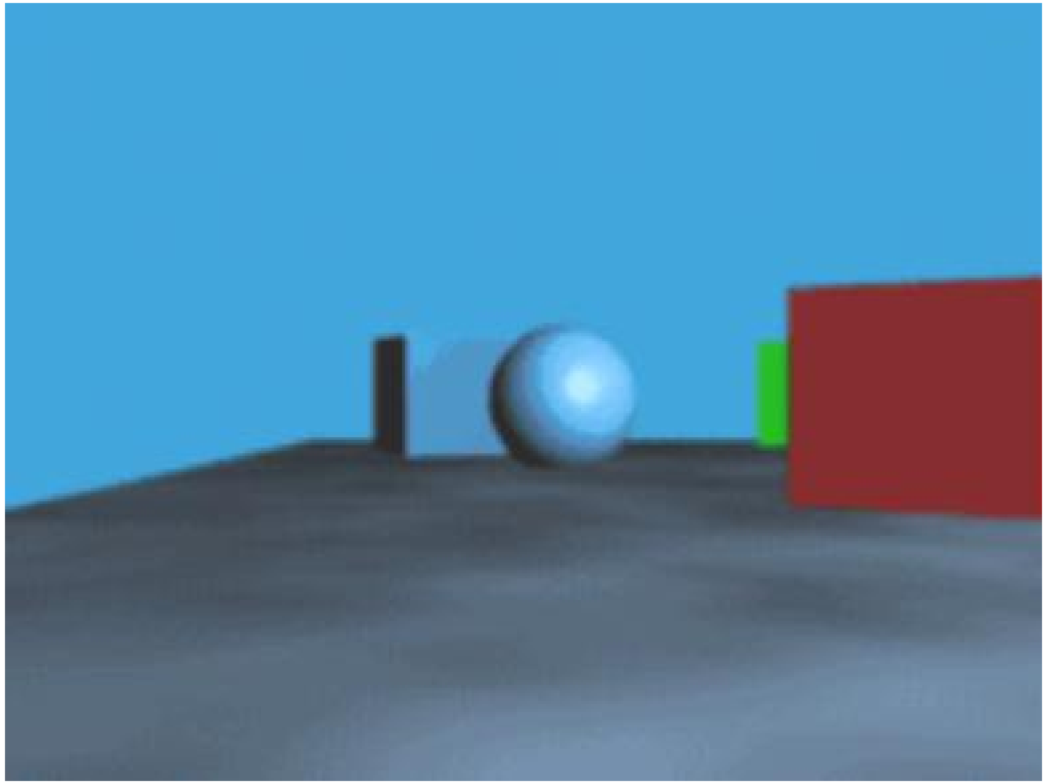
Baseline method with single homography

Misalignments

Parallax



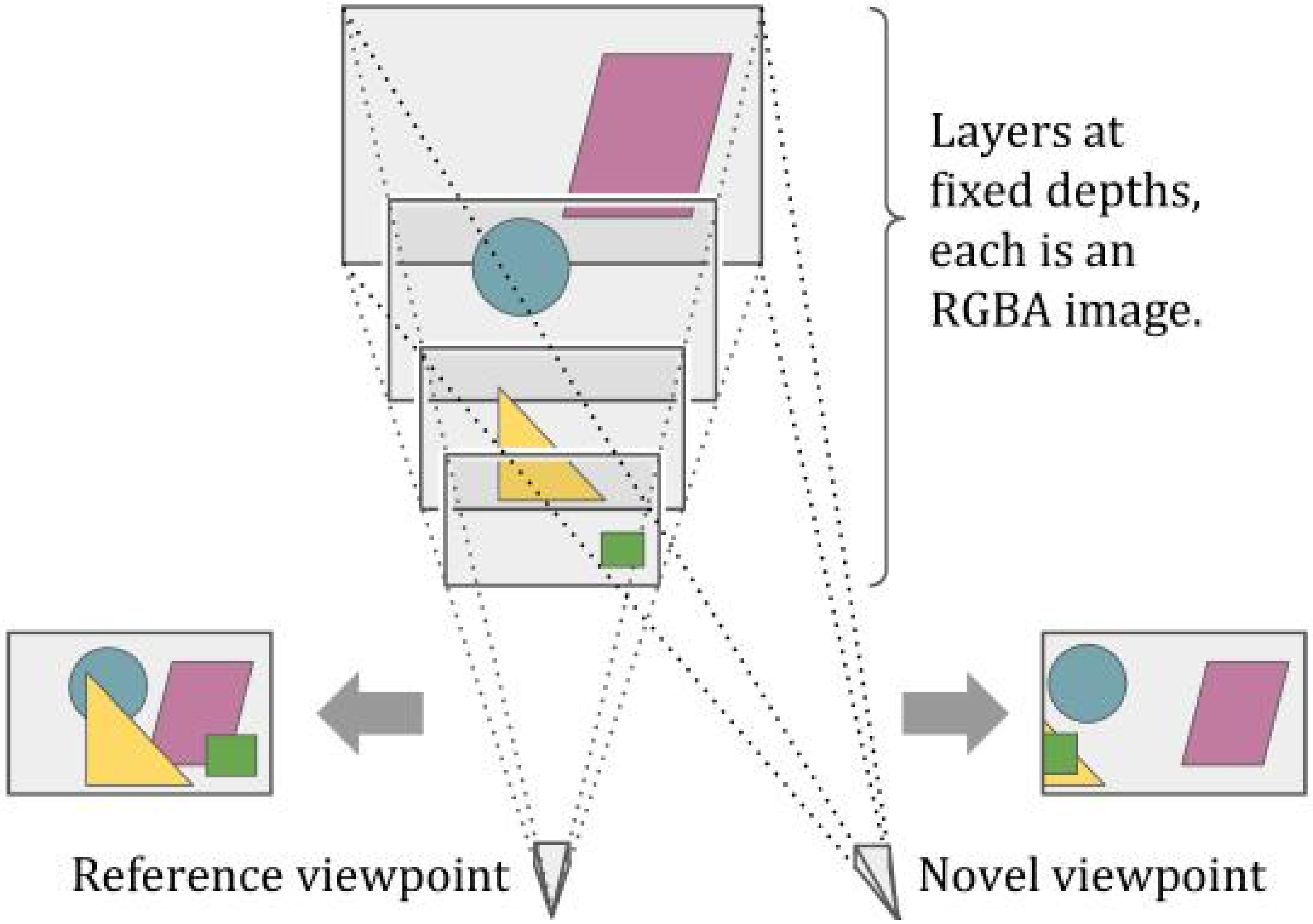
A simplified illustration of the parallax



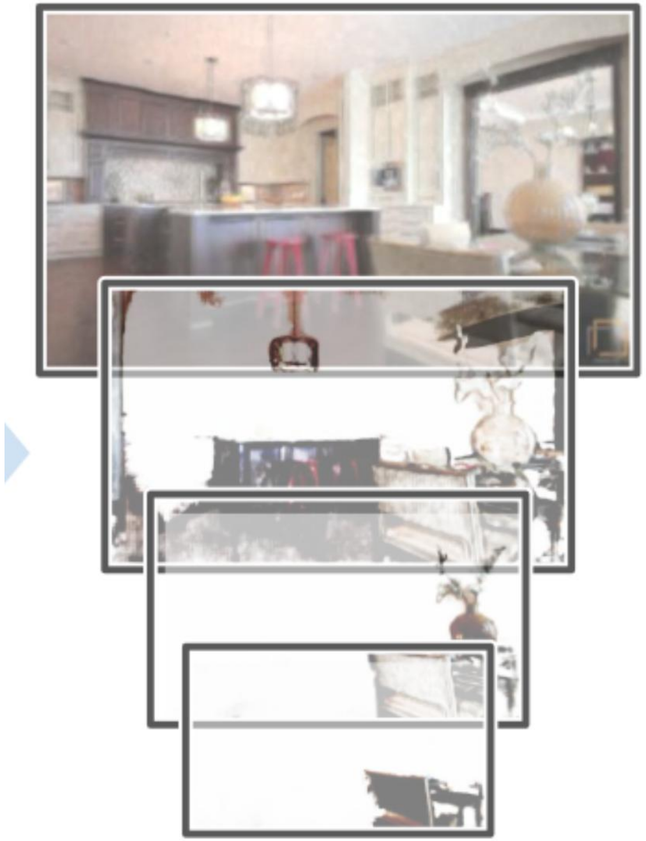
Objects from different depths have different relative motions. Closer objects provide larger parallax

Images from Wikipedia

Multi-homography Proposal Module



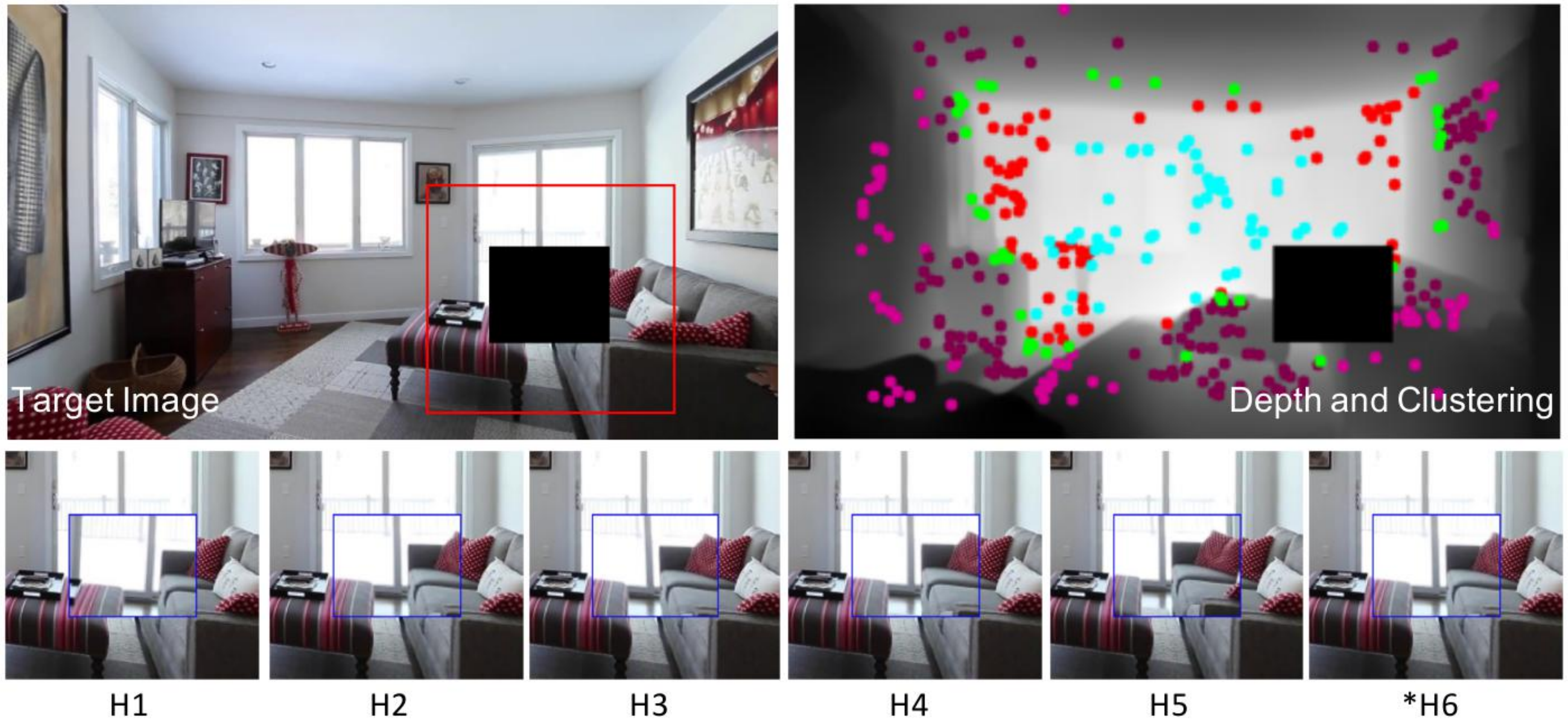
Multiplane Image (MPI)

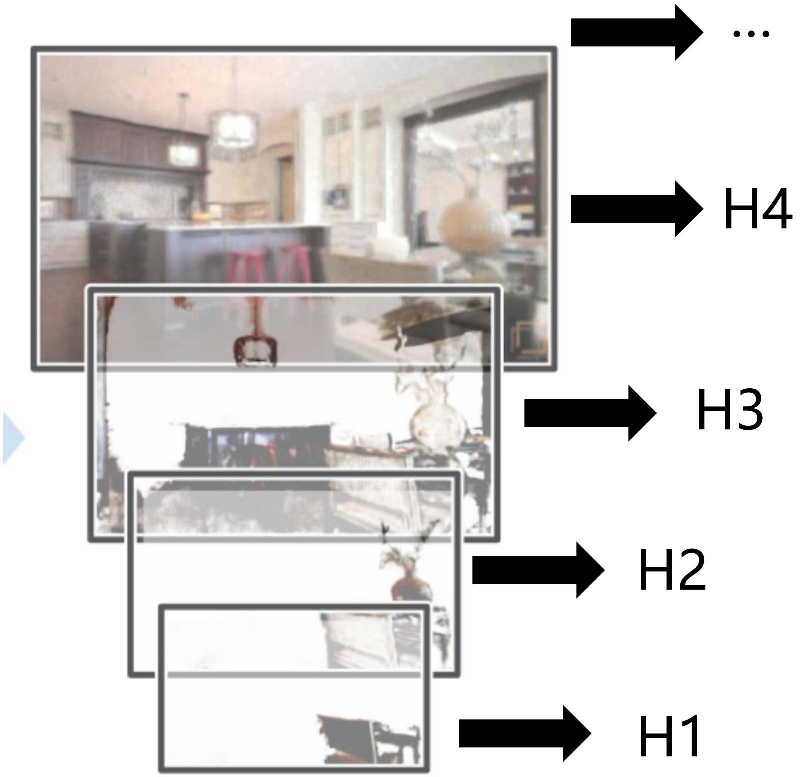


MPI of a scene

Stereo Magnification: Learning View Synthesis using Multiplane Images. *Tinghui Zhou, Richard Tucker, John Flynn, Graham Fyffe, Noah Snavely.* In SIGGRAPH 2018

Multi-homography Proposal Module



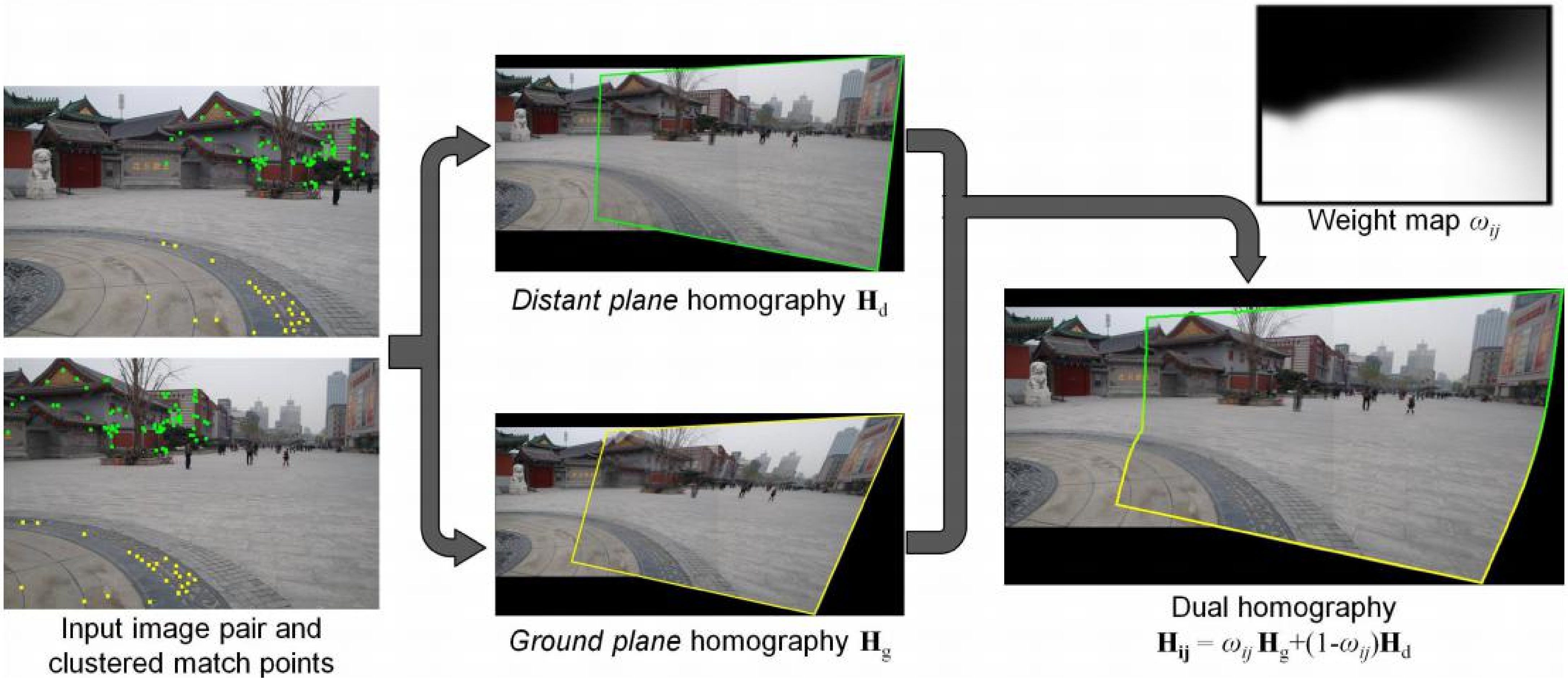


H1-H5: Alignment results from five clustering depths (agglomerative clustering method).
*H6: a homography estimated using all the points

One depth layer for one homography H

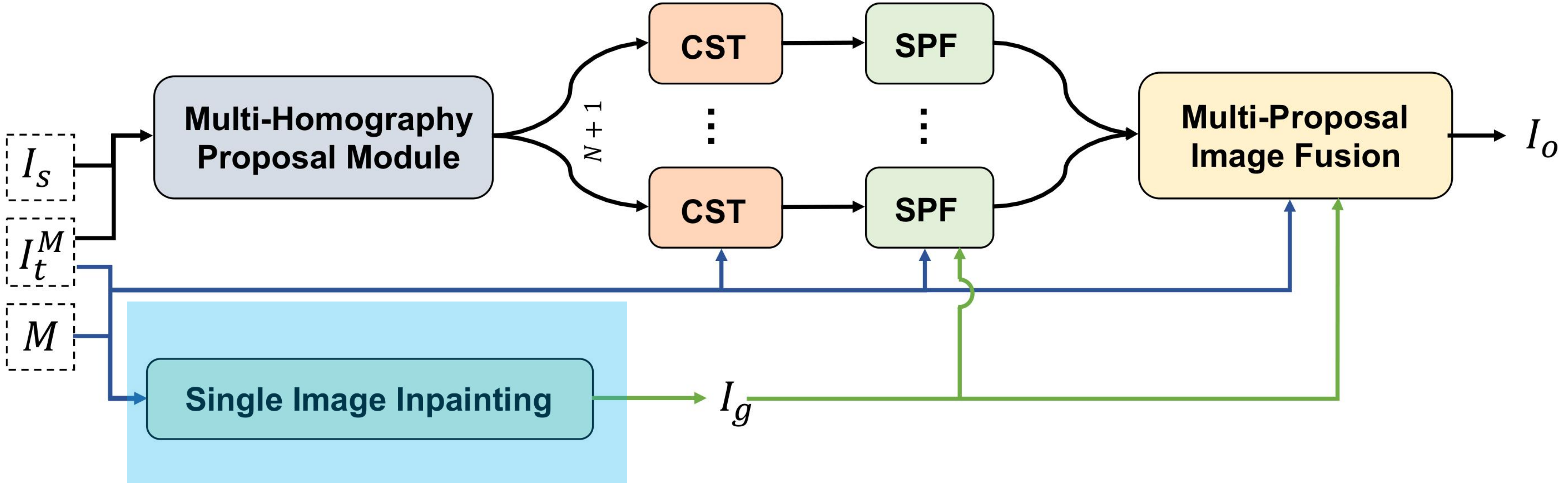
Multi-homography Proposal Module

This strategy is very common in image alignment and image stitching fields



Gao, Junhong, Seon Joo Kim and M. S. Brown. "Constructing image panoramas using dual-homography warping." In *CVPR 2011*.

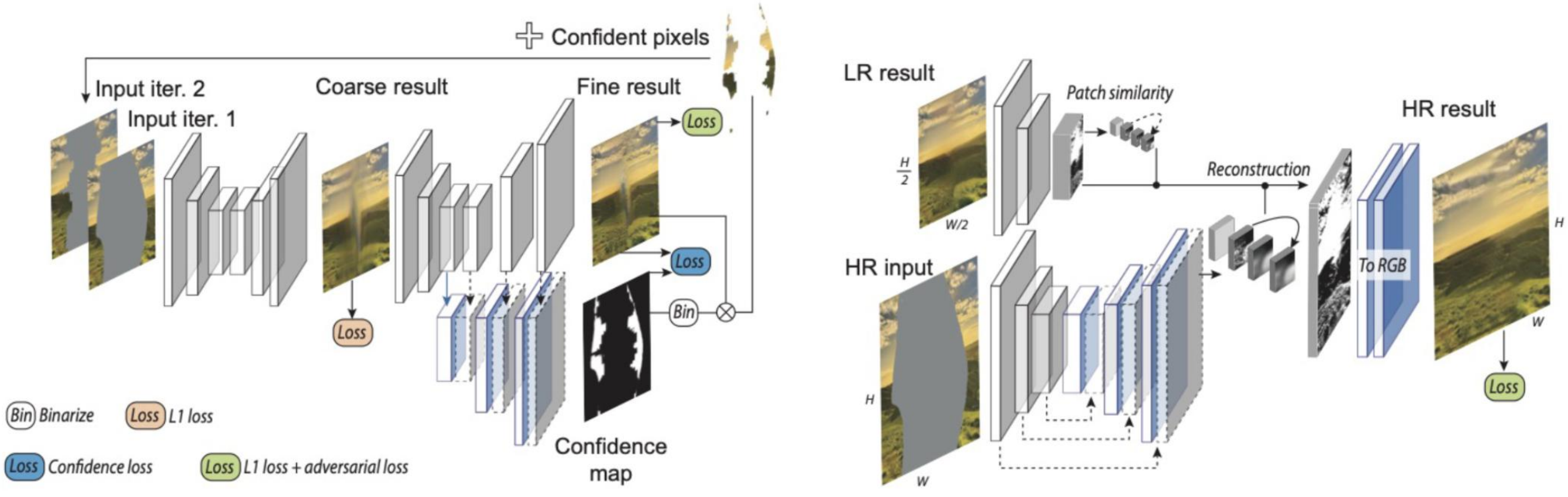
Single Image Inpainting Module



I_s Source/Reference Image I_t^M Target Image with Mask M Mask

CST Color-Spatial Transformer SPF Single-Proposal Fusion

Single Image Inpainting Module



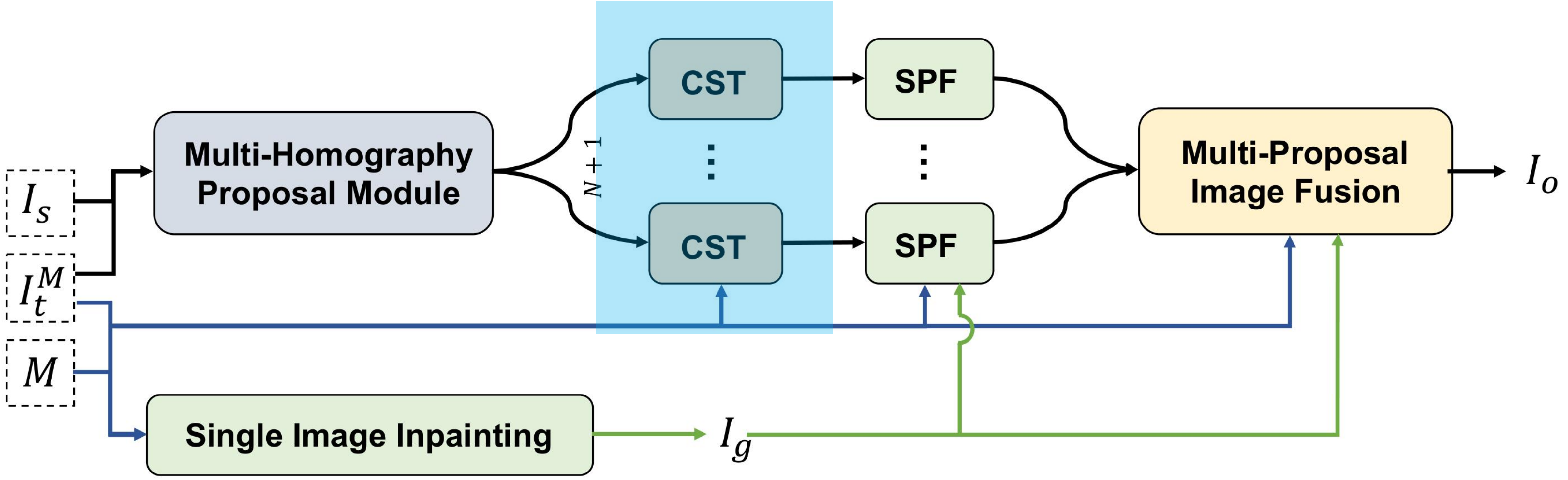
(a) (b)

Fig. 3: The overall structure. (a) Iterative inpainting with confidence feedback. (b) Guided upsampling.

ProFill (ECCV 2020)

Yu Zeng, et al. ProFill: High-Resolution Image Inpainting with Iterative Confidence Feedback and Guided Upsampling, ECCV 2020

Color-Spatial Transformer Module



I_s Source/Reference Image

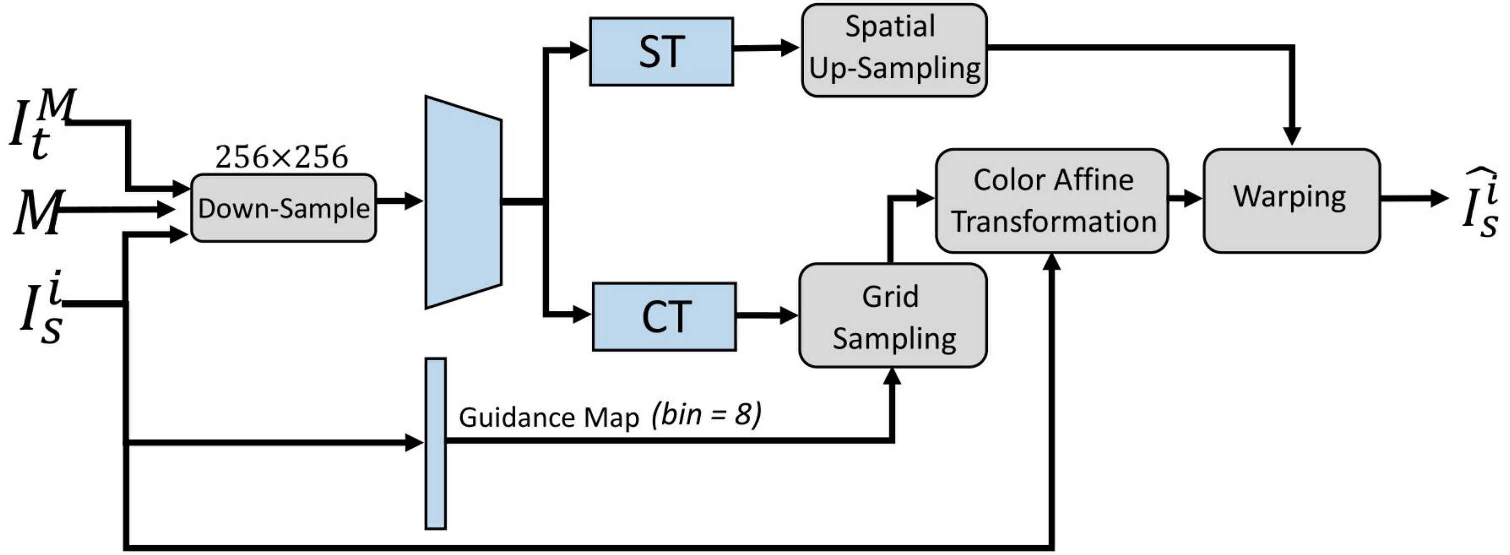
I_t^M Target Image with Mask

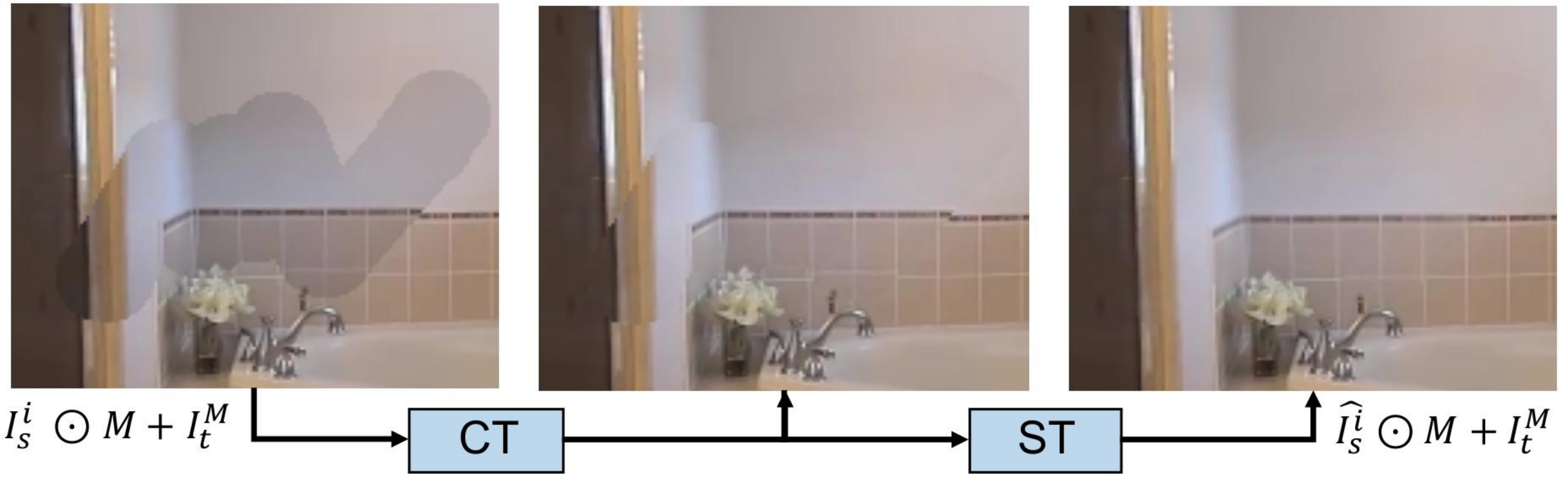
M Mask

CST Color-Spatial Transformer

SPF Single-Proposal Fusion

Color-Spatial Transformer Module





CST

I_t^M **Target Image with Mask**

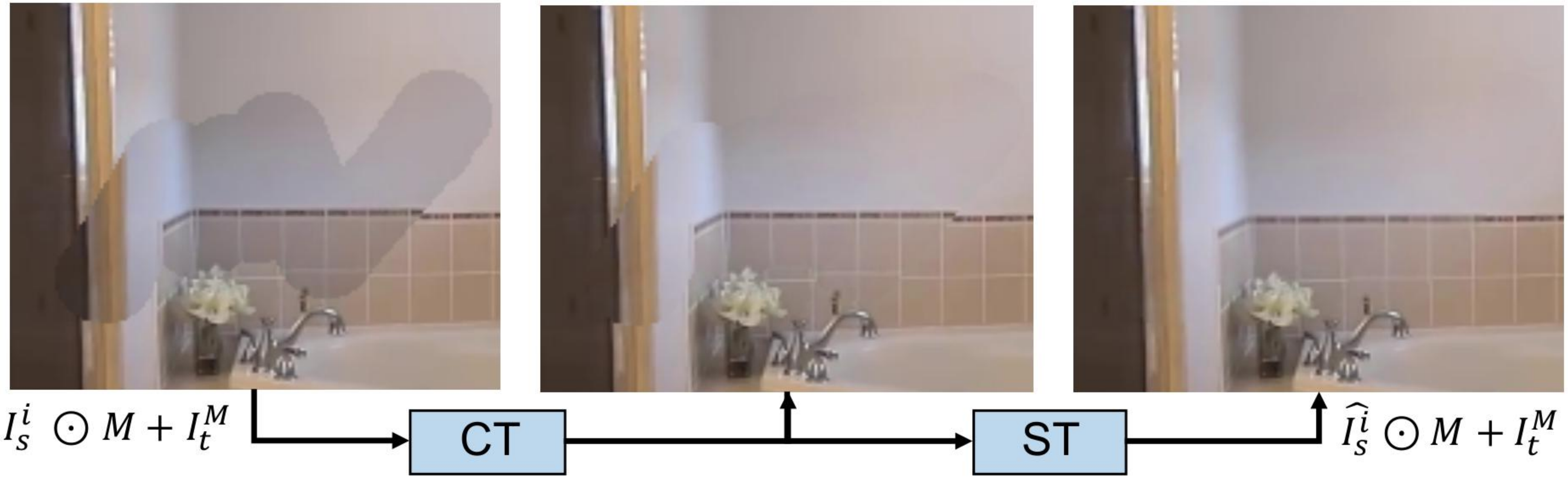
M **Mask**

I_s^i **Aligned source image from i homography**

CT **Color Transformer**

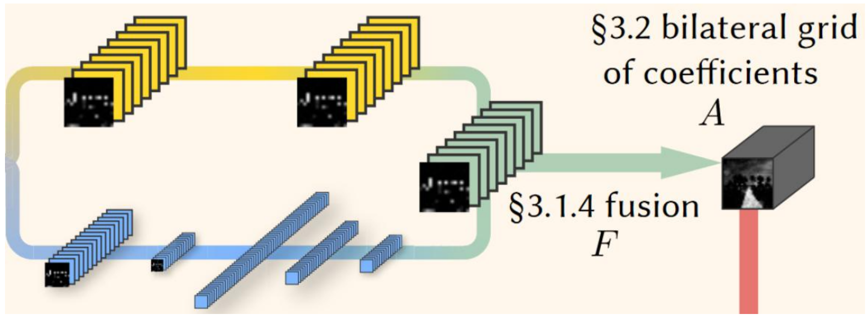
ST **Spatial Transformer**

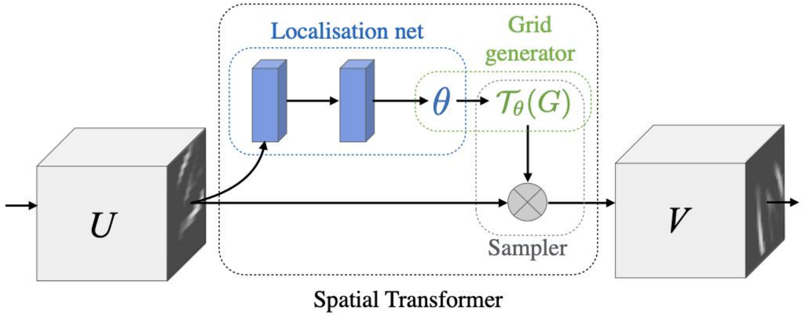
Color-Spatial Transformer Module



Deep Bilateral Filtering
(Siggraph 2017)

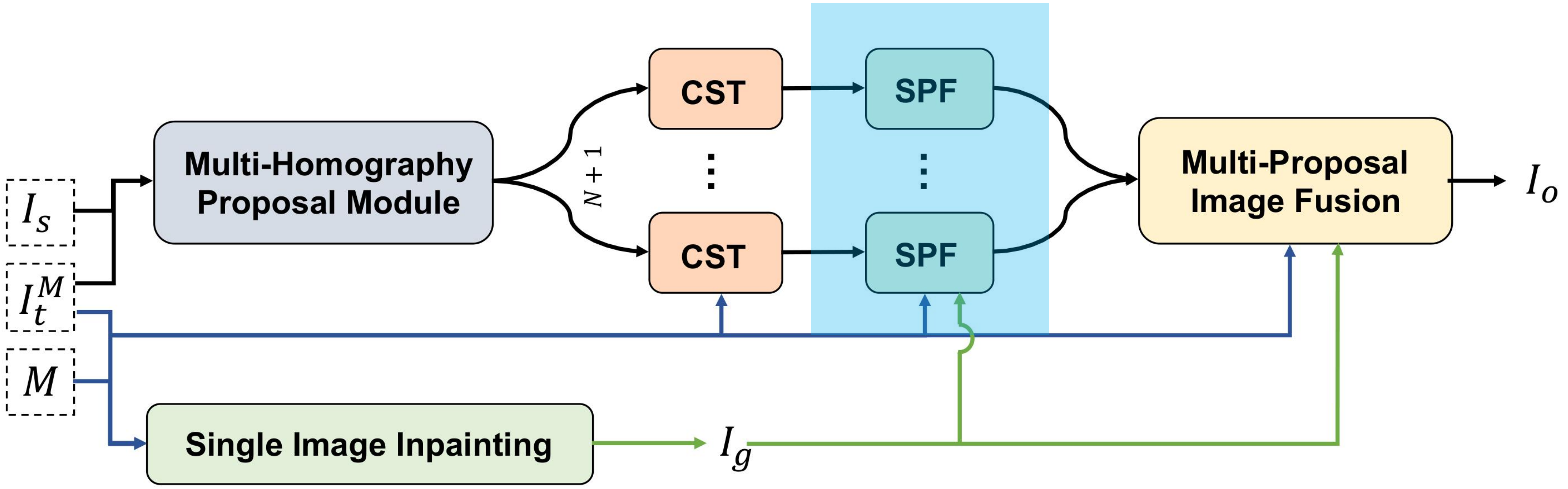
Spatial Transformer Network
(STN, NIPS 2015)





Optical Flow

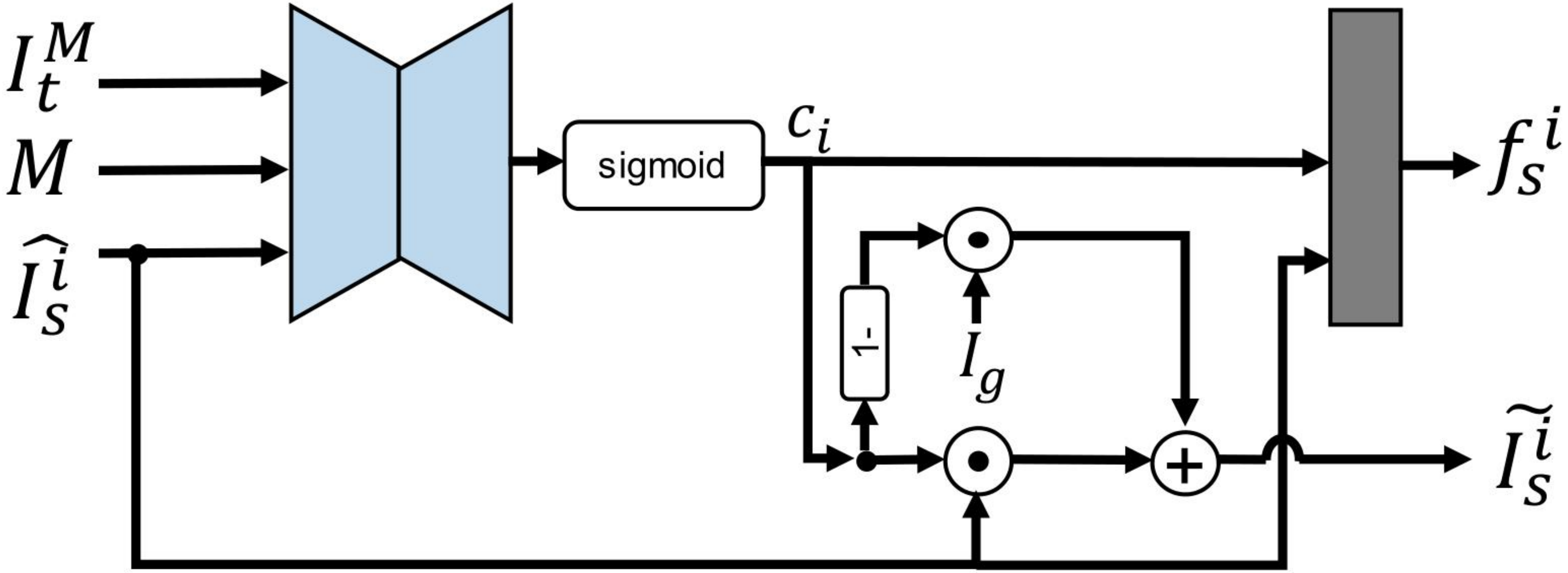
Single-Proposal Fusion Module



I_s Source/Reference Image
I_t^M Target Image with Mask
M Mask

CST Color-Spatial Transformer
SPF Single-Proposal Fusion

Single-Proposal Fusion Module

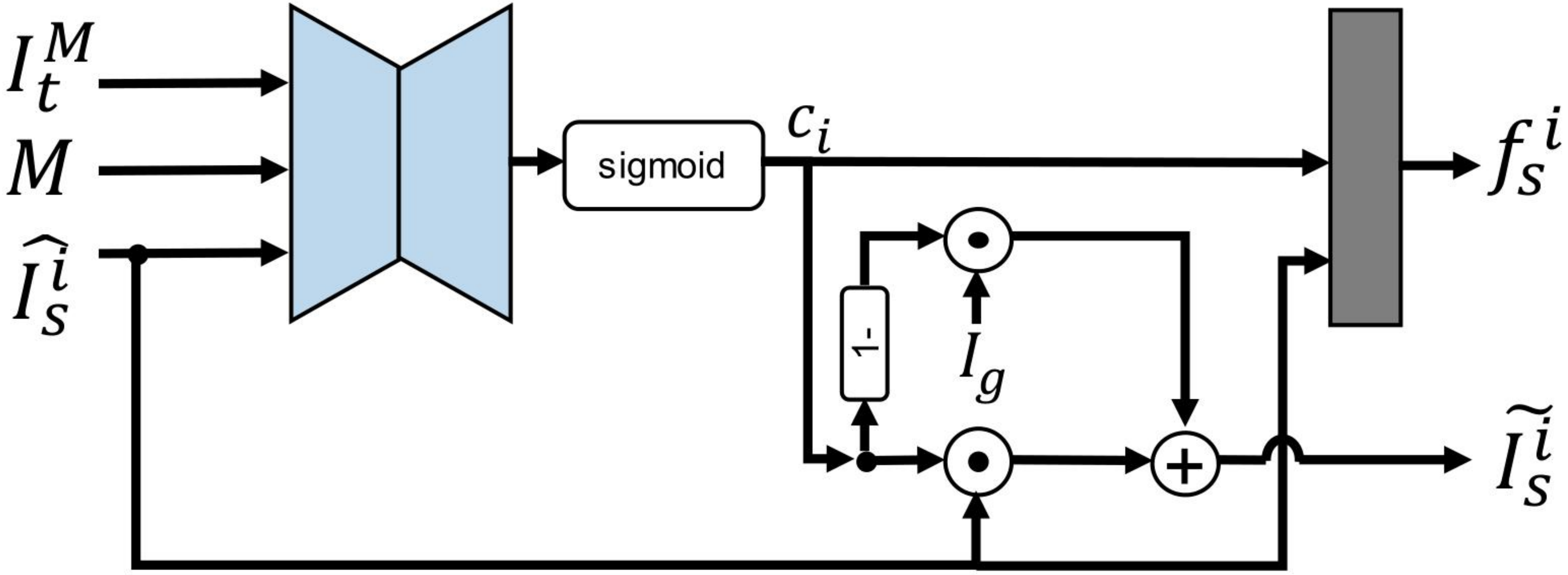


SPF

I_t^M **Target Image with Mask** M **Mask** $\widehat{I_s^i}$ **Refined Source Image** I_g **Inpainting Result from ProFill**

c_i **Confidence Map** $\widetilde{I_s^i}$ **Merged Refined Source Image** f_s^i **Packed Features**

Single-Proposal Fusion Module



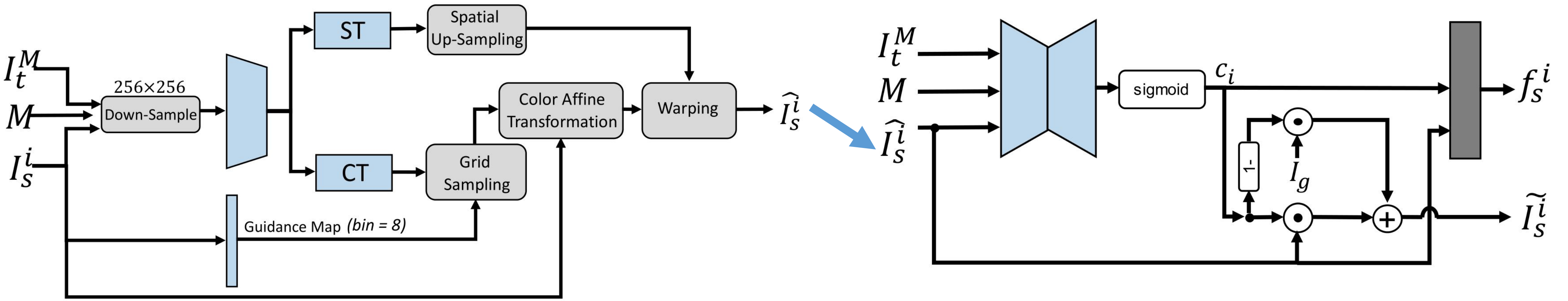
SPF

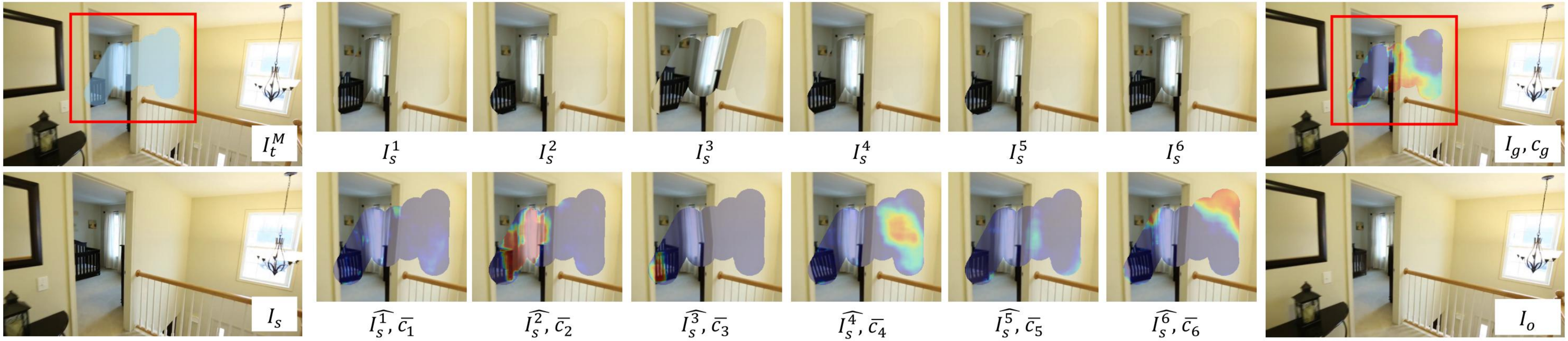
It is easier to understand the SPF from the formulation

$$\tilde{I}_s^i = c_i \odot \hat{I}_s^i + (1 - c_i) \odot I_g$$

Single-Proposal Fusion Module

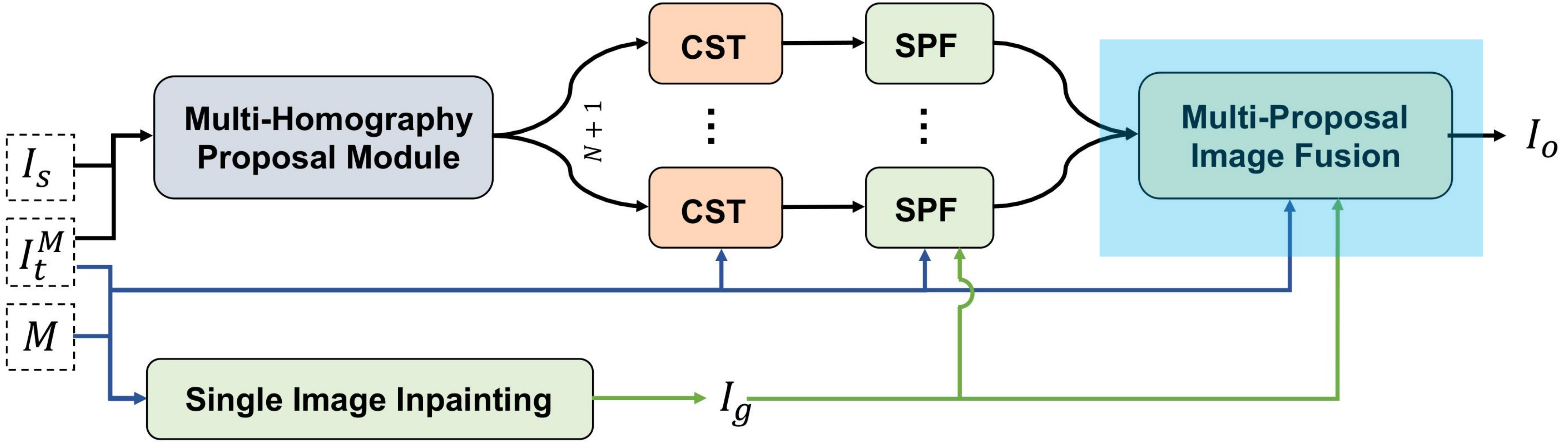
Recall





Intermediate Results

Multi-Proposal Image Fusion Module



I_s Source/Reference Image

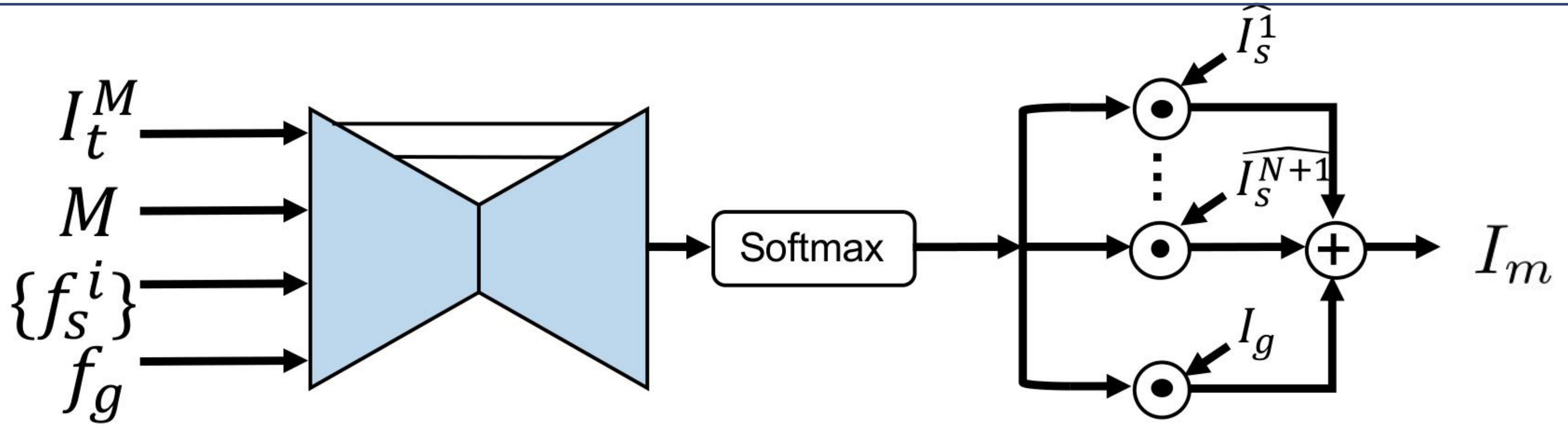
I_t^M Target Image with Mask

M Mask

CST Color-Spatial Transformer

SPF Single-Proposal Fusion

Multi-Proposal Image Fusion Module



MPF

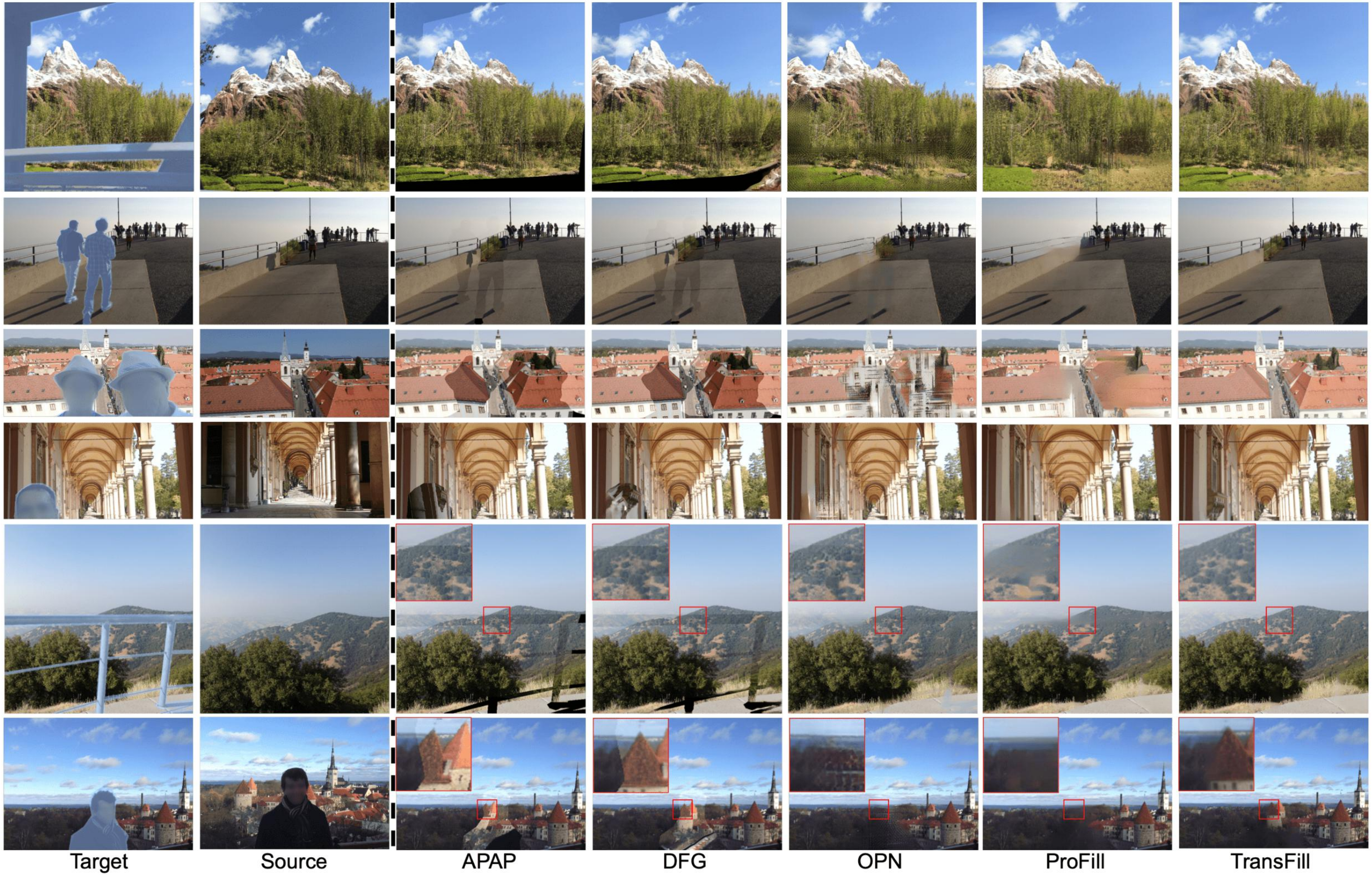
I_t^M **Target Image with Mask** M **Mask** $\widehat{I_s^i}$ **Refined Source Image** I_g **Inpainting Result from ProFill**

f_s^i **Packed Features**

$$I_m = c_g \odot I_g + \sum_{i=1}^{N+1} \overline{c_i} \odot \hat{I}_s^i,$$

Final Result $I_o = I_t^M + M \odot I_m$

Visual Comparison



Ablation Study

Table 2: Ablation Study on Multi-Homography Proposals.

Clustering	N	Outlier Rejection	PSNR↑	SSIM↑	LPIPS↓
Depth	N=5	OANet	**37.576**	**0.9879**	0.0164
Depth	N=5	Ratio Test [34]	37.444	0.9876	0.0168
Random	N=5	OANet	37.499	0.9873	0.0166
Spatial	N=5	OANet	37.384	0.9876	0.0169
Depth	N=3	OANet	37.537	0.9878	**0.0162**
None	N=1	OANet	37.092	0.9868	0.0172

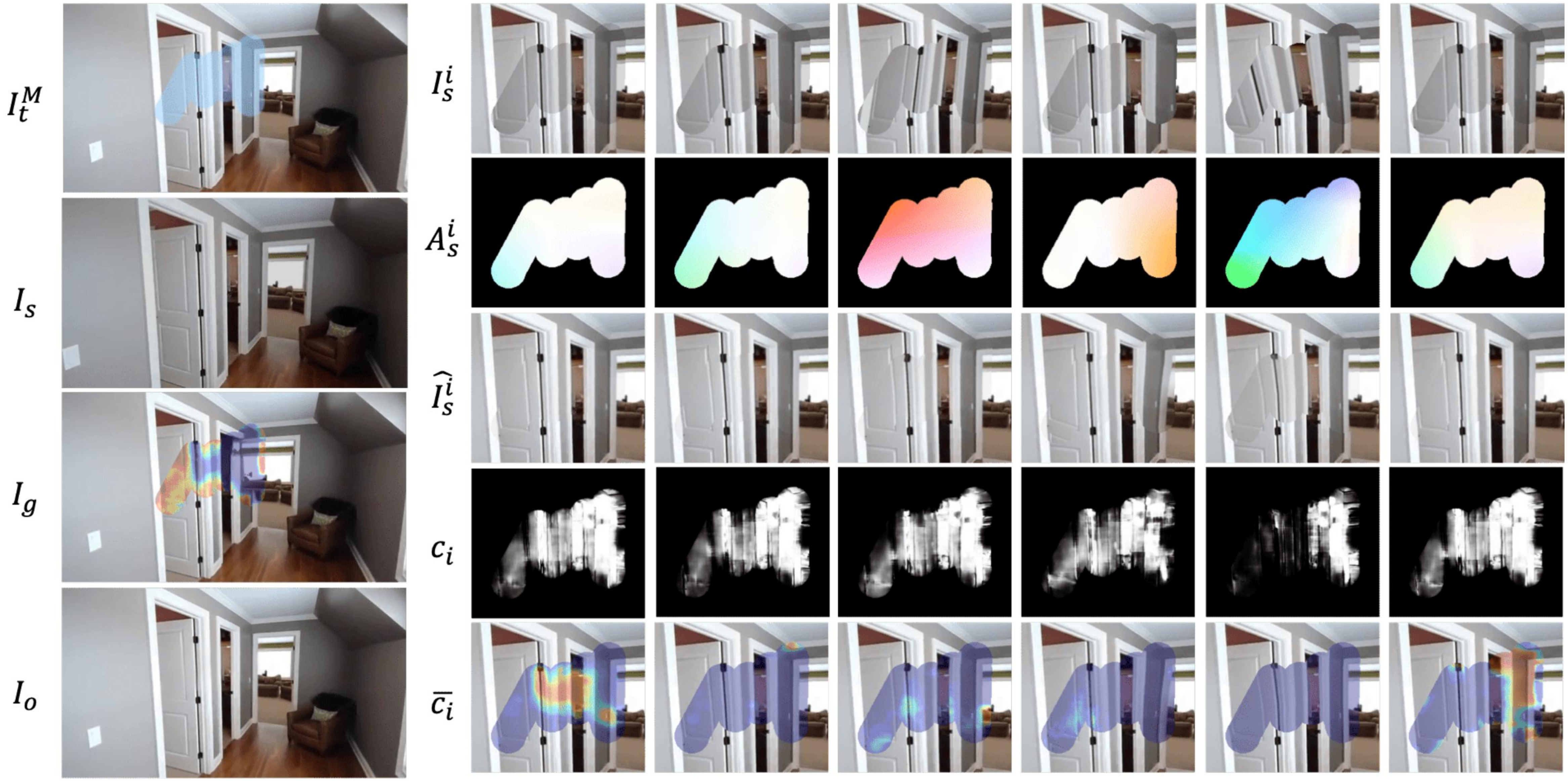
Table 3: Color-Spatial Transformation. **C**: Color, **S**:Spatial

Order	PSNR↑	SSIM↑	LPIPS↓
$C \to S$	**37.576**	**0.9879**	0.0164
$S \to C$	37.566	**0.9879**	**0.0163**
Only S	36.717	0.9866	0.0182
Only C	36.228	0.9849	0.0179

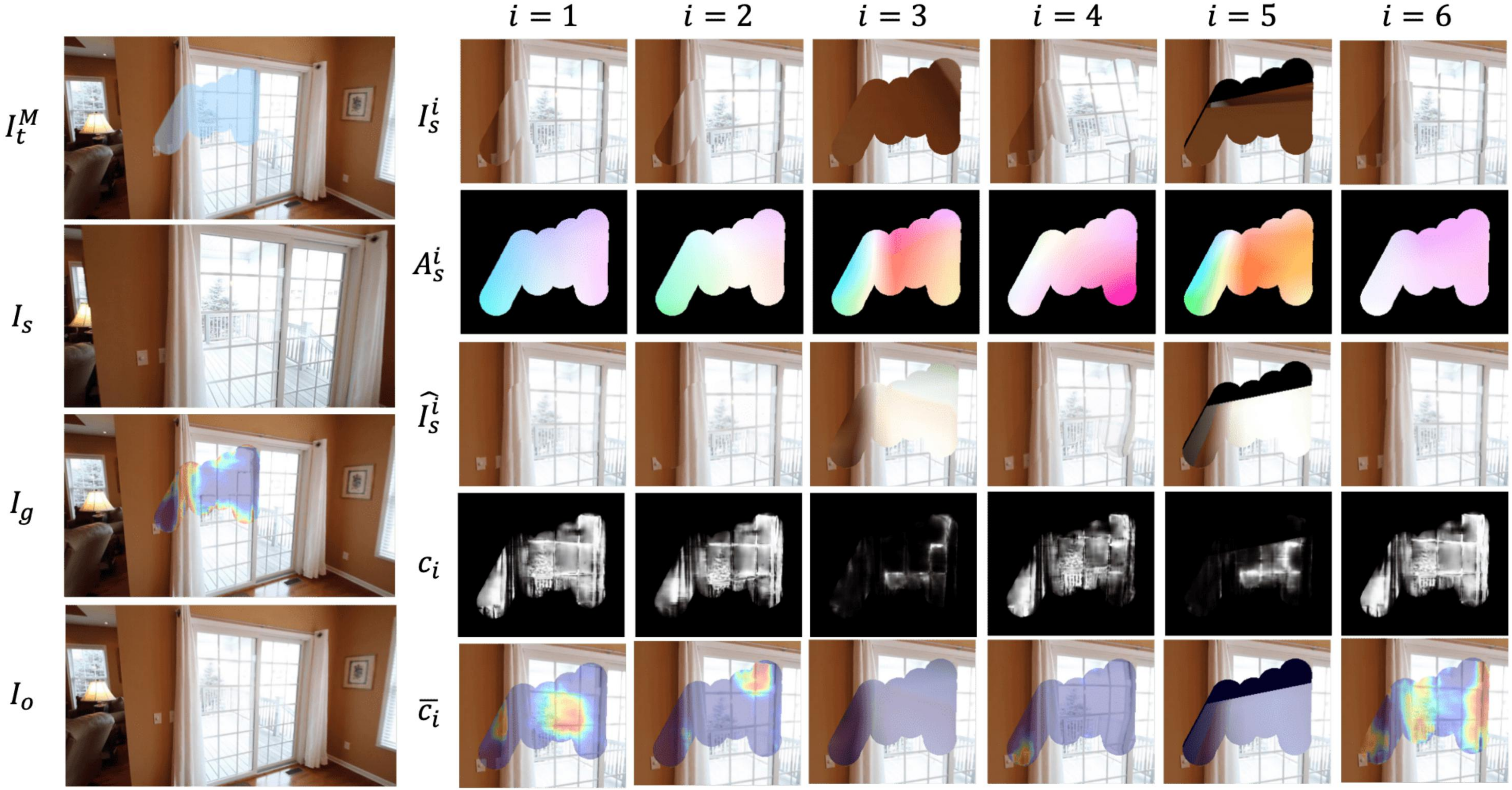
Table 4: Ablation Study on Pipeline Components. **CST**: Color-Spatial Transformer, **SPF**: Single-Proposal Fusion.

CST	SPF	PSNR↑	SSIM↑	LPIPS↓
✓	✓	**37.576**	**0.9879**	**0.0164**
✗	✓	35.579	0.9838	0.0183
✓	✗	36.710	0.9861	0.0188
✗	✗	33.484	0.9782	0.0249

More Intermediate Results

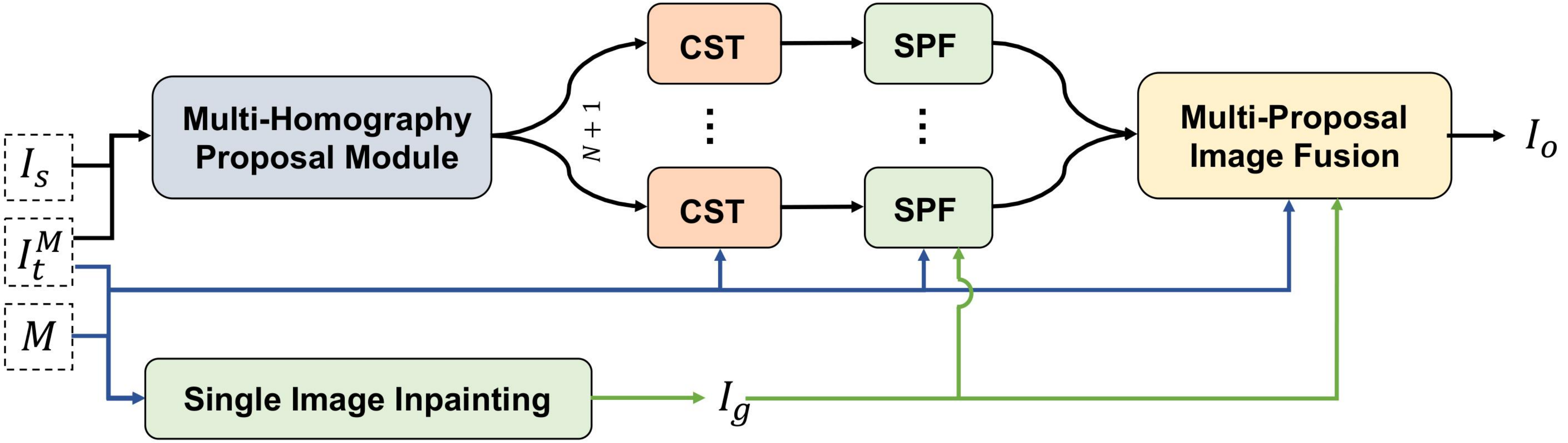


More Intermediate Results



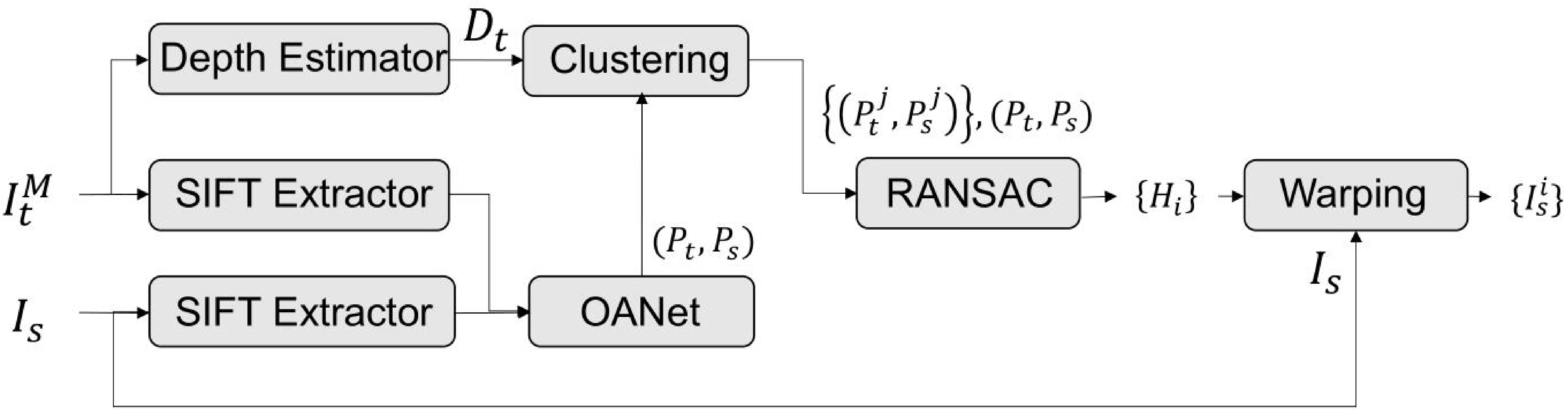
Conclusion

- TransFill, a **multi-homography** estimation pipeline to obtain multiple transformations of the source image, where each **aligns a specific region** to the target image.
- Propose to learn a **color and spatial transformer** to simultaneously perform a color matching and make a per-pixel spatial transformation to **address any residual differences** after the initial alignment.
- Learn weights suitable for **combining all final proposals** with a single image inpainting result



Discussion

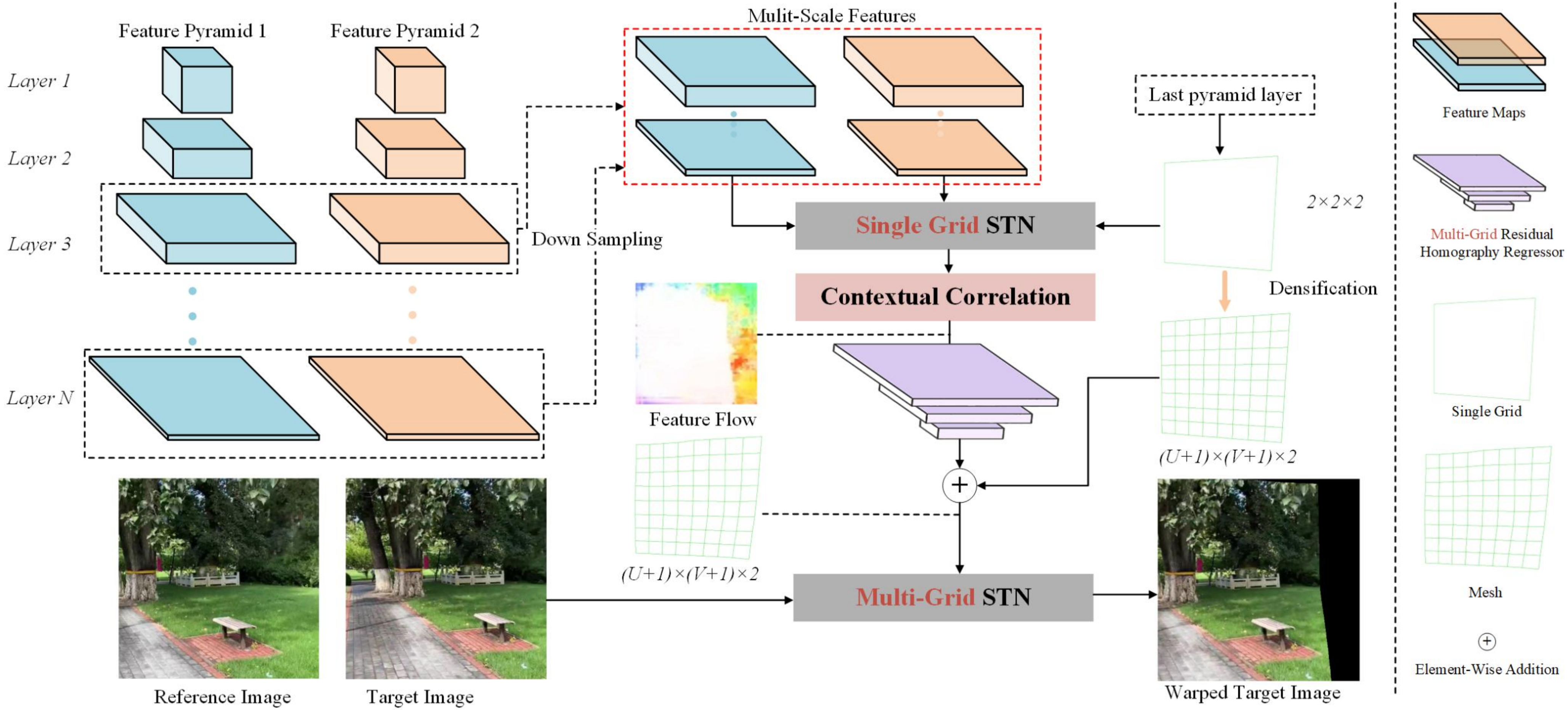
- The model is too large, not very compact, not end-to-end. Used pre-trained models or traditional methods: (1) SIFT, (2) OANet, (3) RANSAC, (4) Agglomerative Clustering Method, (5) MonoDepth Estimation, (6) ProFill, (7) Deep Bilateral Filtering



Multi-homography Proposal Module

Discussion

End-to-end, address the parallax using **multi-grid homography**, fast alignment



Nie, L., Lin, C., Liao, K., Liu, S., & Zhao, Y. (2021). Depth-Aware Multi-Grid Deep Homography Estimation with Contextual Correlation. *ArXiv, abs/2107.02524*.

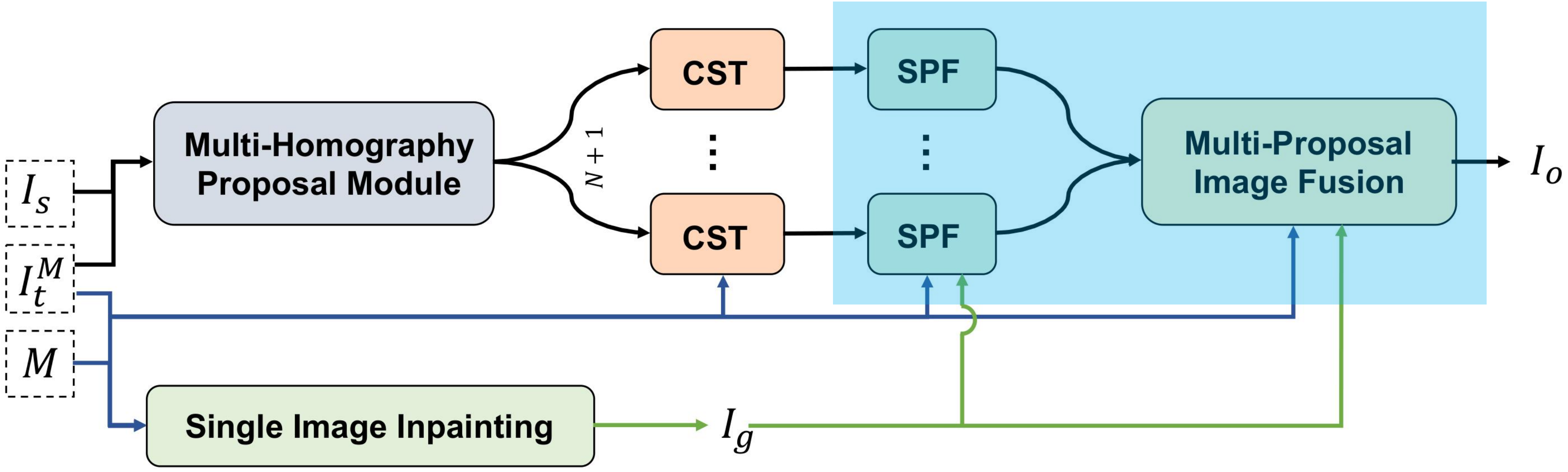
Discussion

Inputs SIFT+RANSAC LB-UDHN APAP (100×100) AANAP (100×100) robust ELA (100×100) Ours (8×8)

Nie, L., Lin, C., Liao, K., Liu, S., & Zhao, Y. (2021). Depth-Aware Multi-Grid Deep Homography Estimation with Contextual Correlation. *ArXiv, abs/2107.02524*.

Discussion

- Merge the multiple fusion strategies



北京交通大學

Thanks

Email: kang_liao@bjtu.edu.cn