2021 IEEE International Conference on Image Processing
19-22 September 2021 • Anchorage, Alaska, USA
Imaging Without Borders

Paper ID: #3160

IEEE Signal Processing Society

FAST AND ACCURATE HOMOGRAPHY ESTIMATION USING EXTENDABLE COMPRESSION NETWORK

Yilei Chen†, Guoping Wang†, Ping An†, Zhixiang You†, Xinpeng Huang†*
†School of Communication and Information Engineering, Shanghai University, Shanghai, China

PROBLEM

Fast and accurate homography estimation between images is crucial for relative pose estimation in autonomous exploration. Recently, learning-based methods have been proposed to solve challenging cases like large displacements, where traditional methods may degrade, with semantic information.

However, the following issues make most of them infeasible in practical applications.

- Low inference speed
- Large model size
- Degraded accuracy under large displacements

CONTRIBUTION

To solve the above problems, we introduce ShuffleNetV2 compressed units [1] to build our network named ShuffleHomoNet, which consists of one basic compressed network and two extended versions.

The main contributions are

- A basic ShuffleHomoNet is proposed based on ShuffleNetV2 compressed units, which can greatly accelerate the homography estimation process and reduce the model size.
- A multiscale weight-shared form and a recurrent coarse-to-fine form are extended from the basic network. The former additionally processes the half-scale input for further dealing with the large displacements, and the latter achieves the optimal performance in the case of sufficient computational resources.
- Experimental results show that our extendable networks can well balance the accuracy and inference speed compared to other methods, and the sizes of all models are less than 9MB.

METHOD

Problem Formulation

A homography matrix $H \in \mathbb{R}^{3\times3}$ between two images represents the planar projective transformation of matched points on the same plane. In this paper, we follow the work of [2], which calculate H_{4pt} instead of H for easier convergence. Our target is to minimize the errors between estimated $\widehat{H}_{4pt}$ and its ground truth.

H_{4pt} is equivalent to the original homography matrix, which is transformed from H using four pairs of matched points:

$$H_{4pt} = \begin{pmatrix} \Delta u_1 & \Delta u_2 & \Delta u_3 & \Delta u_4 \\ \Delta v_1 & \Delta v_2 & \Delta v_3 & \Delta v_4 \end{pmatrix}^T, \quad (1)$$

where $(\Delta u, \Delta v)$ are the offsets of the matched point between two images after multiplying by H.

Basic Compressed ShuffleHomoNet

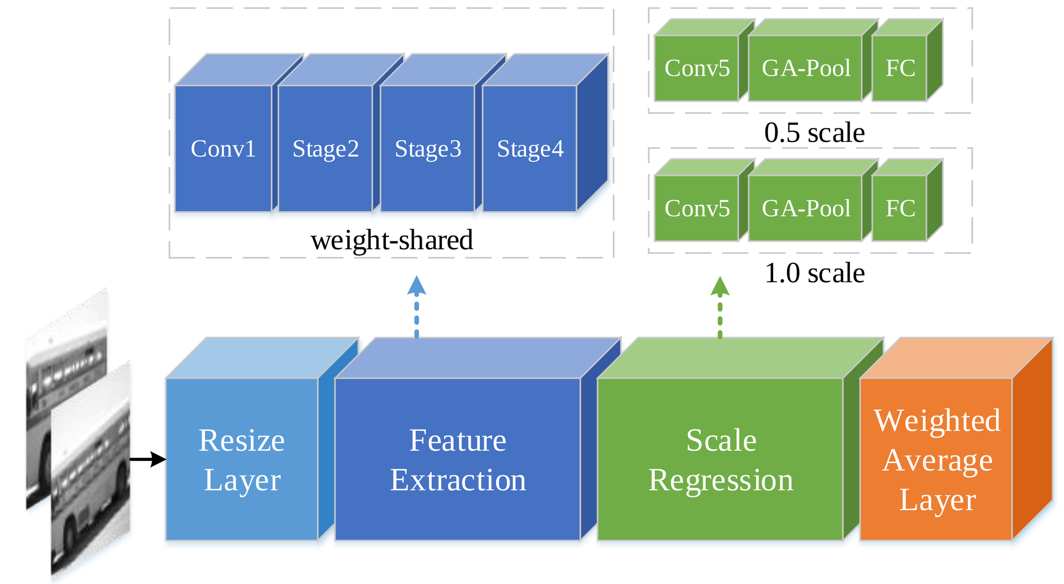
Layer	Output size	KSize	Stride	Repeat	Channels 0.5×	Channels 1×
Conv1	128×128	3×3	2	1	24	24
LA-Pool	64×64	2×2	2	1	24	24
Stage2	32×32 32×32		2 1	1 3	64	128
Stage3	16×16 16×16		2 1	1 7	128	256
Stage4	8×8 8×8		2 1	1 3	256	512
Conv5	4×4	1×1	1	1	1024	1024
GA-Pool	1×1	4×4		1	1024	1024
FC				1	8	8

The overall architecture of the basic network is shown above, which extracts features with the ShuffleNet compressed units, and regresses the homography matrix after the fully-connected (FC) layer.

Specifically, the feature extraction mainly consists of three stages, each of which is built by the repeated ShuffleNet compressed units. Due to its depth-wise separable convolution and group convolution, more basic units can be added to extract better features without increasing complexity.

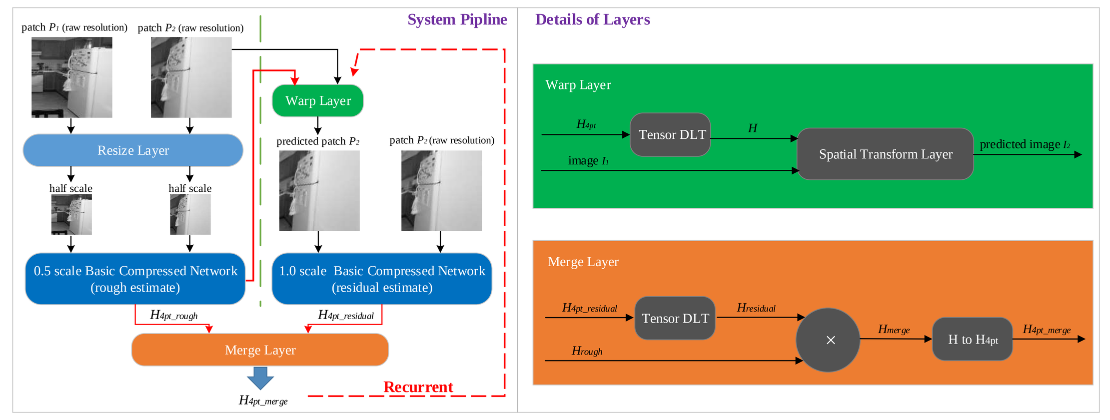
Two Extended Forms

1) Multiscale Weight-shared ShuffleHomoNet



To further deal with the large displacements, we extend the basic network to a multiscale weight-shared form, as shown above. The model additionally takes the half-scale patches as input, which can make the large displacements become small. The generalized features of multiscale input are extracted through the weight-shared feature extraction module. The weighted average layer finally fuses the multiscale predictions of the regression modules according to the H_{4pt} scale property (see details in paper).

2) Recurrent Coarse-to-fine ShuffleHomoNet



In the case of sufficient computational resources, iterative optimization is adopted for the optimal performance. As shown above, the basic network is first applied at 0.5 scale to estimate the rough homography, and then the predicted patch P_2 warped by the patch P_1 with this rough estimation is processed by the weight-shared network to estimate the residual matrix.

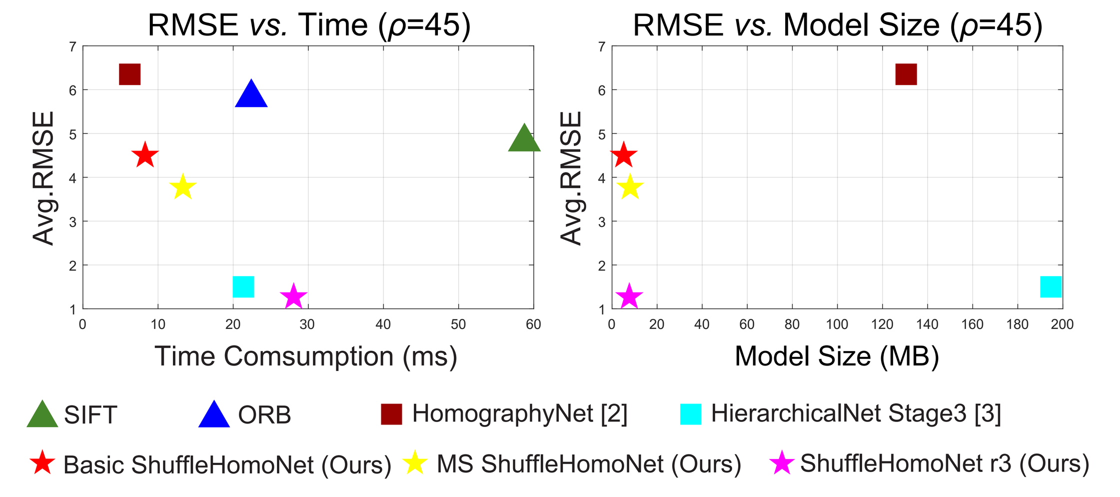
RESULTS

The performance of the proposed networks is evaluated on the synthetic dataset as [2], and $\rho = 32/45$ pixels represents normal and large displacements.

1) H_{4pt} RMSE performance

ρ	Network	Top30%	Top60%	Top90%	Avg
32	SIFT	0.59	0.70	0.89	3.23
	ORB	0.74	1.06	1.38	3.89
	HomographyNet [2]	1.82	2.31	2.95	3.41
	HierarchicalNet stage3 [3]	0.22	0.28	0.35	0.43
	Basic ShuffleHomoNet	**1.12**	**1.45**	**1.87**	**2.21**
	MS ShuffleHomoNet	**0.79**	**1.11**	**1.53**	**1.89**
	Rec ShuffleHomoNet r1	1.01	1.32	1.74	2.07
	Rec ShuffleHomoNet r2	0.30	0.42	0.59	0.78
	Rec ShuffleHomoNet r3	**0.14**	**0.19**	**0.27**	**0.38**
45	SIFT	0.67	0.91	2.83	4.81
	ORB	0.91	1.39	3.10	5.82
	HomographyNet [2]	3.38	4.32	5.52	6.35
	HierarchicalNet stage3 [3]	0.56	0.77	1.06	1.50
	Basic ShuffleHomoNet	**2.32**	**2.98**	**3.82**	**4.50**
	MS ShuffleHomoNet	**1.61**	**2.46**	**2.94**	**3.77**
	Rec ShuffleHomoNet r1	2.00	2.62	3.49	4.28
	Rec ShuffleHomoNet r2	0.71	1.01	1.47	2.10
	Rec ShuffleHomoNet r3	**0.35**	**0.51**	**0.78**	**1.27**

2) RMSE vs. Inference Time / Model Size



REFERENCES

[1] N. Ma, X. Zhang, H. Zheng, and J. Sun, "Shufflenet v2: Practical guidelines for efficient cnn architecture design," in *Proc. Europ. Conf. Comp. Vis.*, 2018, pp. 122-138.

[2] D. DeTone, T. Malisiewicz, and A. Rabinovich, "Deep image homography estimation," *arXiv preprint arXiv:1606.03798*, 2016.

[3] F. Nowruzi, R. Laganiere, and N. Japkowicz, "Homography estimation from image pairs with hierarchical convolutional networks," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops*, 2017, pp. 913-920.