Semi-supervised Learning of Camera Motion from a Blurred Image

Nimisha T M[*],Vijay Rengarajan[†], and Rajagopalan Ambasamudram[*]
[*]Indian Institute of Technology Madras, [†]Carnegie Mellon University

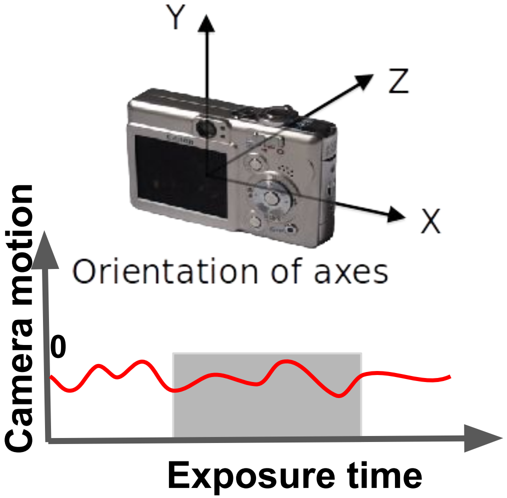
Presented by: Subeesh Vasu
IPCV Lab (http://www.ee.iitm.ac.in/ipcvlab/)
Indian Institute of Technology Madras

Problem Statement

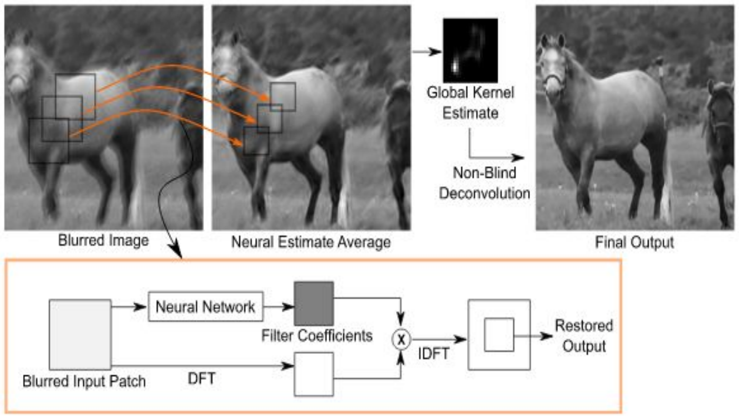
Given a single space-variant motion blurred image, estimate motion underwent by camera during the exposure time.

Space variant blurred image

Estimate camera motion



Prior Works



Motion Estimation

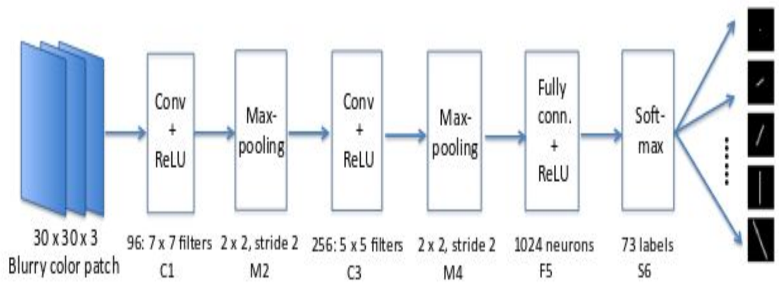
[1] Ayan ECCV 2016

- Estimates uniform motion kernels

[1] Ayan Chakrabarti, “A neural approach to blind motion deblurring,” ECCV 2016.

Prior Works

Motion Estimation

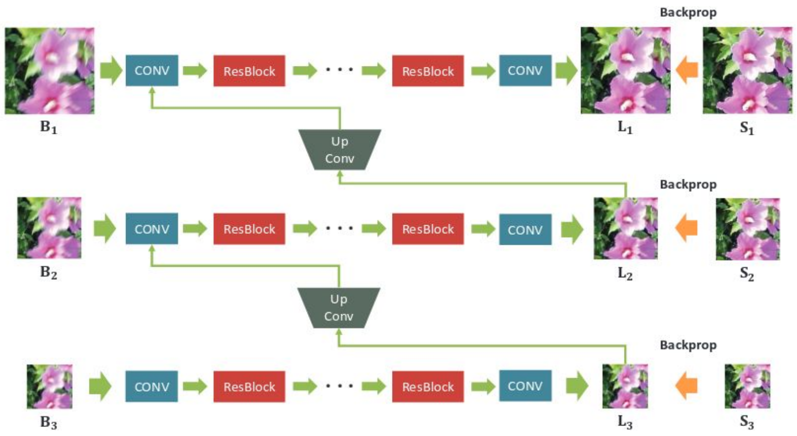


[2] Sun et al. CVPR 2015

- Assumes parametric kernels
- Performs kernel classification over patches

[2] Jian Sun, Wenfei Cao, Zongben Xu, and Jean Ponce, “Learning a convolutional neural network for non-uniform motion blur removal,” CVPR 2015.

Prior Works



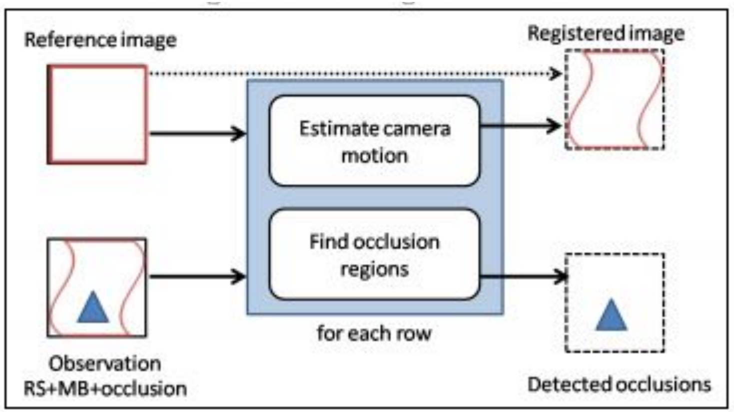
End-to-End Deblurring

[3] Nah et al. CVPR 2017

- No motion estimation step involved

[3] Nah, S., Kim, T.H., Lee, K.M., “ Deep multi-scale convolutional neural network for dynamic scene deblurring”, CVPR 2017

Prior Works



Change Detection

[4,5] Vijay et al. ECCV 2014/TPAMI 2016

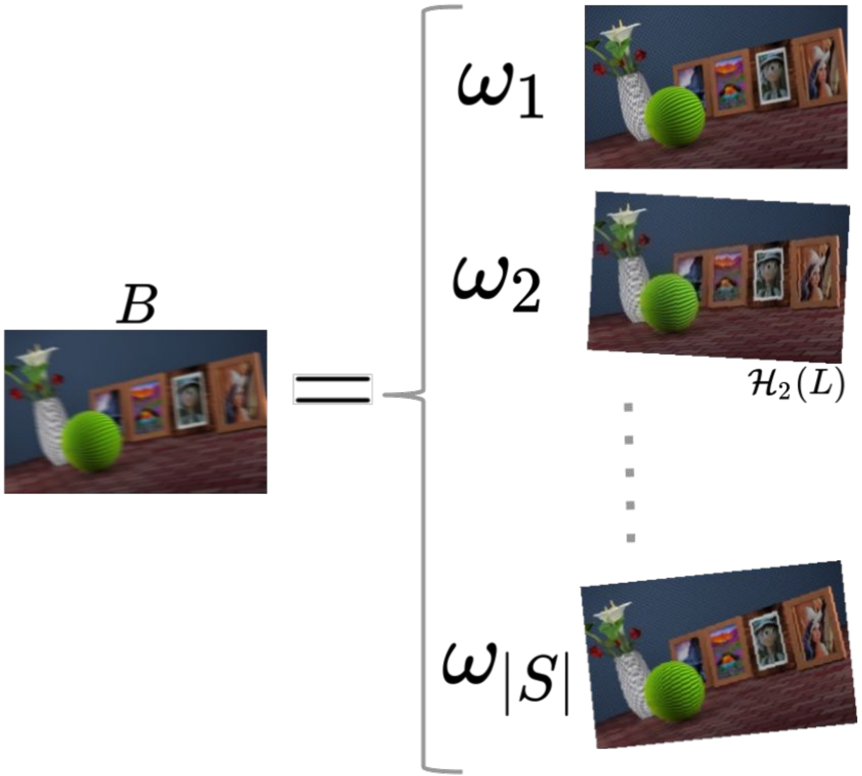
[4] Vijay Rengarajan, Rajagopalan A N and Aravind R, “Change detection in the presence of motion blur and rolling shutter effect,” ECCV 2014.

[5] Vijay Rengarajan, Rajagopalan A N, Aravind R, and Guna Seetharaman, “Image registration and change detection under rolling shutter motion blur,” TPAMI 2017.

Contributions

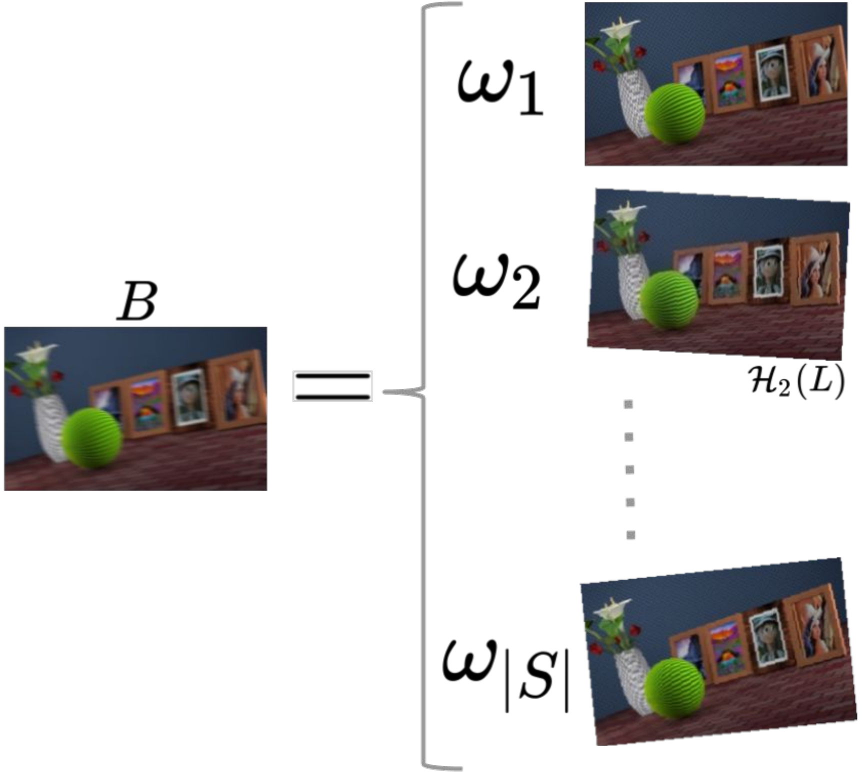
- Estimate **global camera motion** rather than local blur kernels
- **Leverage** the availability of clean image and ground truth motion during training
- Show applications of the estimated motion in **Deblurring** and **Change detection**

Image Formation Model



Blurred image is a weighted sum of warped instances of its clean version.

Image Formation Model

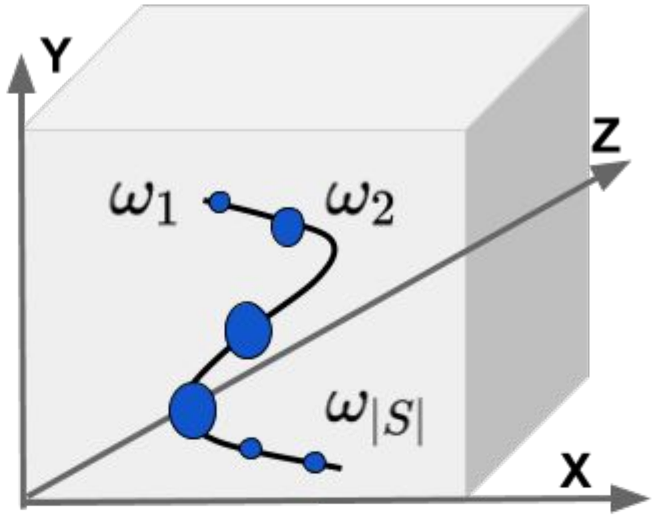


Blurred image is a weighted sum of warped instances of its clean version.

$$B = \sum_{k=1}^{|S|} \omega_k \mathcal{H}_k(L)$$

S	Discretized pose space
$\mathcal{H}$	Warping matrix
ω	Pose space weights

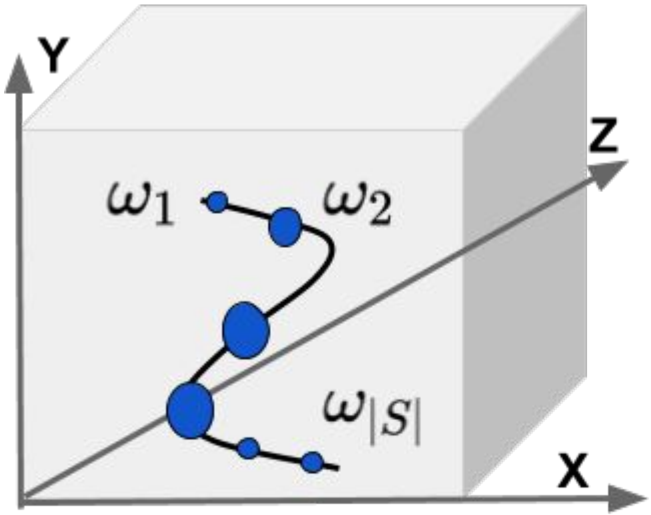
Camera Pose Weights



$$\Omega = [\omega_1, \ldots, \omega_{|S|}]^T$$

- Camera Pose Weights indicates the fraction of the exposure time spend at each pose

Camera Pose Weights



$$\Omega = [\omega_1, \ldots, \omega_{|S|}]^T$$

- Camera Pose Weights indicates the fraction of the exposure time spend at each pose

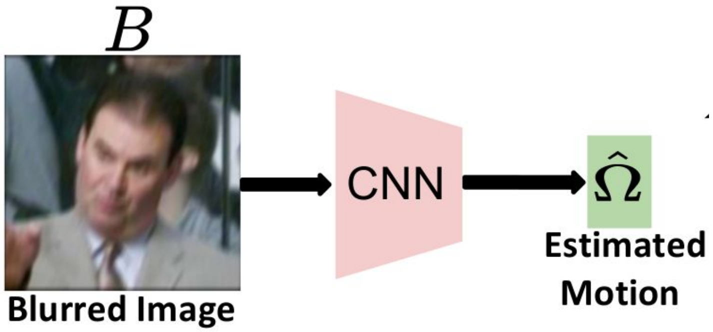
Properties:

- Energy preserving

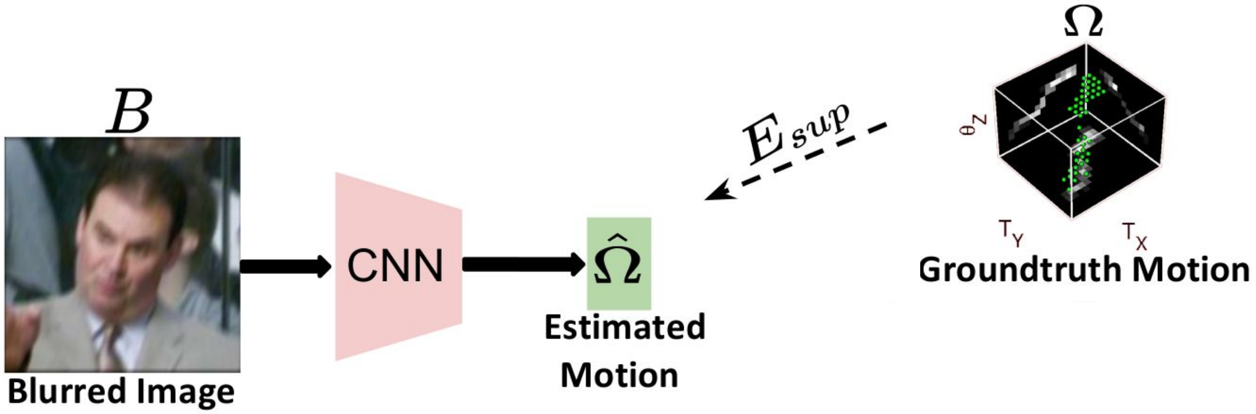
$$\sum_{k=1}^{|S|} \omega_k = 1$$

- Sparse

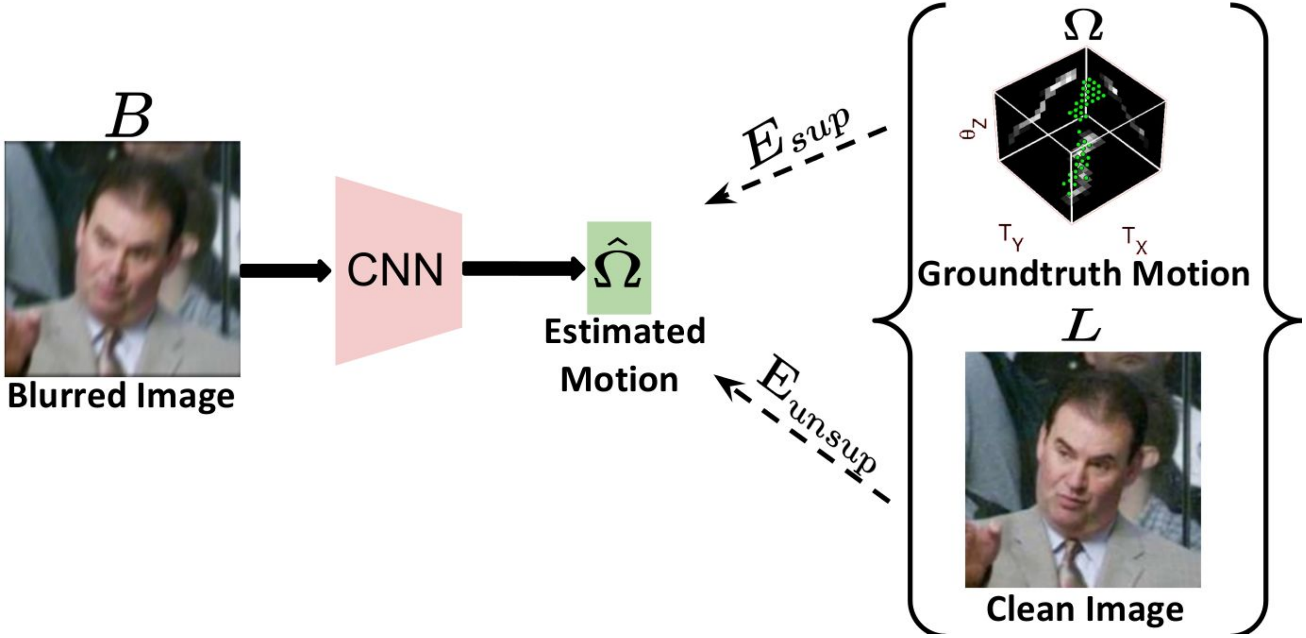
Our Approach



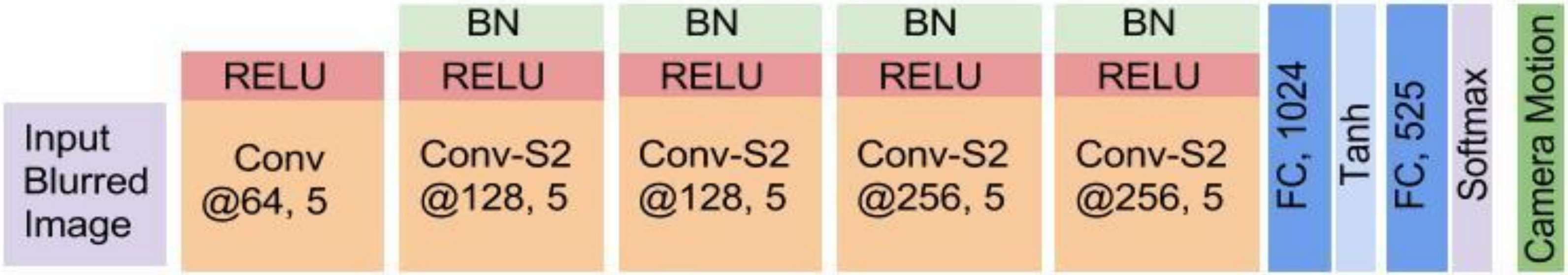
Our Approach



Our Approach



Network Architecture



Cost Functions

Supervised cost

- The predicted pose weights are compared with GT using ℓ_2 loss

$$E_{mse} = \|\widehat{\Omega} - \Omega_{orig}\|_2^2$$

Cost Functions

Supervised cost

- Camera sees only a sparse set of poses from the entire camera pose space
- Hence, impose sparsity constraint

$$E_{spar} = \|\widehat{\Omega}\|_1$$

Cost Functions

Supervised cost

- Camera sees only a sparse set of poses from the entire camera pose space
- Hence, impose sparsity constraint

$$E_{spar} = \|\widehat{\Omega}\|_1$$

$$E_{sup} = \lambda_1 E_{mse} + \lambda_2 E_{spar}$$

Cost Functions

Unsupervised cost

- Exploit the association of the latent image and camera motion for better convergence

$$E_{unsup} = \lambda_3 \| B - \underbrace{\textstyle\sum_{k=1}^{|S|} \widehat{\omega}_k \mathcal{H}_k(L)}_{\boxed{\text{Image formation forward model}}} \|_2^2$$

Cost Functions

Unsupervised cost

- Exploit the association of the latent image and camera motion for better convergence

$$E_{unsup} = \lambda_3 || B - \underbrace{\textstyle\sum_{k=1}^{|S|} \widehat{\omega}_k \mathcal{H}_k(L)}_{\text{Image formation forward model}} ||_2^2$$

$$E_{unsup} = \lambda_3 || b - A\widehat{\Omega} ||_2^2$$

Lexicographically ordered column vectors of blurred image

Training Details

- Prepared the training and validation datasets from PASCAL VOC dataset

Training Details

- Prepared the training and validation datasets from PASCAL VOC dataset
- Resized images to 128 x 128

Training Details

- Prepared the training and validation datasets from PASCAL VOC dataset
- Resized images to 128 x 128
- Blurred images are generated using 200k 3D camera motion trajectories

Training Details

- Prepared the training and validation datasets from PASCAL VOC dataset
- Resized images to 128 x 128
- Blurred images are generated using 200k 3D camera motion trajectories
- Training dataset size of 200k **space-variantly** blurred images

Training Details

3D camera motion estimation:

- Assumed in-plane translations (t_x and t_y) ranging from [−2 : 2] pixels with a step size of one pixel

Training Details

3D camera motion estimation:

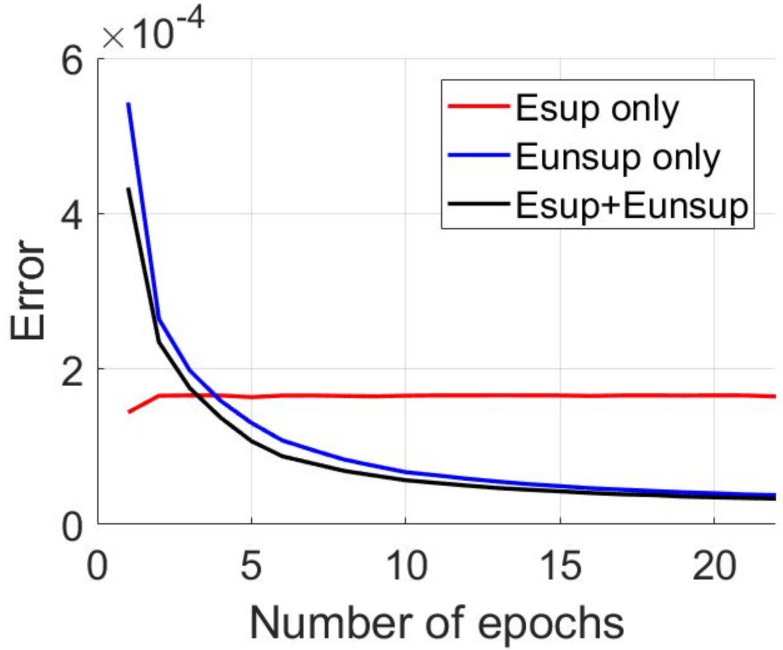
- Assumed in-plane translations (t_x and t_y) ranging from [−2 : 2] pixels with a step size of one pixel
- In-plane rotation $r_z \in [-5 : 5]^\circ$ with a step size of 0.5

Training Details

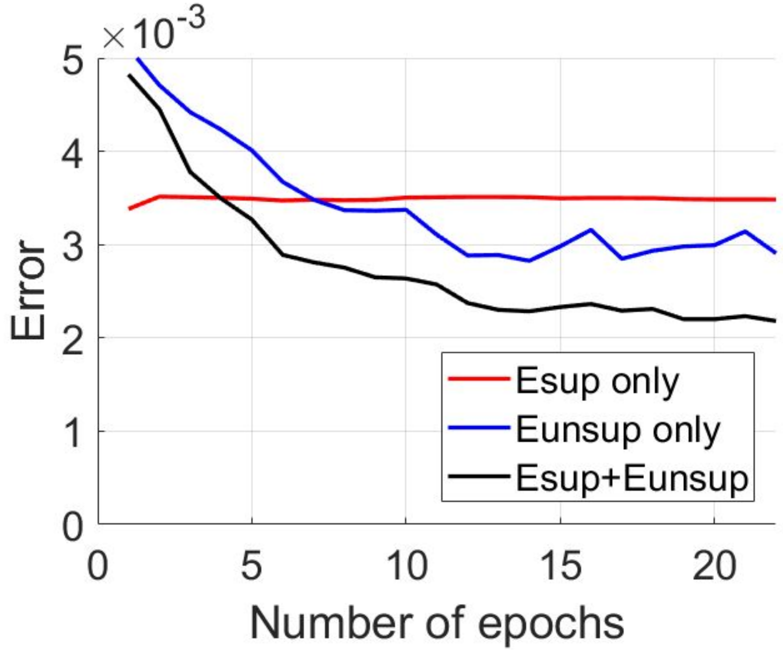
3D camera motion estimation:

- Assumed in-plane translations (t_x and t_y) ranging from [−2 : 2] pixels with a step size of one pixel
- In-plane rotation $r_z \in [-5 : 5]^\circ$ with a step size of 0.5
- Total pose space containing $|S| = 525$ poses.

Error plots



×10^-4
6
4
2
0
Error
0
5
10
15
20
Number of epochs
Esup only
Eunsup only
Esup+Eunsup



×10^-3
5
4
3
2
1
0
Error
0
5
10
15
20
Number of epochs
Esup only
Eunsup only
Esup+Eunsup

Experiments

Quantitative metric: Estimated camera motion and ground truth (GT) motion compared using Normalized Cross Correlation (NCC)

Experiments

Quantitative metric: Estimated camera motion and ground truth (GT) motion compared using Normalized Cross Correlation (NCC)

NCC values near to 1 indicates better estimates

Experiments

Quantitative metric: Estimated camera motion and ground truth (GT) motion compared using Normalized Cross Correlation (NCC)

NCC values near to 1 indicates better estimates

Visual results in applications of

- Deblurring
- Change detection

Quantitative Evaluation of Motion Estimation

- Took 10 different GT camera motions from the dataset in [6]

[6] Rolf Köhler, Michael Hirsch, Betty Mohler, Bernhard Schölkopf, and Stefan Harmeling, “Recording and playback of camera shake: Benchmarking blind deconvolution with a real-world database,” ECCV 2012.

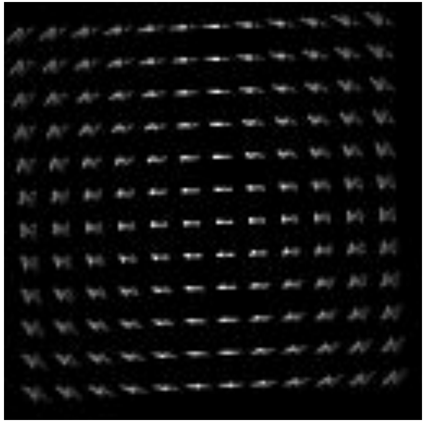
Quantitative Evaluation of Motion Estimation

- Took 10 different GT camera motions from the dataset in [6]
- Clean images from the test set of PASCAL VOC dataset are blurred using these motion

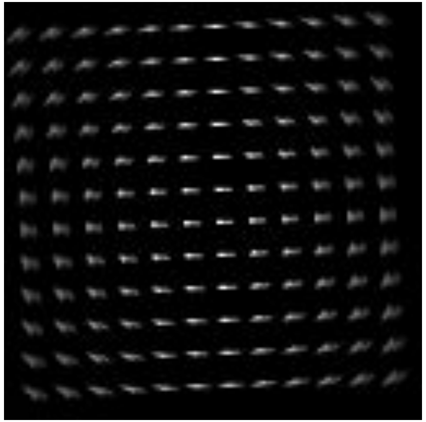
Quantitative Evaluation of Motion Estimation

- Took 10 different GT camera motions from the dataset in [6]
- Clean images from the test set of PASCAL VOC dataset are blurred using these motion
- Predicted the camera motion using our network

Visual Result

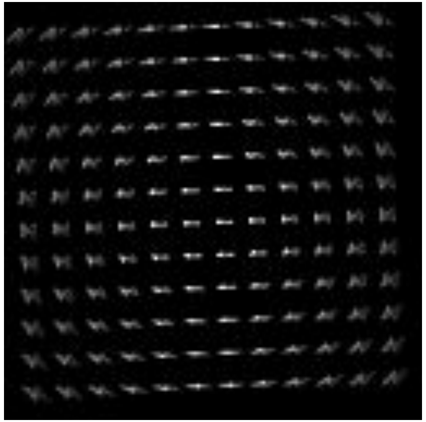


Original blur kernels

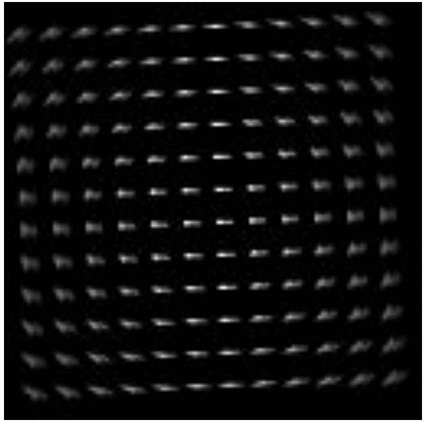


Estimated blur kernels

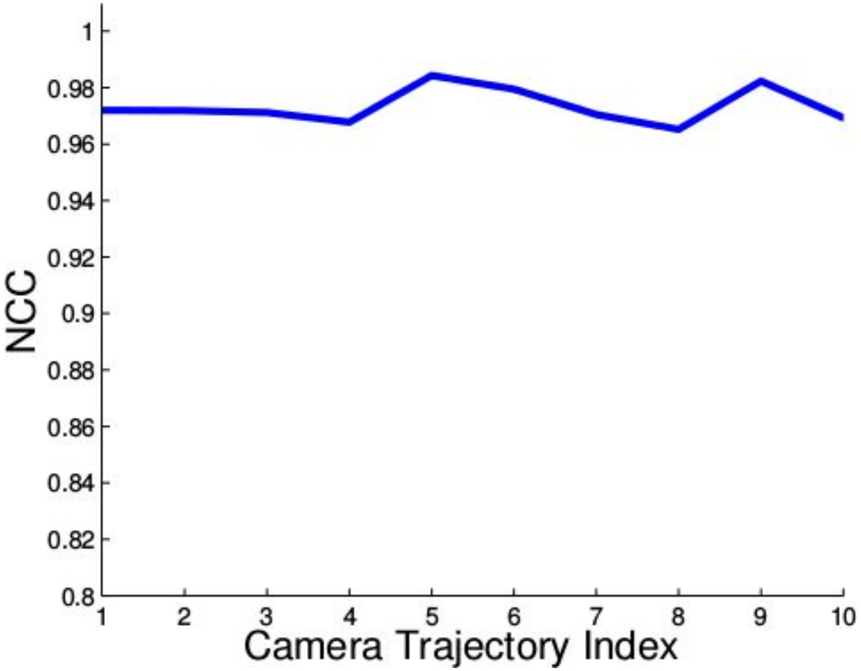
NCC Plot



Original blur kernels



Estimated blur kernels



Deblurring

From estimated camera motion, a non-blind deblurring can be carried out as

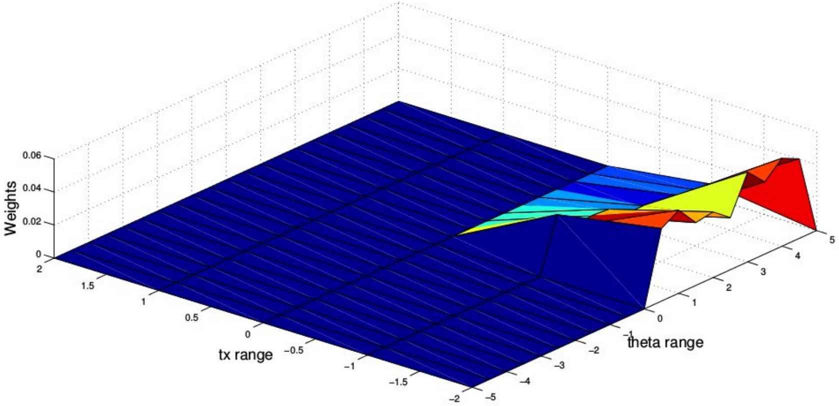
$$\min_l \{ \|b - \widehat{\mathcal{M}} l\|_2^2 + \lambda |\nabla l|_1 \}$$

Formed from $\hat{\Omega}$

Results

Clean image

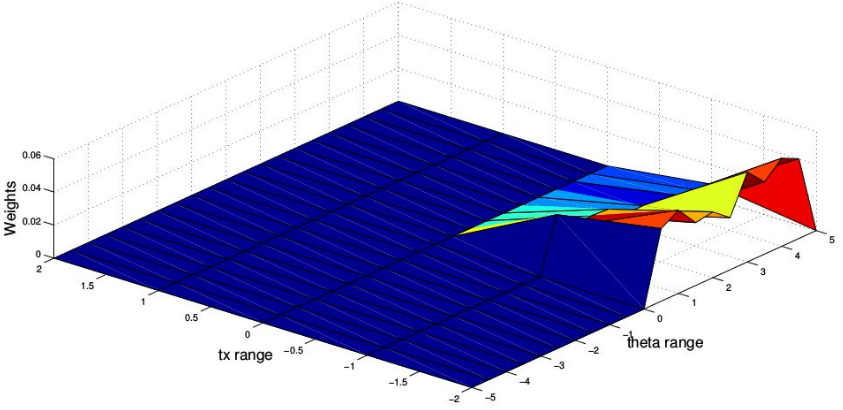
Blurred input



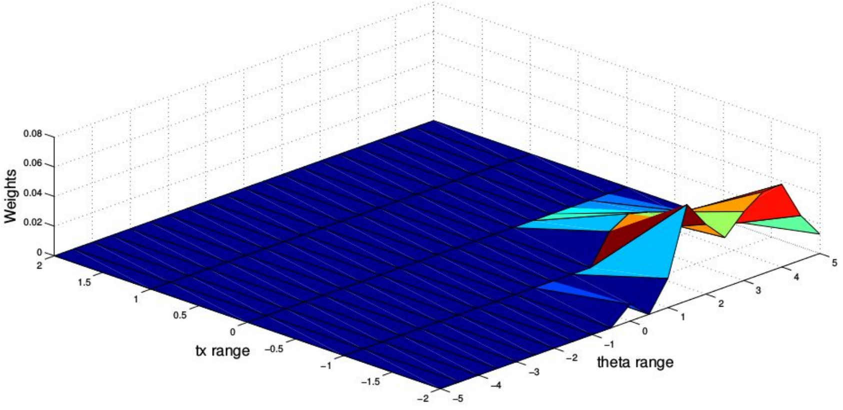
GT camera motion

Results

Clean image **Blurred input**



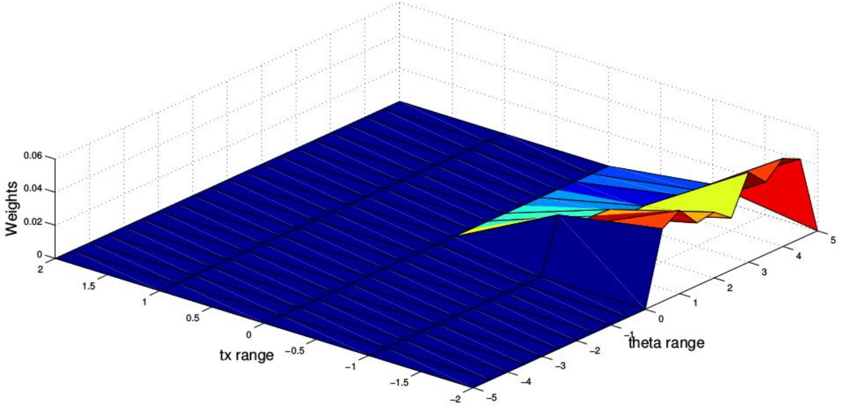
GT camera motion



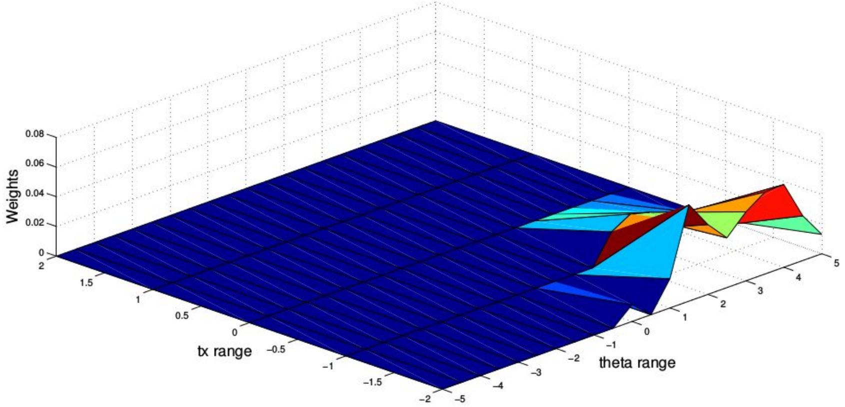
Estimated camera motion

Results

Clean image Blurred input Deblurred output



GT camera motion



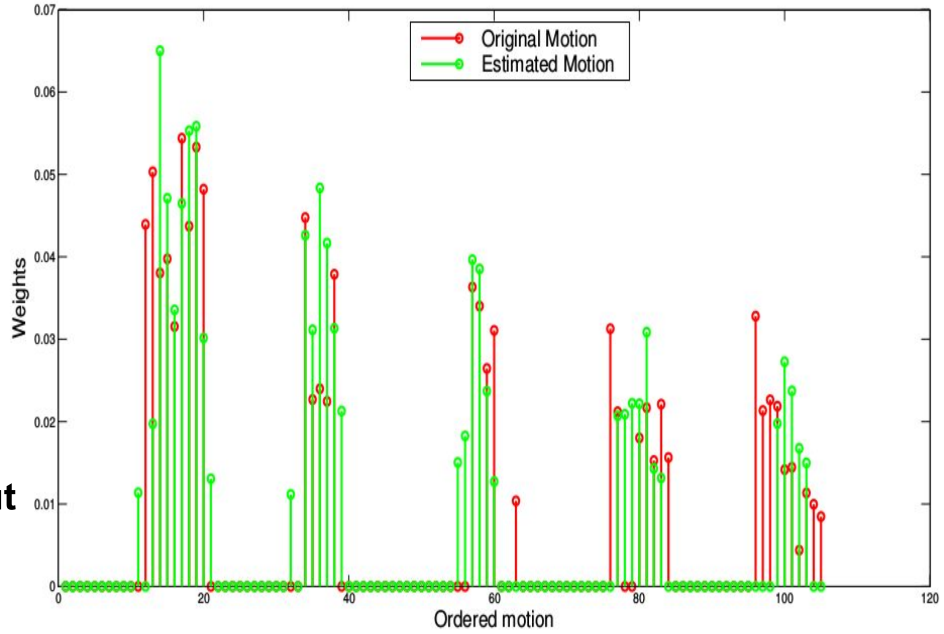
Estimated camera motion

Results

Clean image

Blurred input

Deblurred output



Ordered GT and estimated camera motion

Results

Blurred input

Deblurred output using Nah et al. [3]

[3] Nah, S., Kim, T.H., Lee, K.M, “ Deep multi-scale convolutional neural network for dynamic scene deblurring”, CVPR 2017.

Results

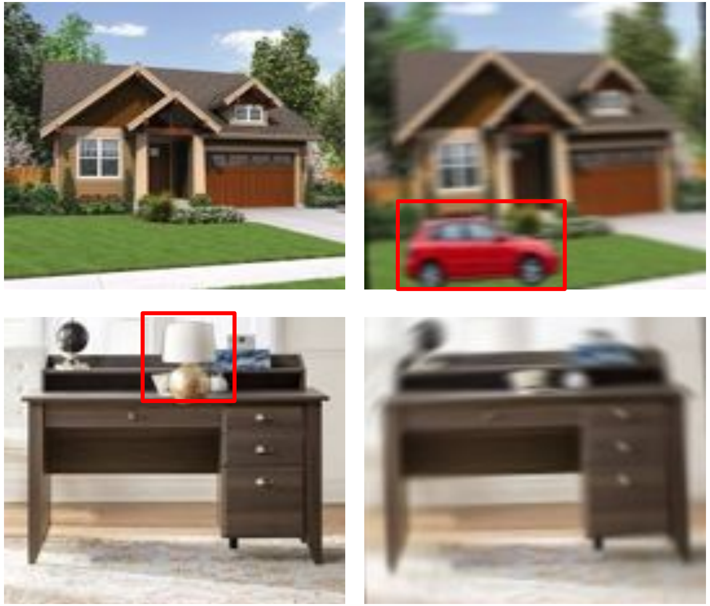
Blurred input | **Deblurred output using Nah et al. [3]** | **Our result**

Change Detection

Input 1

Input 2

Change Detection



Input 1 **Input 2**

Change Detection

Input 1 Input 2 Direct differencing

Change Detection

Input 1 (I_1) | Input 2 (I_2) | abs (I_2 Nah et al. [3] - I_1)

Change Detection

Input 1 **Input 2** **Result obtained from Vijay et al. [4]**

[4] Vijay Rengarajan, Rajagopalan A N and Aravind R, “Change detection in the presence of motion blur and rolling shutter effect,” ECCV 2014.

Change Detection

Input 1 **Input 2** **Our result** obtained by forward blurring Input 1 using the camera motion estimated from Input 2 followed by differencing

Quantitative Comparison for Change Detection

Methods	PCC	JC	YC
Ours	99.31	0.6808	0.7488
I_2 Nah et al. [3] - I_1	89.50	0.1198	0.1193
Vijay et al. [4]	94.87	0.2613	0.2731

T/F : True/False, P/N : Positive/Negative

PCC	Percentage of correct classification : (TP+TN)/(TP+TN+FP+FN)
JC	Jaccard coefficient : TP/(TP+FP+FN)
YC	Yule coefficient : | TP/(TP+FP) + TN/(TN+FN)-1 |

Limitations

- Input size is limited to 128 x 128

Limitations

- Input size is limited to 128 x 128

Reason:

❑ FC layers involved in the network

Limitations

- Input size is limited to 128 x 128
- Pose space restricted to 525 poses

Limitations

- Input size is limited to 128 x 128
- Pose space restricted to 525 poses

Reason:

- ❑ The size of A matrix is controlled by input image size and number of poses.
- ❑ Increasing any of these results in increased training time and memory utilization.

Conclusions

- Proposed a network to estimate global camera motion from blurred image
- Used cost functions that make use of both clean image and ground truth motion
- Proposed work achieve comparable performance in motion deblurring and state-of-the-art results in change detection

Thank You