# “Image de-fencing using RGB-D data”

Vikram Voleti
IIT-Hyderabad, India

Masters thesis at
IIT Kharagpur, India

(2013-2014)

Supervisor:

Prof. Rajiv Ranjan Sahay
Associate Professor,
Electrical Engineering,
IIT Kharagpur, India

Presenting at Max Planck Institute for Informatics, Saarbrücken, Germany

# Abstract

Objective: De-fence images

- Capture multiple views of the same scene
- Use RGB-D data to detect fence
- Inpaint fence regions using Loopy Belief Propagation

# Contents

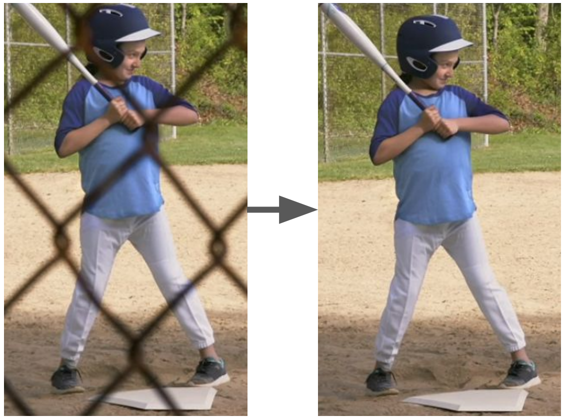
# 1. Introduction

2. Our approach
    a. Capture of multiple images
    b. Fence detection and alignment
    c. Computation of shifts between images
    d. Inpainting of fence region
3. Experiments and results
4. Comparison with other methods
5. Limitations and precautions
6. “Future” work

# 1. Introduction

## Objective:

To remove fence-like occlusions from images



Source: Google I/O 2017

## Use cases:

- Personal videos with fence-like occlusions
- Surveillance: security footage with fence-like occlusions
- Tourist spots like museums or high towers with fences for safety

## Previous work:

- Estimation from single image (not multi-modal)
  - via regular patterns[1] (not all fences),
  - or graph cuts[2] (not all fence pixels),
  - or through learning-based matting[3] (requires heavy user input),
- Estimation from multiple images[4] by inpainting not via an optimization-based framework

## Our novelty:

Use of multi-modal information involving depth maps (from Microsoft Kinect) to detect fences, inpainting via an optimization-based framework

[1] M. Park, K. Brocklehurst, Robert Collins, and Yanxi Liu, "Deformed lattice detection in real-world images using mean-shift belief propagation," IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), vol. 31, no. 10, pp. 1804–1816, 2009.
[2] Y. Boykov and V. Kolmogorov, "An experimental comparison of min-cut/max-flow algorithms," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 26, no. 9, pp. 1124–1137, 2004.
[3] Y. Zheng and C. Kambhamettu, "Learning based digital matting," in IEEE International Conference on Computer Vision (ICCV), 2009, pp. 889–896
[4] M. Park, K. Brocklehurst, R. T. Collins, and Y. Liu. "Image de-fencing revisited." In Asian Conference on Computer Vision, pages 422-434, 2011.

1. Introduction

# 2. Our approach

a. Capture of multiple images

b. Fence detection

c. Computation of shifts between images

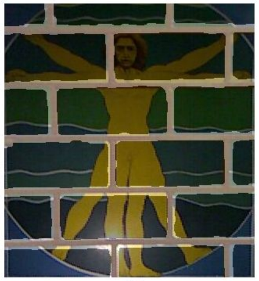
d. Inpainting of fence region

3. Experiments and results
4. Comparison with other methods
5. Limitations and precautions
6. “Future” work

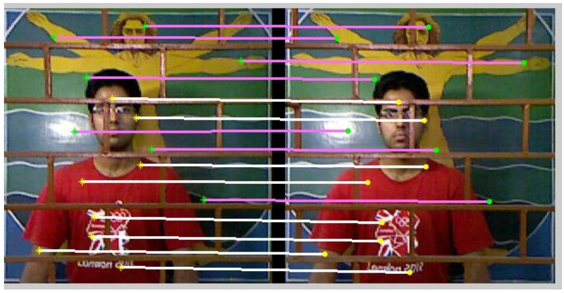
# 2. Our approach

Capture of multiple RGB-D images → Fence detection and alignment → Computation of pixel shift → Inpainting

with slight shift in camera position
(choose one to de-fence)



using affine transformation



using ASIFT

using LBP

# 2. Our approach - Capture of multiple RGB-D images

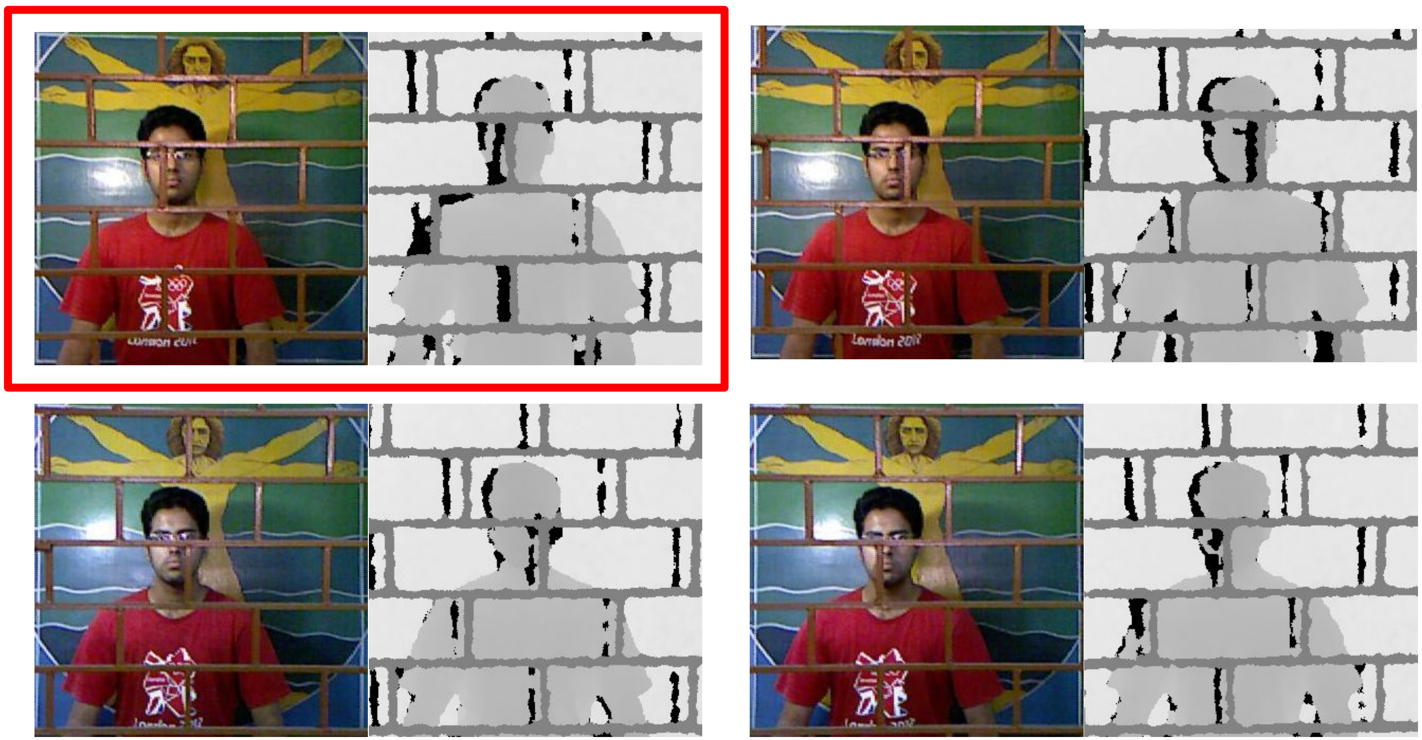
## Capture of multiple RGB-D images

Fence detection and alignment

Computation of pixel shift

Inpainting

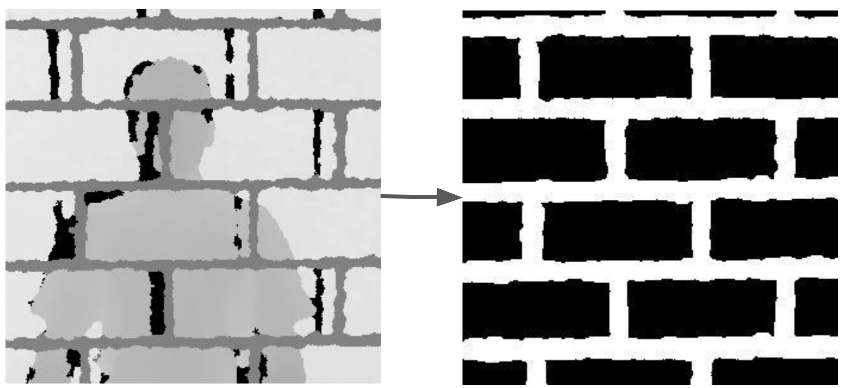
- Multiple images of the same scene are captured with slight positional shift in the camera.
- We choose one image to de-fence, and use the rest to borrow information.
- Objects occluded in one image are revealed in another:



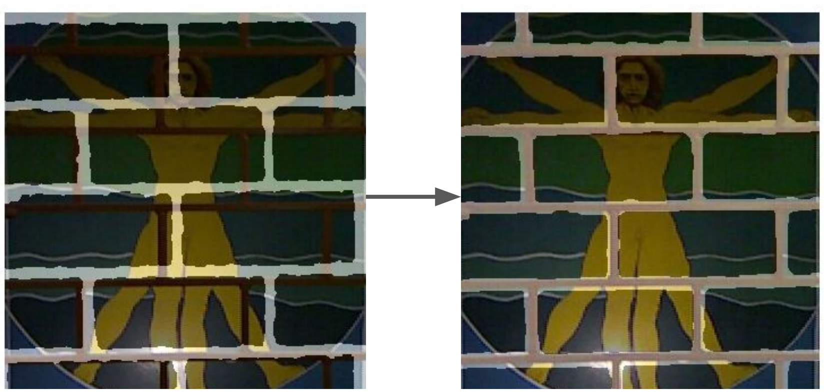
# 2. Our approach - Fence detection and alignment

Capture of multiple RGB-D images ⇨ **Fence detection and alignment** ⇨ Computation of pixel shift ⇨ Inpainting

## Fence detection



## Fence alignment



Fence is detected by filtering the depth map.

But detected fence does not coincide with fence in color image, due to spatial difference in RGB and D cameras.

Detected fence is aligned using homography.

<u>Special mention</u>: Cross-modal stereo by using Kinect - Wei-Chen Chiu, Ulf Blanke, and Mario Fritz

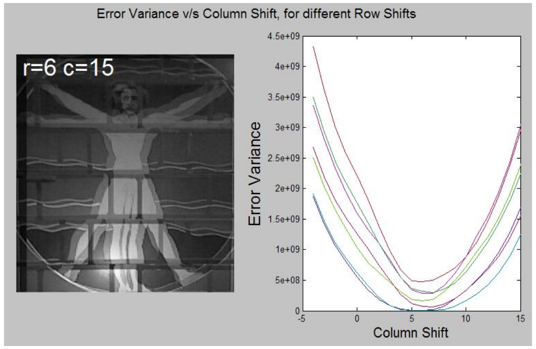
# 2. Our approach - Computation of pixel shift

Capture of multiple RGB-D images ⇨ Fence detection and alignment ⇨ **Computation of pixel shift** ⇨ Inpainting

- To borrow information of an occluded pixel from another image, the pixel shift between the images is needed

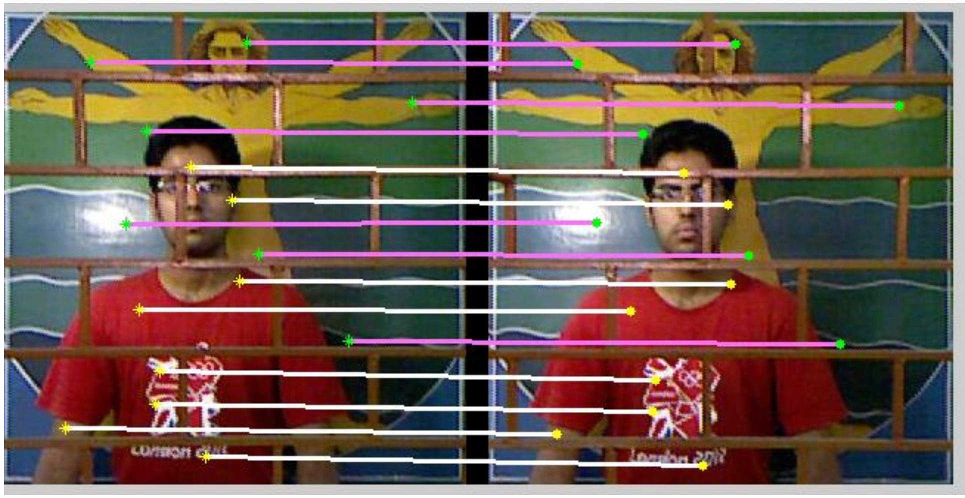
**Assumption**: Global shift

1) Naive way - computationally intensive




2) Optical flow - cannot handle occlusions

3) **ASIFT**[5]: [online demo]
   - State-of-the-art (at the time)
   - Can handle affine variations




[5] J. M. Morel and G.Yu, "ASIFT: A new framework for fully affine invariant image comparison," SIAM Journal on Imaging Sciences, vol. 2, no. 2, pp. 438–469, 2009.

# 2. Our approach - Inpainting

Capture of multiple RGB-D images ⇨ Fence detection and alignment ⇨ Computation of pixel shift ⇨ **Inpainting**

- De-fenced image is modelled as a Markov Random Field
- **Degradation Model**: $y_m = O_m . W_m . x + n_m$

  $y_m$ is the observed image,
  $x$ is the de-fenced image,
  $W_m$ is the warp matrix: it describes the Pixel Shift,
  $O_m$ is the occlusion matrix: it crops the non-fence regions from *x*,
  $n_m$ is noise, assumed to be Gaussian.

- **Objective function**:

$$\mathbf{J}(\mathbf{x}) = ||\mathbf{y}_\mathrm{m} - \mathbf{O}_\mathrm{m}.\mathbf{W}_\mathrm{m}.\mathbf{x}||^2 + \beta \sum_{c \in C} \mathbf{V}_\mathrm{c}(\mathbf{x})$$

  $V_c(x)$ is the clique potential function:

$$V_c(x) = |x_{i,j} - x_{i-1,j}| + |x_{i,j} - x_{i+1,j}| + |x_{i,j} - x_{i,j-1}| + |x_{i,j} - x_{i,j+1}|$$

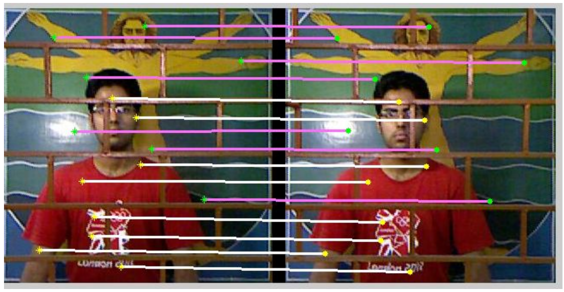
- $\mathbf{J}(\mathbf{x})$ is optimized to obtain MAP estimate $x$ using **Loopy Belief Propagation**

# 2. Our approach - recap

Capture of multiple RGB-D images → Fence detection and alignment → Computation of pixel shift → Inpainting

with slight shift in camera position
(choose one to de-fence)

using affine transformation

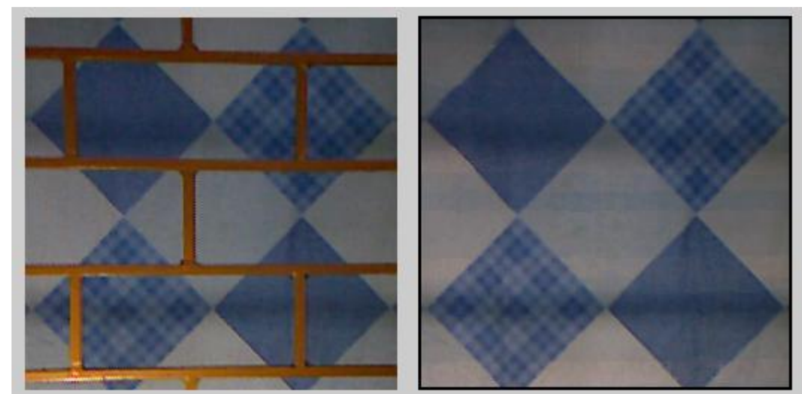


using ASIFT

using LBP

# 3. Experiments and results

- Rectangular fence and single object, with varying material of object

RMSE ≈ 3.1; pSNR ≈ 38dB; SSIM ≈ 0.87

# 3. Experiments and results

- Rectangular and diagonal fences with two objects:

Using two pixel shifts - one for the person, and one for the background.

RMSE ≈ 3.9

pSNR ≈ 36dB

SSIM ≈ 0.79

1. Introduction
2. Our approach
    a. Capture of multiple images
    b. Fence detection
    c. Computation of shifts between images
    d. Inpainting of fence region
3. Experiments and results

# 4. Comparison with other methods

5. Limitations and precautions
6. “Future” work

# 4. Comparison with other methods

Using Total Variation Inpainting using Split Bregman[5]

- Uses single image to inpaint, fails for thick fences

Using Learning-Based Digital Matting[6] on multiple images

- Uses single pixel shift, requires heavy user input, cannot handle multiple planes of depth

Using our approach

- Can handle thick fences and multiple planes of depth

[5] Getreuer, P., "Total variation inpainting using Split Bregman," *Image Processing On Line*, 2:147-157, 2012.

[6] [14] Khasare, V., and Sahay, R., "Seeing through the fence: image de-fencing using a video sequence," *2013 IEEE International Conference on Image Processing*, Restoration and Enhancement III.

1. Introduction
2. Our approach
    a. Capture of multiple images
    b. Fence detection
    c. Computation of shifts between images
    d. Inpainting of fence region
3. Experiments and results
4. Comparison with other methods

# 5. Limitations and precautions

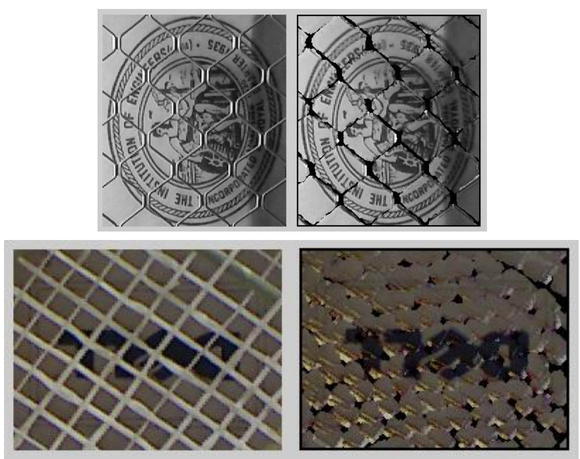
6. “Future” work

# 5. Limitations and precautions

- Sufficient indoor illumination - Kinect fails to record depth in outdoor settings

Colour image    Depth image

- Sufficient camera translation - to obtain full information



- Sufficient fence dilation - to avoid errors in depth map


- Limited camera translation - to avoid perspective distortion
  - If camera translation is too high, inpainting leads to a blurred image

1. Introduction
2. Our approach
    a. Capture of multiple images
    b. Fence detection
    c. Computation of shifts between images
    d. Inpainting of fence region
3. Experiments and results
4. Comparison with other methods
5. Limitations and precautions

# 6. “Future” work

# 6. “Future” work

<u>Some improvements</u> -

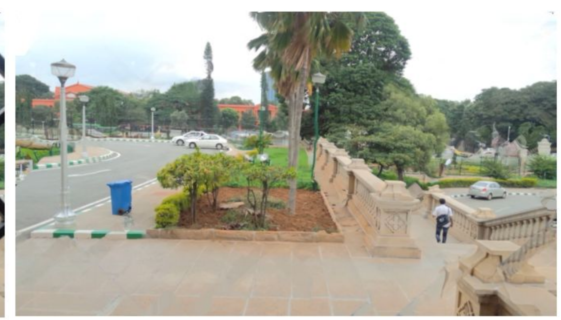
- perform depth completion by fusing data from multiple recorded depth maps affected by occlusions

Conference Paper - S. Jonna, V. S. Voleti, R. R. Sahay, and M. S. Kankanhalli, “A multimodal approach for image de-fencing and depth inpainting,” in Proc. Int. Conf. Advances in Pattern Recognition, 2015, pp. 1-6.

<u>More work</u> -

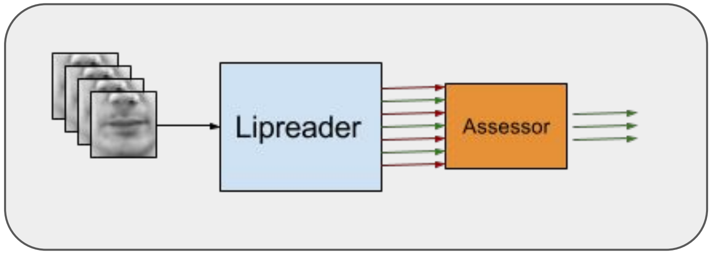
- made a dataset of captured occluded and unoccluded images
- proposed a robust algorithm to estimate optical flow under known fences/occlusions
- used the total variation of the unoccluded image and the completed depth map as regularization constraints
- used Split-Bregman iterative method to simultaneously de-fence and inpaint

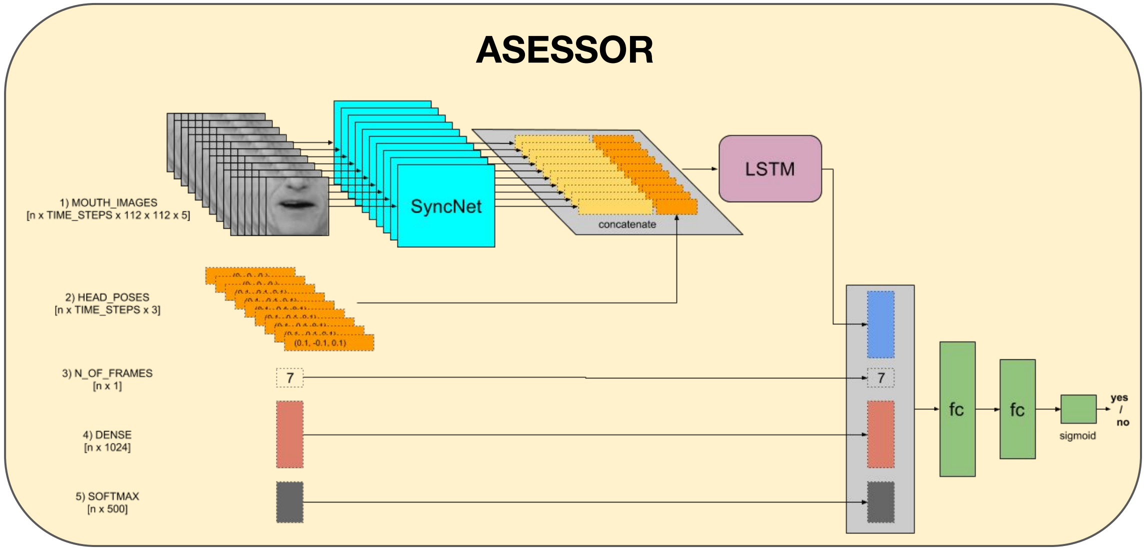
Journal paper - S. Jonna, S. Satapathy, V. S. Voleti, R. R. Sahay, “Unveiling the scene - A Multimodal Framework for Simultaneous Image Disocclusion and Depth Map Completion using Computational Cameras”

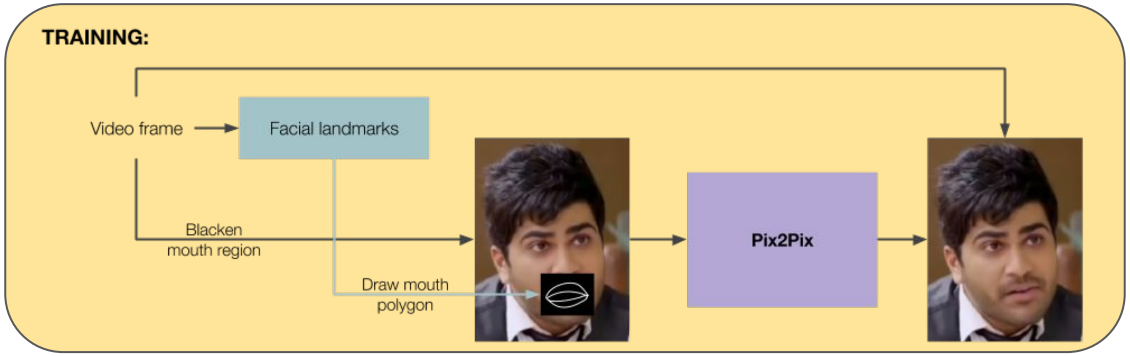
# @IIIT Hyderabad, India

- Using an assessor to filter video retrieval results




- Movie translation with lip-sync

Thank you.

voletiv.github.io

# APPENDIX

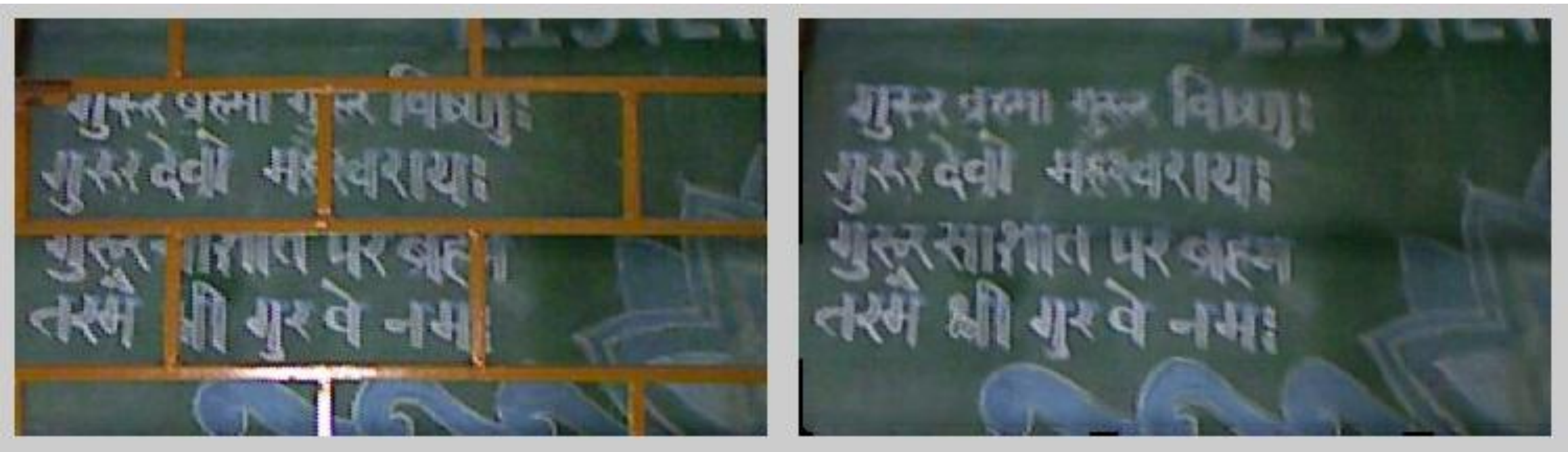
# Computation of Pixel Shifts

Superimposition of Reference Image and Shifted Test Image

Since the superimposed image is not too blurry, we can conclude that the pixel shifts have been calculated accurately.

# Experiments and results

## Rectangular fence with single object: Board

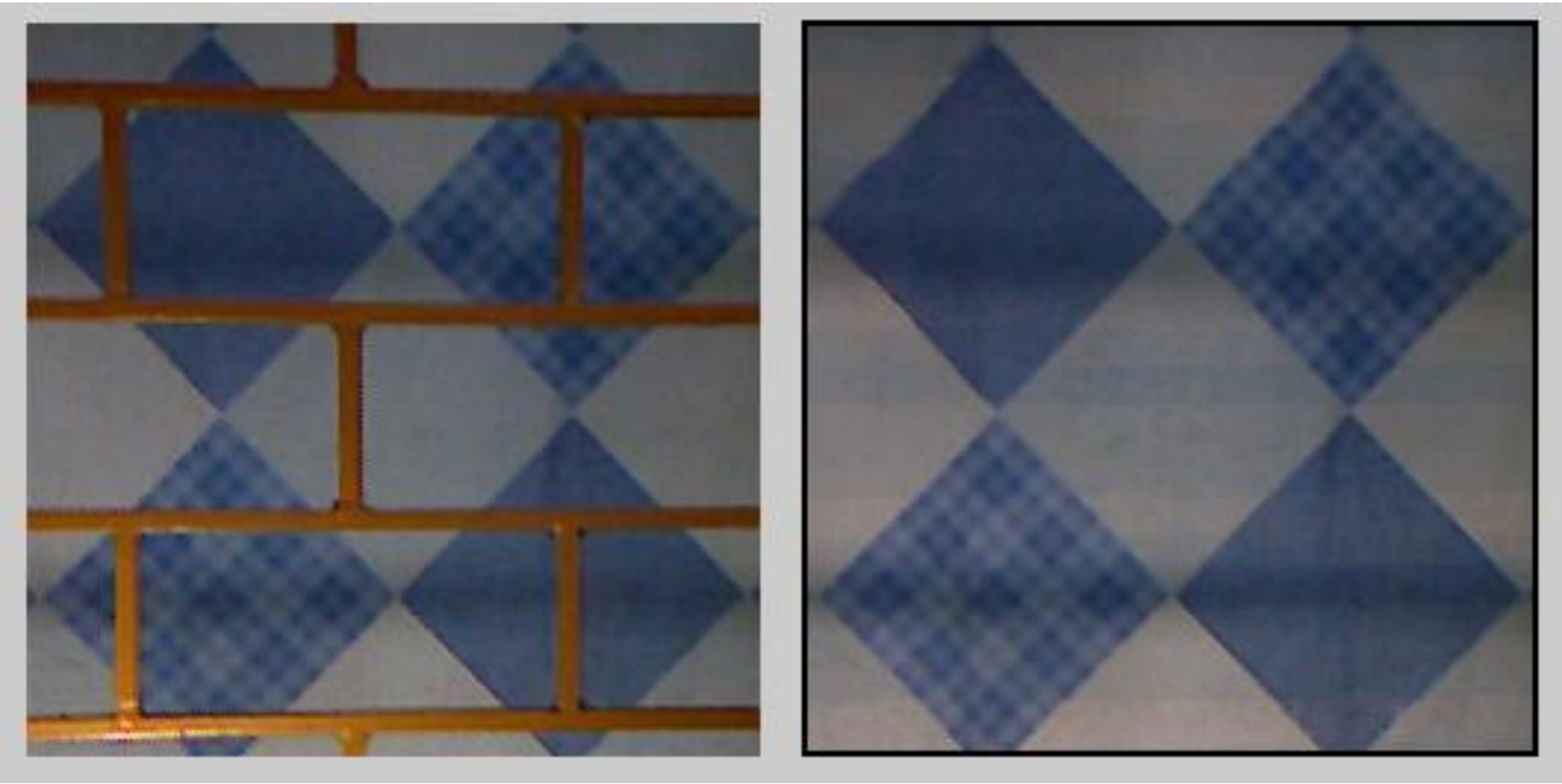


# Experiments and results

## Rectangular fence with single object: Poster

# Experiments and results

## Rectangular fence with single object: Bedsheet



# Experiments and results

## Rectangular fence with single object: Painting

# Experiments and results

## Rectangular fence with two objects

# Experiments and results

## Diagonal fence with two objects

# Comparison

## Total Variation Inpainting using Split Bregman

Using Total Variation Inpainting using Split Bregman

Using Proposed Algorithm

# Comparison

## De-fencing using Learning-Based Matting

Using De-fencing using Learning-Based Matting

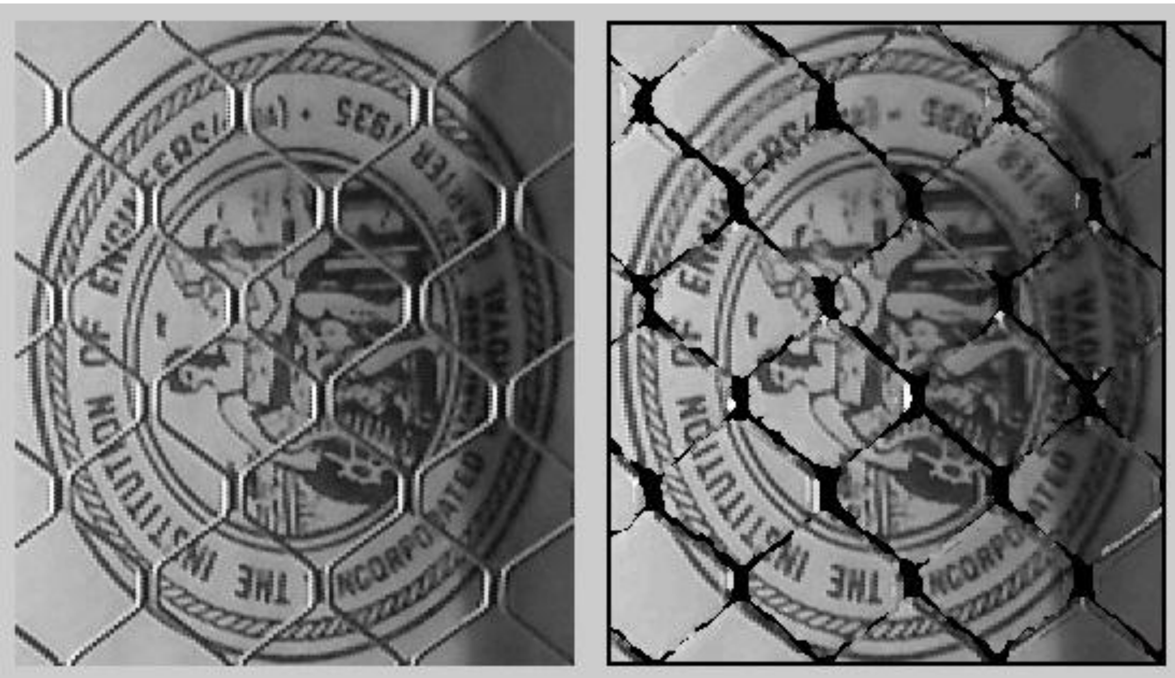
Using Proposed Algorithm

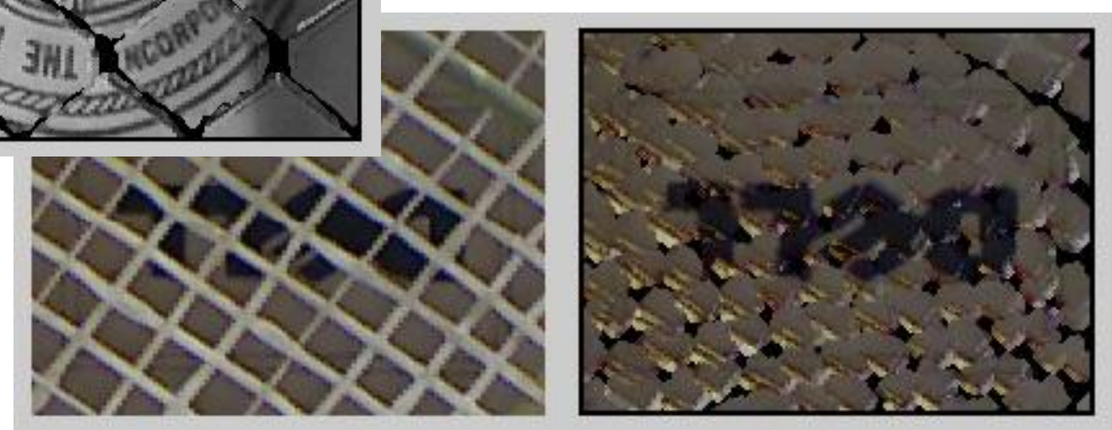
# Precautions

## Sufficient fence dilation


# Precautions

## Sufficient camera translation





# Precautions

## Limited camera translation