Blind Image Deblurring Based on Sparse Representation and Structural Self-Similarity

Jing Yu[1],Zhenchun Chang[2],Chuangbai Xiao[1],Weidong Sun[2]

[1]Beijing University of Technology,[2]Tsinghua University

Motion Blur

Motion Blur has been one of the most common artifacts in digital imaging.

Motion Blur

Motion Blur has been one of the most common artifacts in digital imaging.

Motion Blur

Motion Blur has been one of the most common artifacts in digital imaging.

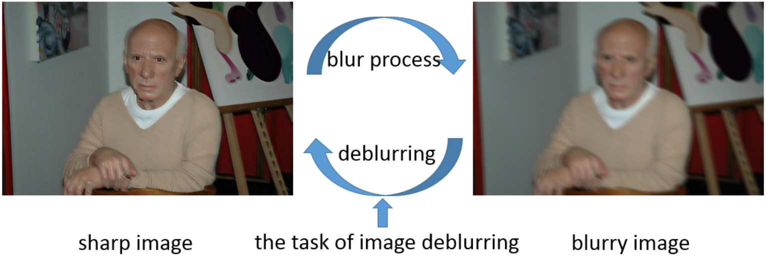
Motion Blur

Motion Blur has been one of the most common artifacts in digital imaging.

More recently, deblurring has received renewed attention due to the emerging need for removing motion blur in images captured by mobile phones.

Deblurring

Deblurring attempts to reconstruct or recover a blurry image by modeling the degradation and applying the inverse process.

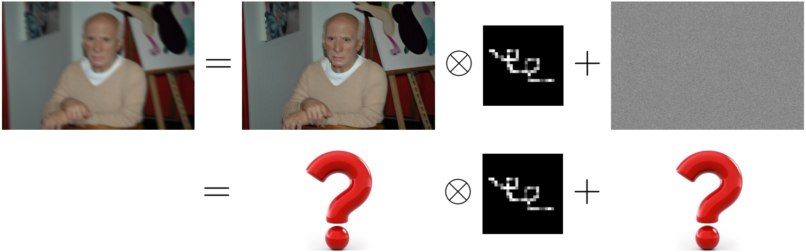


Blur Model

If the blur is uniform across the image,the blur can be expressed as a convolution of a sharp image and a blur kernel.

Blur Model

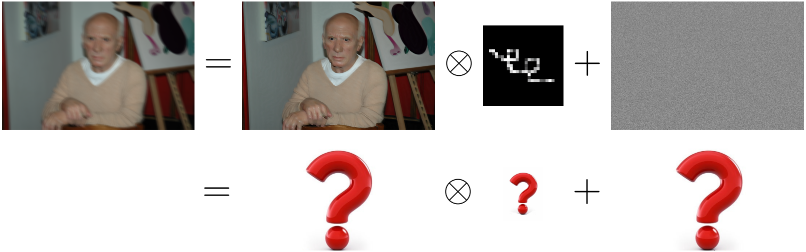
If the blur is uniform across the image,the blur can be expressed as a convolution of a sharp image and a blur kernel.



- the blur kernel is known: **non blind deconvolution**, only need to estimate the sharp image

Blur Model

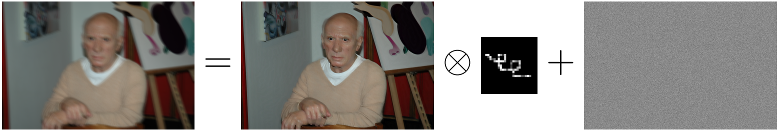
If the blur is uniform across the image,the blur can be expressed as a convolution of a sharp image and a blur kernel.



- the blur kernel is known: **non blind deconvolution**, only need to estimate the sharp image
- the blur kernel is unknown: **blind deconvolution**, need to estimate both the sharp image and blur kernel

Challenge:ill-posed

In blind deconvolution, the number of unknowns is larger than the number of constraints and the problem is ill posed.



Challenge:ill-posed

In blind deconvolution, the number of unknowns is larger than the number of constraints and the problem is ill posed.

Challenge:ill-posed

In blind deconvolution, the number of unknowns is larger than the number of constraints and the problem is ill posed.

Challenge:ill-posed

In blind deconvolution, the number of unknowns is larger than the number of constraints and the problem is ill posed.

Additional priors about the sharp image or the blur kernel are needed.

Structural Self-Similarity Prior

Multi-scale structural self-similarity refers to that similar image structures, both within the same scale and across different scales, frequently recur in natural images explicitly or implicitly.

Structural Self-Similarity Prior

Multi-scale structural self-similarity refers to that similar image structures, both within the same scale and across different scales, frequently recur in natural images explicitly or implicitly.

sharp image($\boldsymbol{X}$)

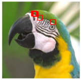
sharp patch($\mathbf{Q}\boldsymbol{X}$)

- $\boldsymbol{X}$: the vector form of sharp image
- $\mathbf{Q}\boldsymbol{X}$: patch extracted from sharp image$\boldsymbol{X}$

Structural Self-Similarity Prior

Multi-scale structural self-similarity refers to that similar image structures, both within the same scale and across different scales, frequently recur in natural images explicitly or implicitly.

sharp image($\boldsymbol{X}$)



down-sampled image($\boldsymbol{X}^{\alpha}$)

sharp patch($\mathbf{Q}\boldsymbol{X}$)

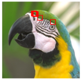
similar patches($\mathbf{R}_i\boldsymbol{X}^{\alpha}$)

- $\boldsymbol{X}$: the vector form of sharp image
- $\mathbf{Q}\boldsymbol{X}$: patch extracted from sharp image $\boldsymbol{X}$
- $\boldsymbol{X}^{\boldsymbol{\alpha}}$: the vector form of down-sampled sharp image
- $\mathbf{R}_i\boldsymbol{X}^{\boldsymbol{\alpha}}$: similar patches extracted from $\boldsymbol{X}^{\alpha}$ compared with $\mathbf{Q}\boldsymbol{X}$

Structural Self-Similarity Prior

Multi-scale structural self-similarity refers to that similar image structures, both within the same scale and across different scales, frequently recur in natural images explicitly or implicitly.

sharp image($\boldsymbol{X}$)



down-sampled image($\boldsymbol{X}^{\alpha}$)

sharp patch($\mathbf{Q}\boldsymbol{X}$)

similar patches($\mathbf{R}_i\boldsymbol{X}^{\alpha}$)

- $\boldsymbol{X}$: the vector form of sharp image
- $\mathbf{Q}\boldsymbol{X}$: patch extracted from sharp image$\boldsymbol{X}$
- $\boldsymbol{X}^{\boldsymbol{\alpha}}$: the vector form of down-sampled sharp image
- $\mathbf{R}_i\boldsymbol{X}^{\boldsymbol{\alpha}}$: similar patches extracted from $\boldsymbol{X}^{\alpha}$ compared with $\mathbf{Q}\boldsymbol{X}$

The linear combination of the L most similar patches of $\mathbf{Q}\boldsymbol{X}$ (put into the set $\mathcal{S}$) can be used to predict $\mathbf{Q}\boldsymbol{X}$:

$$\mathbf{Q}\boldsymbol{X} \approx \sum_{i\in\mathcal{S}} w_i \mathbf{R}_i \boldsymbol{X}^{\alpha}$$

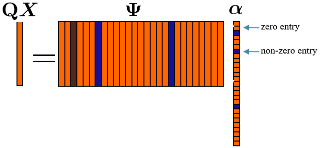
$$w_i = \frac{\exp(-\|\mathbf{Q}\boldsymbol{X} - \mathbf{R}_i\boldsymbol{X}^{\alpha}\|_2^2/h)}{\sum_{l\in\mathcal{S}} \exp(-\|\mathbf{Q}\boldsymbol{X} - \mathbf{R}_l\boldsymbol{X}^{\alpha}\|_2^2/h)}$$

Sparse Representation Prior

Image patches can always be represented well as a sparse linear combination of atoms in an appropriate dictionary.

$$\mathbf{Q}\boldsymbol{X} = \Psi\boldsymbol{\alpha} \quad ||\boldsymbol{\alpha}||_0 \ll n$$

- $\mathbf{Q}\boldsymbol{X} \in \mathbb{R}^n$:the vector form of patch extracted from image $\boldsymbol{X}$
- $\boldsymbol{\Psi} = [\boldsymbol{\psi_1}, \cdots, \boldsymbol{\psi_t}] \in \mathbb{R}^{n \times t}$:dictionary.$\psi_i$ is called the atom of the dictionary
- $\boldsymbol{\alpha} = [\alpha_1, \cdots, \alpha_t]^T \in \mathbb{R}^t$:the representation coefficient of $\mathbf{Q}\boldsymbol{X}$

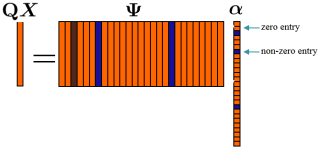


Sparse Representation Prior

Image patches can always be represented well as a sparse linear combination of atoms in an appropriate dictionary.

$$\mathbf{Q}\boldsymbol{X} = \Psi\boldsymbol{\alpha} \quad ||\boldsymbol{\alpha}||_0 \ll n$$

- $\mathbf{Q}\boldsymbol{X} \in \mathbb{R}^n$:the vector form of patch extracted from image $\boldsymbol{X}$
- $\boldsymbol{\Psi} = [\boldsymbol{\psi_1}, \cdots, \boldsymbol{\psi_t}] \in \mathbb{R}^{n\times t}$:dictionary.$\psi_i$ is called the atom of the dictionary
- $\boldsymbol{\alpha} = [\alpha_1, \cdots, \alpha_t]^T \in \mathbb{R}^t$:the representation coefficient of $\mathbf{Q}\boldsymbol{X}$



The choice of the dictionary:

- prespecified transform matrix(e.g.wavelets,curvelets):simple,poor adaptability
- learn the dictionary from training data(e.g.K-SVD):good adaptability,widely used;need to choose good training data

Proposed dictionary learning method

The key issue of sparse representation is to identify an appropriate dictionary.

Proposed dictionary learning method

The key issue of sparse representation is to identify an appropriate dictionary.

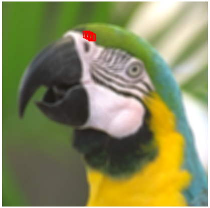
- **Use database consisting of enormous images as training data**:the database needs to provide patches similar to the patches in the sharp image;inefficient

Proposed dictionary learning method

The key issue of sparse representation is to identify an appropriate dictionary.

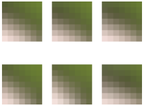
- **Use database consisting of enormous images as training data**:the database needs to provide patches similar to the patches in the sharp image;inefficient
- **Use blurry image as training data**:the blurry image is not quite similar with the sharp image

sharp image



blurry image

sharp patch



similar patches in blurry image

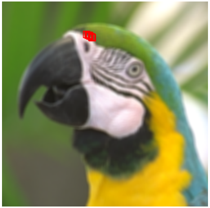
Proposed dictionary learning method

The key issue of sparse representation is to identify an appropriate dictionary.

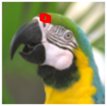
- **Use database consisting of enormous images as training data**:the database needs to provide patches similar to the patches in the sharp image;inefficient
- **Use blurry image as training data**:the blurry image is not quite similar with the sharp image

We use an over-complete dictionary trained on down-sampled blurry patches to help exploit the sparse prior of sharp patches.

sharp image

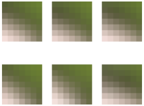


blurry image



down-sampled blurry image

sharp patch



similar patches in blurry image

similar patches in down-sampled blurry image

Objective function

By taking the down-sampled blurry image as dictionary training data and incorporating both sparse representation and structural self-similarity as regularization constraints, we get the following joint minimization problem of both image and blur kernel

$$\min_{\boldsymbol{x},\boldsymbol{h}} \Big\{ \overbrace{\|\nabla \boldsymbol{y} - \nabla \boldsymbol{x} \otimes \boldsymbol{h}\|_2^2}^{\text{observation model}} + \overbrace{\lambda_c \sum_j \|\mathbf{Q}_j \boldsymbol{X} - \Psi \boldsymbol{\alpha}_j\|_2^2}^{\text{sparsity}} + \overbrace{\lambda_s \sum_j \|\mathbf{Q}_j \boldsymbol{X} - \sum_{i \in S_j} w_i^j \mathbf{R}_i \boldsymbol{X}^{\alpha}\|_2^2}^{\text{structural self-similarity}}$$

$$+ \underbrace{\lambda_g \|\nabla \boldsymbol{x}\|_2^2}_{\text{smoothness}} + \underbrace{\lambda_h \|\boldsymbol{h}\|_2^2}_{\text{blur kernel}} \Big\} \quad \text{s.t. } \forall j \ \|\boldsymbol{\alpha}_j\|_0 \leq T$$

- $\boldsymbol{y}$:blurry image
- $\boldsymbol{x}$:sharp image
- $\boldsymbol{h}$:blur kernel
- $\nabla = \{\partial_x, \partial_y\}$:the spatial derivative operator in two directions
- λ_c,λ_s,λ_g,λ_h:regularization weights

Optimization

$$\min_{\boldsymbol{x},\boldsymbol{h}} \Big\{ ||\nabla \boldsymbol{y} - \nabla \boldsymbol{x} \otimes \boldsymbol{h}||_2^2 + \lambda_c \sum_j ||\mathbf{Q}_j \boldsymbol{X} - \Psi \boldsymbol{\alpha}_j||_2^2 + \lambda_s \sum_j ||\mathbf{Q}_j \boldsymbol{X} - \sum_{i \in S_j} w_i^j \mathbf{R}_i \boldsymbol{X}^{\alpha}||_2^2 + \lambda_g ||\nabla \boldsymbol{x}||_2^2 + \lambda_h ||\boldsymbol{h}||_2^2 \Big\} \quad \text{s.t. } \forall j \ ||\boldsymbol{\alpha}_j||_0 \le T$$

$$\min_{\boldsymbol{x},\boldsymbol{h}} \Big\{ ||\nabla \boldsymbol{y} - \nabla \boldsymbol{x} \otimes \boldsymbol{h}||_2^2 + \lambda_c \sum_j ||\mathbf{Q}_j \boldsymbol{X} - \Psi \boldsymbol{\alpha}_j||_2^2 + \lambda_s \sum_j ||\mathbf{Q}_j \boldsymbol{X} - \sum_{i \in S_j} w_i^j \mathbf{R}_i \boldsymbol{X}^\alpha||_2^2 + \lambda_g ||\nabla \boldsymbol{x}||_2^2 + \lambda_h ||\boldsymbol{h}||_2^2 \Big\} \quad \text{s.t. } \forall j \ ||\boldsymbol{\alpha}_j||_0 \le T$$

The above objective function is non-convex.We take an iterative process to solve this problem that alternately optimizes the motion blur kernel and the latent image.

- fix $\hat{\boldsymbol{x}}_k$,update $\hat{\boldsymbol{h}}_k$
- fix $\hat{\boldsymbol{h}}_k$,update $\hat{\boldsymbol{x}}_{k+1}$

Initialization:$k = 0$,$\hat{\boldsymbol{x}}_0 = \boldsymbol{y}$

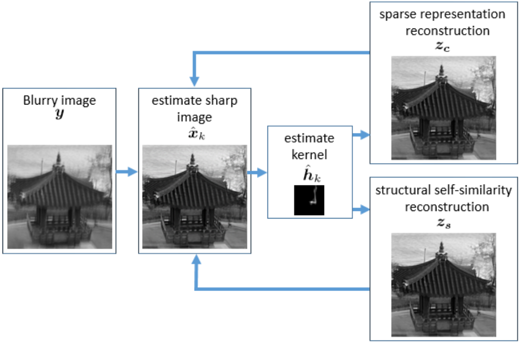
Optimization

$$\min_{\boldsymbol{x},\boldsymbol{h}} \Big\{ ||\nabla \boldsymbol{y} - \nabla \boldsymbol{x} \otimes \boldsymbol{h}||_2^2 + \lambda_c \sum_j ||\mathbf{Q}_j \boldsymbol{X} - \Psi \boldsymbol{\alpha}_j||_2^2 + \lambda_s \sum_j ||\mathbf{Q}_j \boldsymbol{X} - \sum_{i \in S_j} w_i^j \mathbf{R}_i \boldsymbol{X}^{\alpha}||_2^2$$
$$+ \lambda_g ||\nabla \boldsymbol{x}||_2^2 + \lambda_h ||\boldsymbol{h}||_2^2 \Big\} \quad \text{s.t. } \forall j \ ||\boldsymbol{\alpha}_j||_0 \leq T$$

The above objective function is non-convex.We take an iterative process to solve this problem that alternately optimizes the motion blur kernel and the latent image.

- fix $\hat{\boldsymbol{x}}_k$,update $\hat{\boldsymbol{h}}_k$
- fix $\hat{\boldsymbol{h}}_k$,update $\hat{\boldsymbol{x}}_{k+1}$

Initialization:$k = 0$,$\hat{\boldsymbol{x}}_0 = \boldsymbol{y}$



Quantitative Evaluation on Synthetic Dataset

We measure the quality of an estimated blur kernel using the error ratio measure r proposed by Levin et.al.

$$\mathrm{ER} = \frac{||\boldsymbol{x} - \hat{\boldsymbol{x}}_{\hat{\boldsymbol{h}}}||_2^2}{||\boldsymbol{x} - \hat{\boldsymbol{x}}_{\boldsymbol{h}}||_2^2}$$

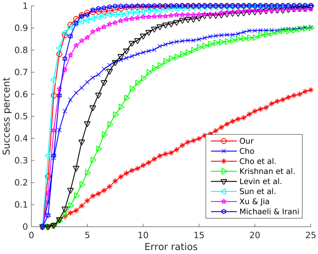
- $\boldsymbol{h}$: real kernel
- $\hat{\boldsymbol{h}}$: estimated kernel
- $\boldsymbol{x}$: real sharp image
- $\hat{\boldsymbol{x}}_{\boldsymbol{h}}$: recovered image with $\boldsymbol{h}$
- $\hat{\boldsymbol{x}}_{\hat{\boldsymbol{h}}}$: restored image with $\hat{\boldsymbol{h}}$

Quantitative Evaluation on Synthetic Dataset

We measure the quality of an estimated blur kernel using the error ratio measure r proposed by Levin et.al.

$$\mathrm{ER} = \frac{||\boldsymbol{x} - \hat{\boldsymbol{x}}_{\hat{\boldsymbol{h}}}||_2^2}{||\boldsymbol{x} - \hat{\boldsymbol{x}}_{\boldsymbol{h}}||_2^2}$$

- $\boldsymbol{h}$: real kernel
- $\hat{\boldsymbol{h}}$: estimated kernel
- $\boldsymbol{x}$: real sharp image
- $\hat{\boldsymbol{x}}_{\boldsymbol{h}}$: recovered image with $\boldsymbol{h}$
- $\hat{\boldsymbol{x}}_{\hat{\boldsymbol{h}}}$: restored image with $\hat{\boldsymbol{h}}$



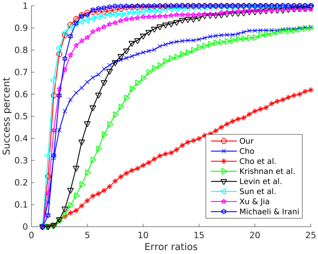
Cumulative error ratio over 640 large natural images provided by Sun[2013]

Quantitative Evaluation on Synthetic Dataset

We measure the quality of an estimated blur kernel using the error ratio measure r proposed by Levin et.al.

$$\mathrm{ER} = \frac{||\boldsymbol{x} - \hat{\boldsymbol{x}}_{\hat{\boldsymbol{h}}}||_2^2}{||\boldsymbol{x} - \hat{\boldsymbol{x}}_{\boldsymbol{h}}||_2^2}$$

- $\boldsymbol{h}$: real kernel
- $\hat{\boldsymbol{h}}$: estimated kernel
- $\boldsymbol{x}$: real sharp image
- $\hat{\boldsymbol{x}}_{\boldsymbol{h}}$: recovered image with $\boldsymbol{h}$
- $\hat{\boldsymbol{x}}_{\hat{\boldsymbol{h}}}$: restored image with $\hat{\boldsymbol{h}}$



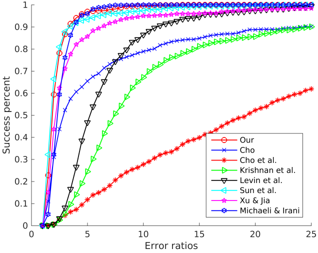
Cumulative error ratio over 640 large natural images provided by Sun[2013]

It is empirically observed by Michaeli & Irani that the deblurring results are still visually pleasing for error ratios $r \leq 5$,thus if $r \leq 5$,the blind deconvolution is regarded successful.

We measure the quality of an estimated blur kernel using the error ratio measure r proposed by Levin et.al.

$$\mathrm{ER} = \frac{||\boldsymbol{x} - \hat{\boldsymbol{x}}_{\hat{\boldsymbol{h}}}||_2^2}{||\boldsymbol{x} - \hat{\boldsymbol{x}}_{\boldsymbol{h}}||_2^2}$$

- $\boldsymbol{h}$: real kernel
- $\hat{\boldsymbol{h}}$: estimated kernel
- $\boldsymbol{x}$: real sharp image
- $\hat{\boldsymbol{x}}_{\boldsymbol{h}}$: recovered image with $\boldsymbol{h}$
- $\hat{\boldsymbol{x}}_{\hat{\boldsymbol{h}}}$: restored image with $\hat{\boldsymbol{h}}$

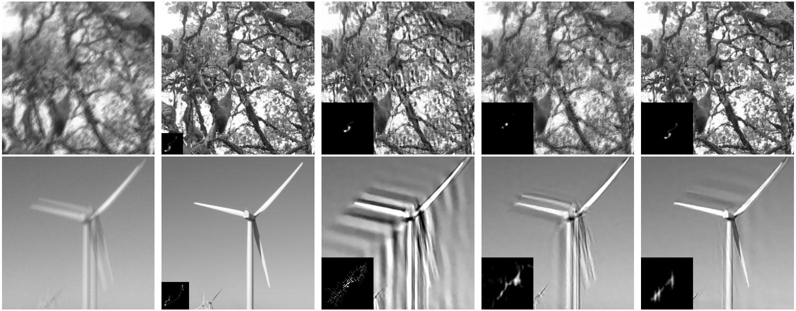


Cumulative error ratio over 640 large natural images provided by Sun[2013]

It is empirically observed by Michaeli & Irani that the deblurring results are still visually pleasing for error ratios $r \leq 5$,thus if $r \leq 5$,the blind deconvolution is regarded successful.

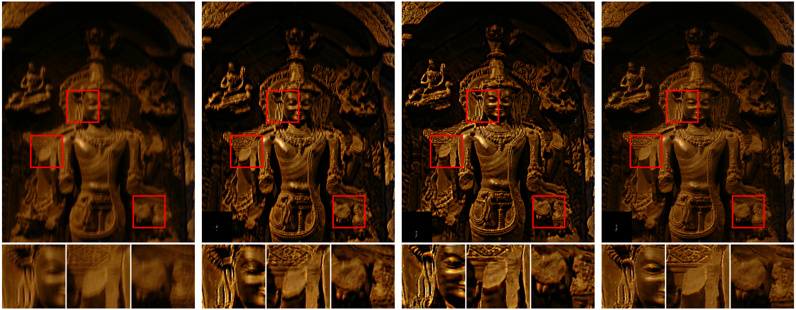
	success rate%	mean error ratio
Ours	96.88	2.2181
Michaeli & Irani[2014]	95.94	2.5662
Sun et al.[2013]	93.44	2.3764
Xu & Jia[2010]	85.63	3.6293
Levin et al.[2011]	46.72	6.5577
Cho & Lee[2009]	65.47	8.6901
Krishnan et al.[2011]	24.49	11.5212
Cho et al.[2011]	11.74	24.7020

Visual Comparison(Synthetic image)



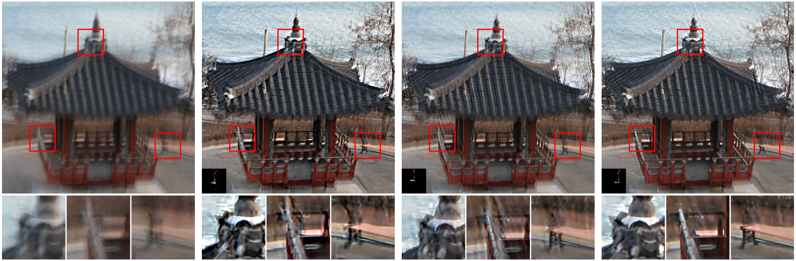
Blurry image	Real sharp image	Sun et al.[2013]	Michaeli & Irani[2014]	Ours

Visual Comparison(Real Image)



Blurry image — Krishnan et al.[2011] — Levin et al.[2011] — Ours

Visual Comparison(Real Image)



Blurry image | Perrone & Favaro [2014] | Michaeli & Irani[2014] | Ours

Thanks!