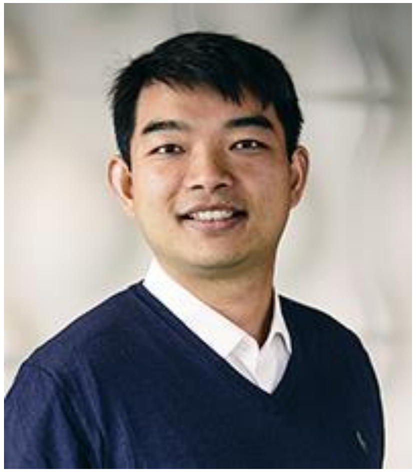
A Two-Step Approach to Hallucinating Faces: Global Parametric Model and Local Nonparametric Model

Ce Liu Heung-Yeung Shum Chang-Shui Zhang

By 吴昊东

AUTHOR

Ce Liu



Researcher

Microsoft Research New England

Research Affiliate

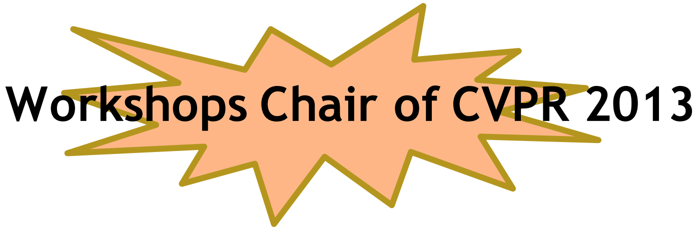
Computer Science and Artificial Intelligence Laboratory (CSAIL)

Massachusetts Institute of Technology

Adjunct Assistant Professor

Computer Science Department, Boston University

Homepage:
http://research.microsoft.com/en-us/um/people/celiu/ (Microsoft)
http:/people.csail.mit.edu/celiu/ (MIT)



FACE HALLUCINATION

- Face super-resolution (FSR)
- Face sketch-photo synthesis (FSPS) techniques

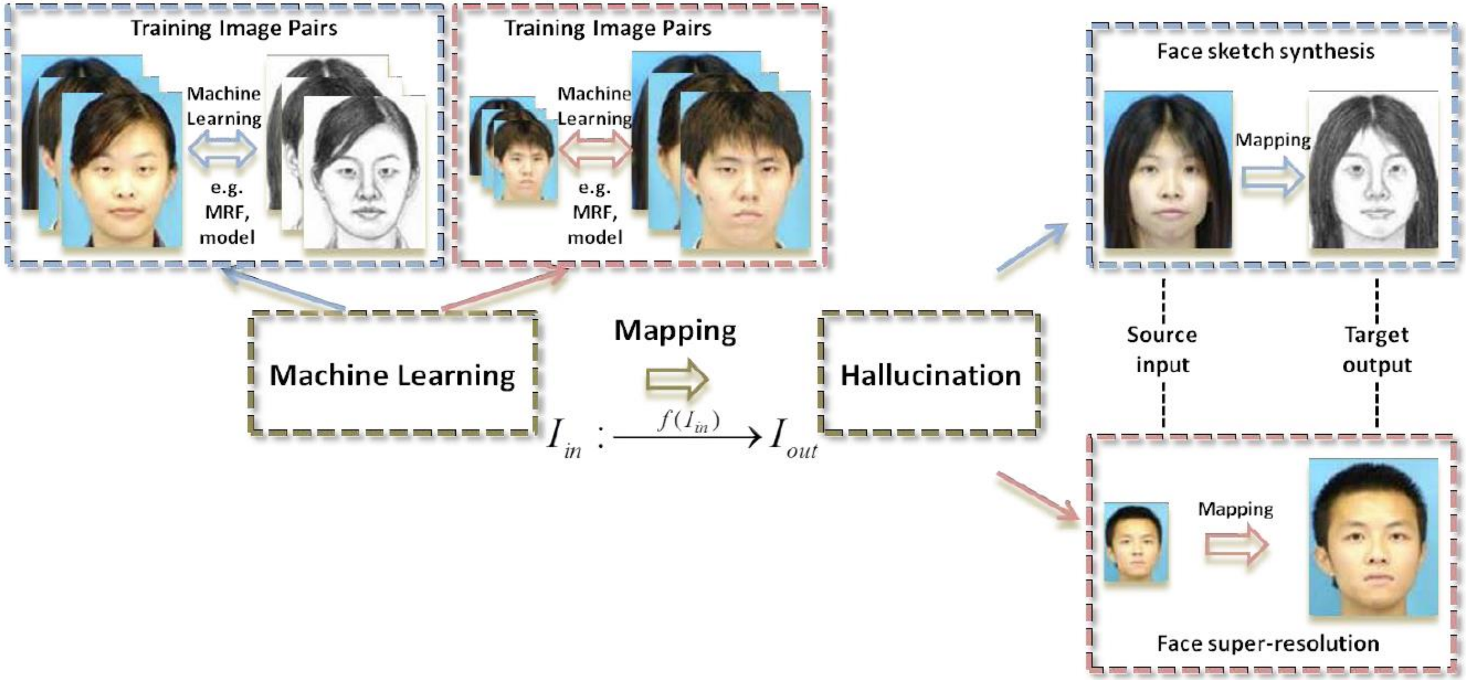


Fig. 1 Diagram of face hallucination

A Comprehensive Survey to Face Hallucination, IJCV 2013

FACE SUPER-RESOLUTION

- Reconstruction-based
  - Input images alone
- Learning-based
  - Learn by other HR & LR pairs of images (or only HR images)

CONSTRAINTS

- **Sanity constraint.** The result must be very close to the input image when smoothed and down-sampled.

- **Global constraint.** The result must have common characteristics of a human face, *e.g*. eyes, mouth and nose, symmetry, etc.

- **Local constraint.** The result must have specific characteristics of this face image with photorealistic local features.

CONSTRAINTS

- Sanity constraint
  - Easily satisfied
- Global constraint
  - Without -> noisy
- Local constraint
  - Without -> too smooth, close to average face

Two-Step Approach

- Title:
  - A **Two-Step** Approach to Hallucinating Faces: **Global** Parametric Model and **Local** Nonparametric Model

- Global + Local

Bayesian Formulation

- Smoothing and down-sampling
  - The same as averaging pooling*

$$I_L(m,n) = \frac{1}{s^2}\sum_{i=0}^{s-1}\sum_{j=0}^{s-1} I_H(sm+i, sn+j) \qquad (3)$$

- If I_L & I_H are vectors,

$$I_L = AI_H \qquad (4)$$

- Reverse it!

Bayesian Formulation

- maximum a posteriori (MAP) criterion

$$p(I_H|I_L) = \frac{p(I_L|I_H)p(I_H)}{p(I_L)}.$$

$$I_H^* = \arg\max_{I_H} p(I_L|I_H)p(I_H). \tag{5}$$

Global and Local Modeling of Face

- Global + Local

$$I_H = I_H^l + I_H^g. \quad (6)$$

- Since I_L is the low-frequency part of I_H

$$AI_H^g = AI_H, \ AI_H^l = 0. \quad (7)$$

$$p(I_H) = p(I_H^l, I_H^g) = p(I_H^l | I_H^g) p(I_H^g). \quad (8)$$

- Regard $p(I_L|I_H)$ as soft constraint to I_H

$$p(I_L | I_H) = \frac{1}{Z} \exp\{-\|AI_H - I_L\|^2 / \lambda\}, \quad (9)$$

$$p(I_L | I_H) = \frac{1}{Z} \exp\{-\|AI_H^g - I_L\|^2 / \lambda\} = p(I_L | I_H^g). \quad (10)$$

Target

$$I_H^* = \arg\max_{I_H} p(I_L|I_H)p(I_H). \quad (5)$$

$$p(I_L|I_H) = \frac{1}{Z}\exp\{-\|AI_H^g - I_L\|^2/\lambda\} = p(I_L|I_H^g). \quad (10)$$

$$I_H^* = \arg\max_{I_H^g, I_H^l} p(I_L|I_H^g)p(I_H^g)p(I_H^l|I_H^g). \quad (11)$$

$$p(I_L|I_H^g)p(I_H^g) \Uparrow \longrightarrow I_H^{g*}$$

$$p(I_H^l|I_H^g) \Uparrow \longrightarrow I_H^{l*}$$

$$\longrightarrow \quad I_H^* = I_H^{g*} + I_H^{l*}$$

The Method

- Decouple high-resolution face image to two parts

 =  + 

I_H $\qquad$ I_H^g $\qquad$ I_H^l

I_H— high resolution face image $\quad$ I_H^g — global face $\quad$ I_H^l — local face

- Two-step Bayesian inference

I_L

?

$$I_H^* = \arg\max_{I_H} p(I_L \mid I_H)p(I_H)$$
$$= \arg\max_{I_H^g, I_H^l} p(I_L \mid I_H^g, I_H^l)p(I_H^g, I_H^l)$$
$$= \arg\max_{I_H^g, I_H^l} p(I_L \mid I_H^g)p(I_H^g)p(I_H^l \mid I_H^g)$$

1. Inferring global face

$$I_H^{g*} = \arg\max_{I_H^{g*}} p(I_L \mid I_H^g)p(I_H^g)$$

2. Inferring local face

$$I_H^{l*} = \arg\max_{I_H^{l*}} p(I_H^l \mid I_H^{g*})$$

Finally adding them together $I_H^* = I_H^{l*} + I_H^{g*}$

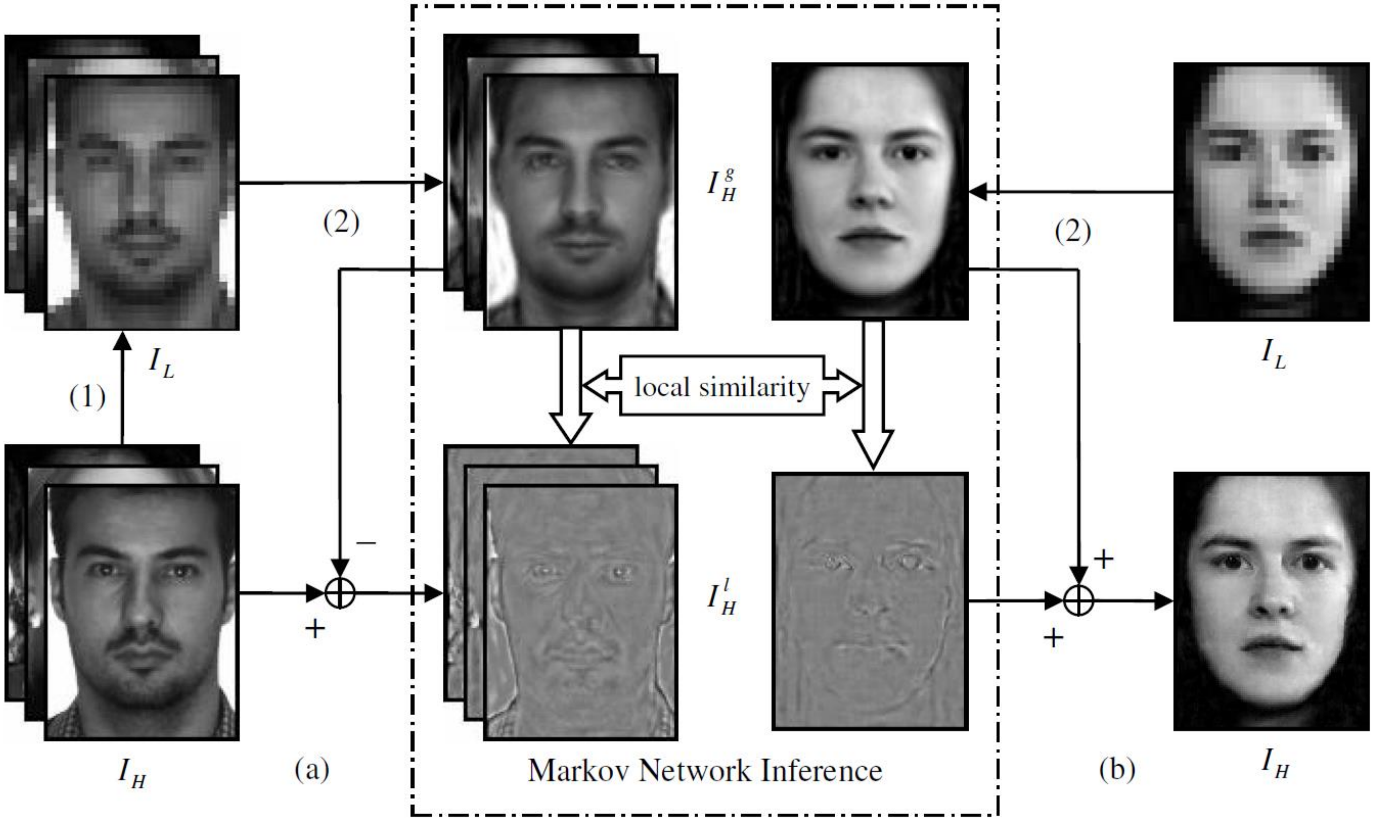


Figure 2. The function of Markov network in our model. (a) is the training process and (b) the hallucinating process. (1): smooth and down-sampling. (2): MAP inference to get the optimal global face I_H^{g*}. The Markov network finds the optimal local feature image I_H^{l*} by energy minimization.

Global Modeling

- Apply PCA to training face image $\{I_H^{(i)}\}_{i=1}^k$

$$I_H^g = BX + \mu, X = B^T(I_H - \mu), \qquad (12)$$

$$p(I_L|I_H^g)p(I_H^g) \longleftrightarrow p(I_L|X)p(X)$$

$$p(X) = \frac{1}{Z'}\exp\{-X^T\Lambda^{-1}X\}, \qquad (13)$$

$$p(I_L|X) = \frac{1}{Z}\exp\{-\|A(BX+\mu)-I_L\|^2/\lambda\}. \qquad (14)$$

Global Modeling

$$\text{maximize } p(I_L|X)p(X)$$

$$X^* = \arg\min_X \lambda X^T \Lambda^{-1} X + \|A(BX+\mu)-I_L\|^2, \qquad (15)$$

$$X^* = (B^T A^T A B + \lambda\Lambda^{-1})^{-1} B^T A^T (I_L - A\mu). \qquad (16)$$

$$I_H^{g*} = BX^* + \mu$$

- I_H^{g*} is very close to human face with some smoothness
- Calculate quickly using input I_L

Local Modeling

- Patch-based nonparametric Markov network
- Square patch size: $w \times w$ & h overlap
- $\vec{v}$ means (i, j)
- Δ_x means (1,0)
- Δ_y means (0,1)

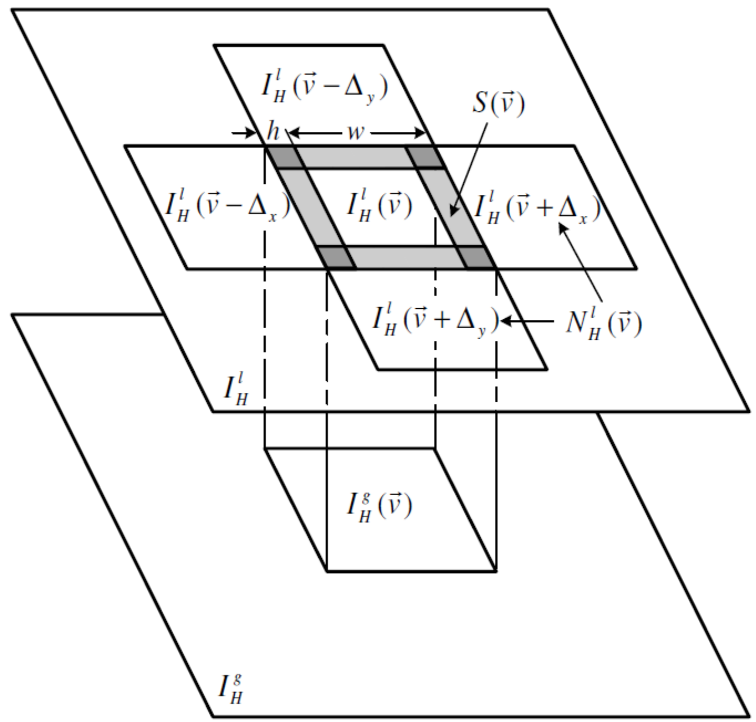


Figure 3. Illustration of the patch-based Markov network.

Local Modeling

- Assume that the above network is a Markov network

$$p(I_H^l(\vec{v})\,|\,I_H^{l-}(\vec{v}), I_H^g) = p(I_H^l(\vec{v})\,|\,N_H^l(\vec{v}), I_H^g(\vec{v})), \qquad (17)$$

- Suppose $p(I_H^l(\vec{v})\,|\,N_H^l(\vec{v}), I_H^g(\vec{v}))$ Gibbs distribution

$$p(I_H^l(\vec{v})\,|\,N_H^l(\vec{v}), I_H^g(\vec{v})) \propto \exp\{-E_G(I_H^l(\vec{v}), N_H^l(\vec{v}), I_H^g(\vec{v}))\} \quad (18)$$

where $E_G(\cdot)$ is the Gibbs potential function to describe how likely a patch $I_H^l(\vec{v})$ connects to $I_H^g(\vec{v})$ and is surrounded by $N_H^l(\vec{v})$.

Local Modeling

$$p(I_H^l(\vec{v})|N_H^l(\vec{v}), I_H^g(\vec{v})) \propto \exp\{-E_G(I_H^l(\vec{v}), N_H^l(\vec{v}), I_H^g(\vec{v}))\} \quad (18)$$

$N_H^l(\vec{v})$ and $I_H^g(\vec{v})$ independently.

$$\begin{aligned} & E_G(I_H^l(\vec{v}), N_H^l(\vec{v}), I_H^g(\vec{v})) \\ = \; & E_G^{int}(I_H^l(\vec{v}), N_H^l(\vec{v})) + E_G^{ext}(I_H^l(\vec{v}), I_H^g(\vec{v})) \\ \equiv \; & E_G^{int}(\vec{v}) + E_G^{ext}(\vec{v}) \end{aligned} \qquad (19)$$

Local Modeling

- $E_G^{ext}(\vec{v})$
- Training pairs $\{I_H^{l(i)}(\vec{v}), I_H^{g(i)}(\vec{v})\}_{i=1}^{k}$

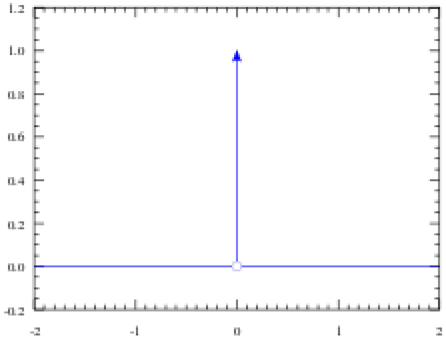
$$E_G^{ext}(\vec{v}) = \frac{1}{\lambda'} \sum_{i=1}^{k} \delta[I_H^l(\vec{v}) - I_H^{l(i)}(\vec{v})] d^2[I_H^g(\vec{v}), I_H^{g(i)}(\vec{v})], \quad (20)$$

$d(\cdot)$ is the distance metric between two patches in I_H^g

- Laplacian image L_H^g of I_H^g

$$d^2[I_H^g(\vec{v}), I_H^{g(i)}(\vec{v})] = \|L_H^g(\vec{v}) - L_H^{g(i)}(\vec{v})\|^2. \qquad (21)$$

$\delta(\cdot)$ is the dirac function



Local Modeling

- $E_G^{int}(\vec{v})$

$$E_G^{int}(\vec{v}) = \frac{1}{\lambda''} \sum_{\mathbf{u} \in S(\vec{v})} [\, I_H^l(\mathbf{u}) - N_H^l(\mathbf{u}) \,]^2, \qquad (22)$$

- Total Energy

$$E_{MN} = \sum_{\vec{v}} (E_G^{int}(\vec{v}) + E_G^{ext}(\vec{v})).$$

$p(I_H^l | I_H^{g*})$

$$I_H^{l*} = \arg\min_{I_H^l} E_{MN}. \qquad (24)$$

Simulated Annealing

MinimizeMarkovNetworkEnergy

Loop until $T < \epsilon$

Loop $\vec{v}$ *sequentially visiting all patches*

compute the energy of each patch $I_H^{l(i)}(\vec{v})$

set $I_H^l(\vec{v}) = I_H^{l(i)}(\vec{v})$ *with probability* (25)

decrease T

$I_H^{l*} = I_H^l$

$$exp\{-(E_G^{int}(\vec{v}) + E_G^{ext}(\vec{v}))/T\}, \qquad (25)$$

Experimental Results

- Database
  - FERET data set
  - AR data set
  - Other collections
- Align the face image

Experimental Results

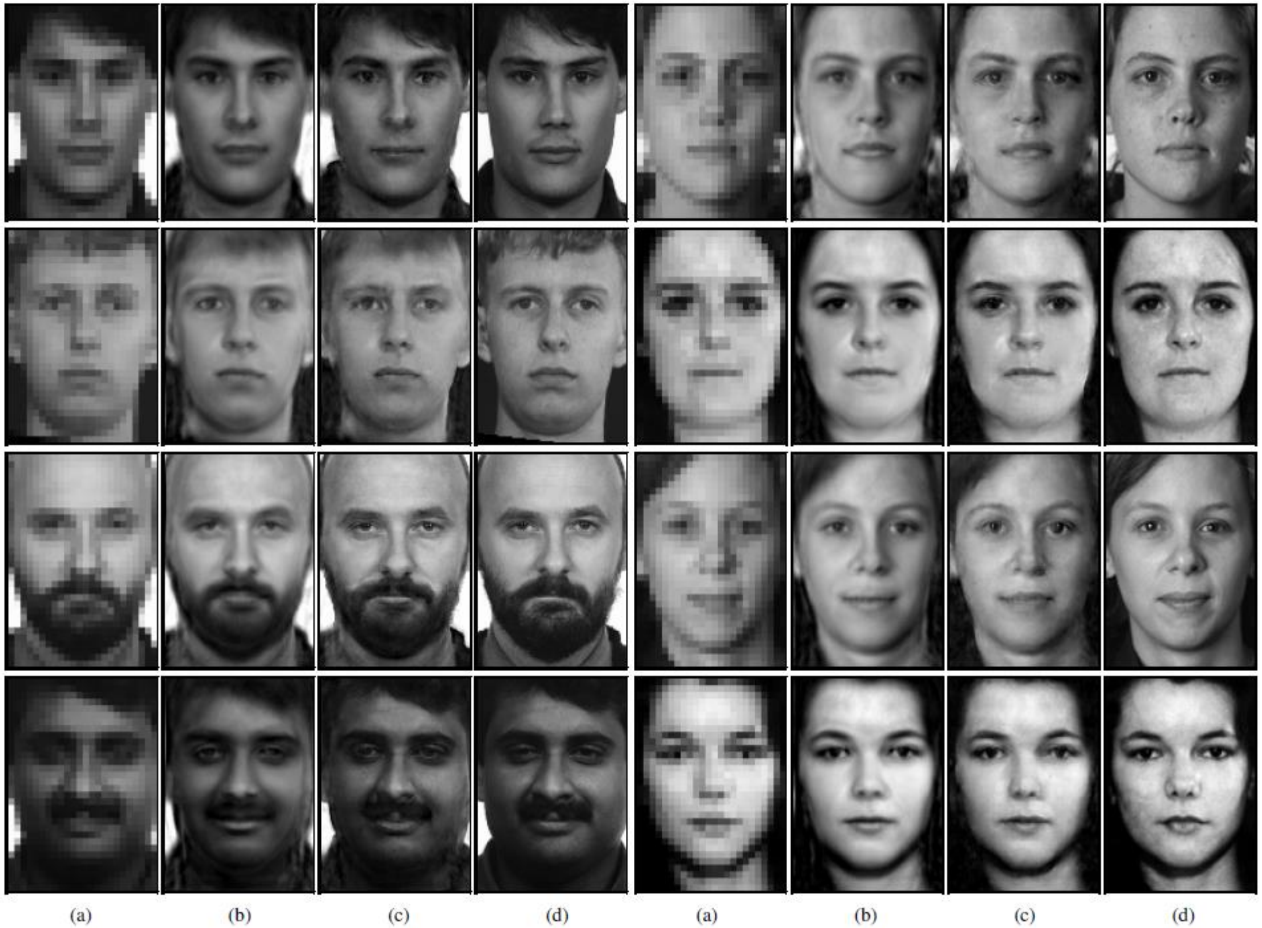


Figure 4. The hallucination results. (a) is the low-resolution 24×32 input. (b) is the inferred global face I_H^{g*} from (a). (c) is the final result $I_H^* = I_H^{g*} + I_H^{l*}$. I_H^{l*} is inferred from I_H^{g*} by Markov network. (d) is the original high-resolution 96×128 image.

Experimental Results

- Compete with other paper

(a) Input 24×32 (b) Our method (c) Cubic B-Spline (d) Hertzmann et al. (e) Baker et al. (f) Original 96×128

Figure 5. Comparison between our method and others.

Conclusion

- Combine global and local constrains
  - robustness and efficiency
- Why not complex model
  - the error can be compensated in the local model
- Patch-based nonparametric Markov network
  - Nonparametric -> accuracy
  - patch-based -> high efficiency
- Simulated Annealing
  - Similar to Gibbs sampling but converges more quickly
- Generalization -> Other super-resolution problem

The End

THANK YOU!

Q&A