VST Workshop on

Computer vision, graphics and Image processing

Super resolution of images and videos

Dr. Uma Mudenagudi

Professor,

Department of Electronics and Communication,

BVB College of Engineering and Technology, Hubli

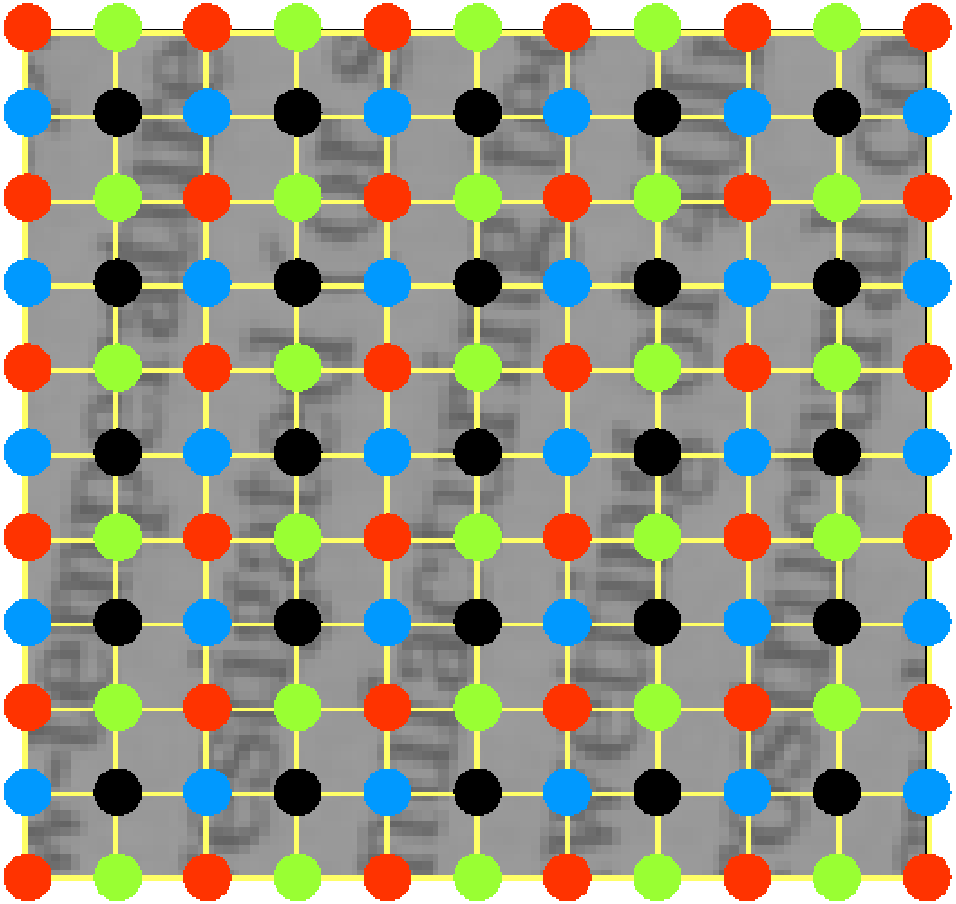
Class-14 15 22-06-2010 10.00-1.15pm

OUTLINE

- Introduction
- Motivation
- Space SR
- Spatio-temporal SR
- SR from images of 3D scene
- Conclusions

What is SR?

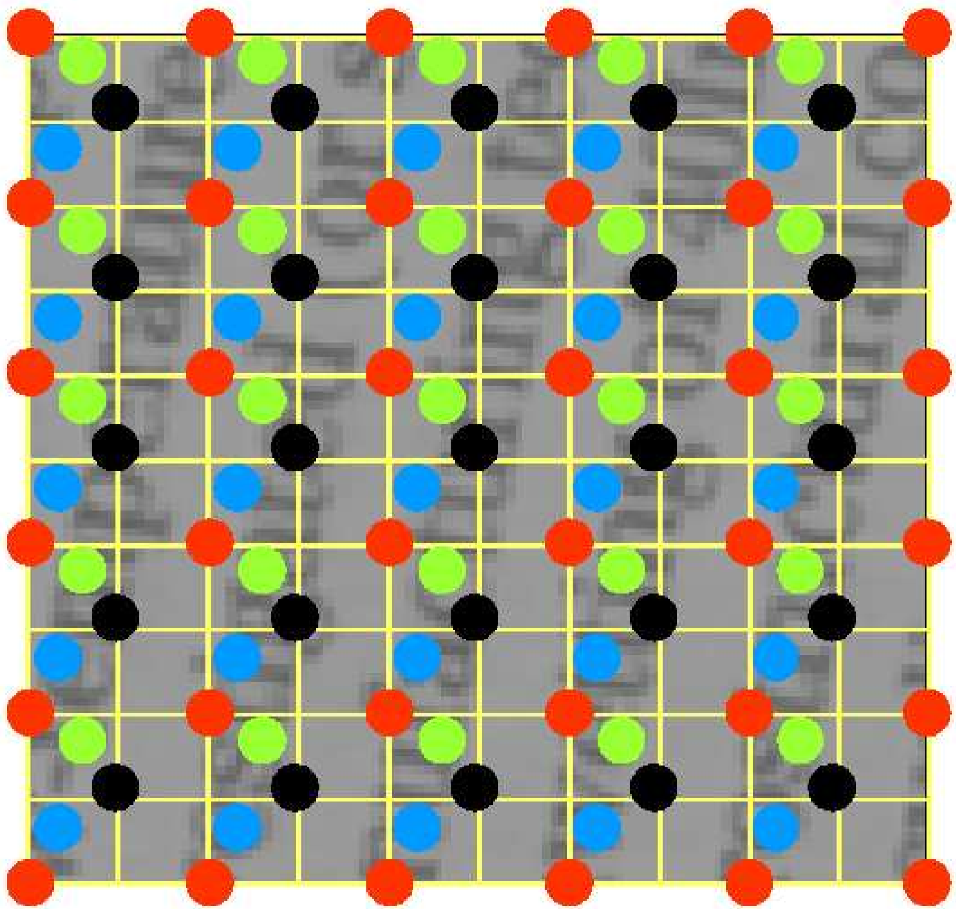
Spatial resolution - fuse four images



SR in its simplest form

What is SR?

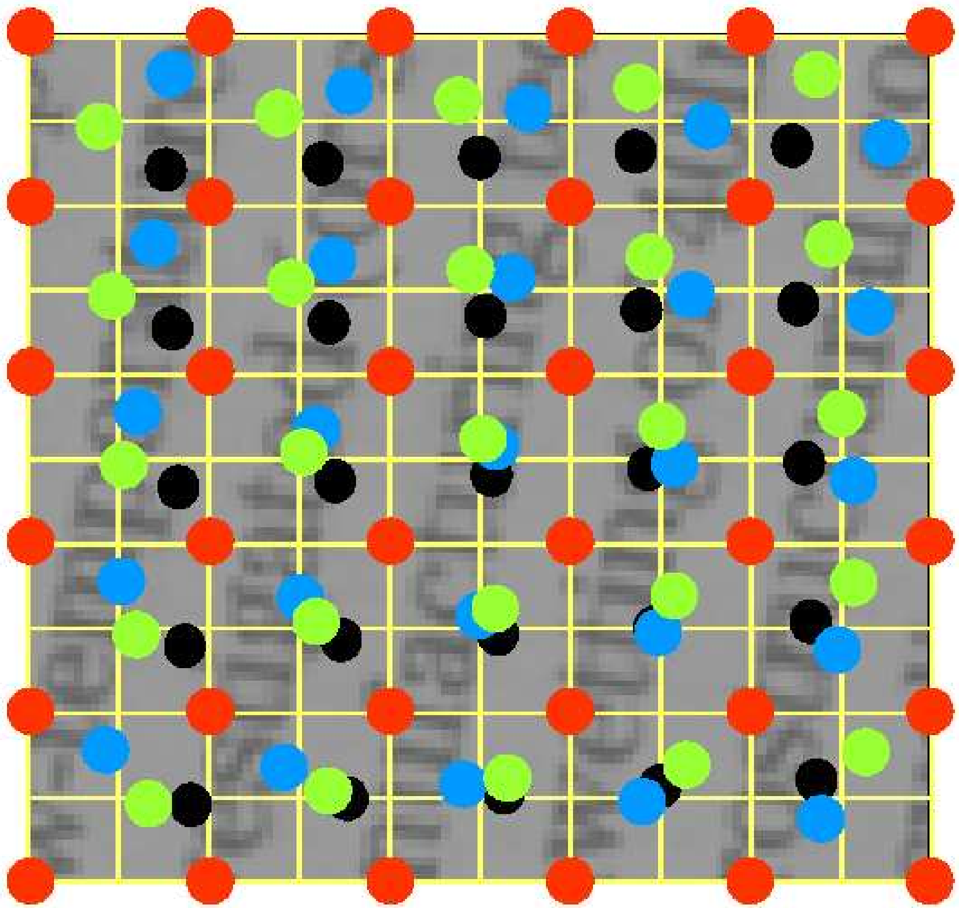
- Camera displacements are uncontrolled



SR with nonuniform sampling

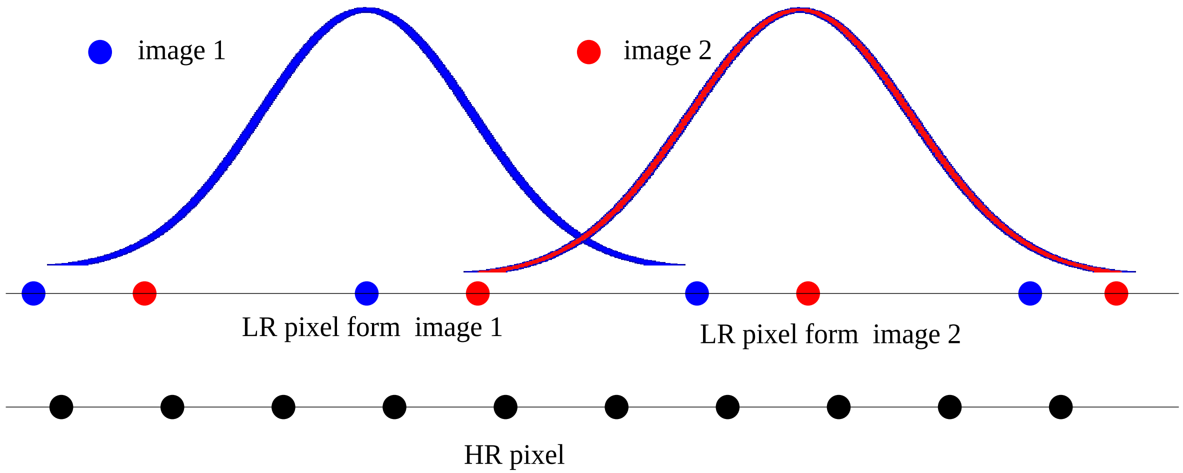
Why SR is difficult?

- Camera rotation and zoom



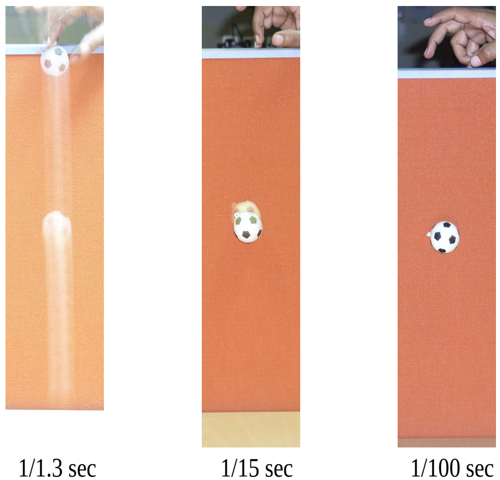
Requires accurate registration

More complications!

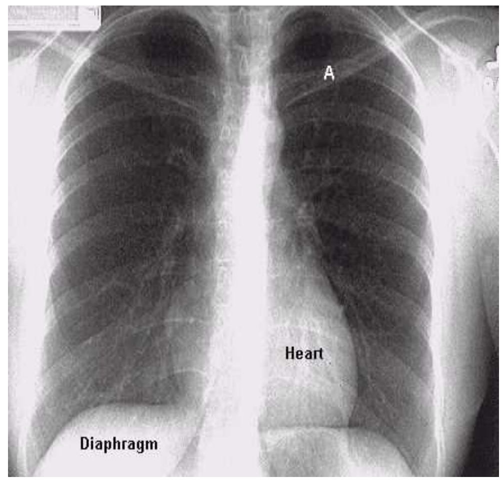


Sampling is not a point operation

Camera PSF



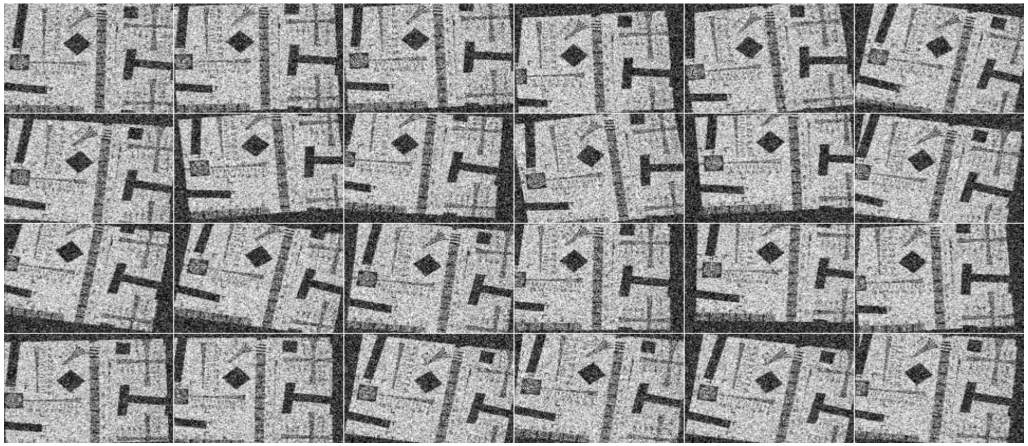
Motion blur



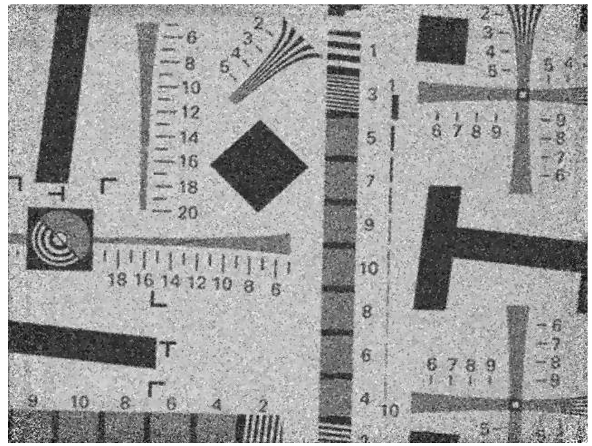
Noise

Spatial SR problem

- The super resolution process
  - given a set of noisy, blurred and transformed images and an estimated PSF, generate a high resolution image.



Noisy input images



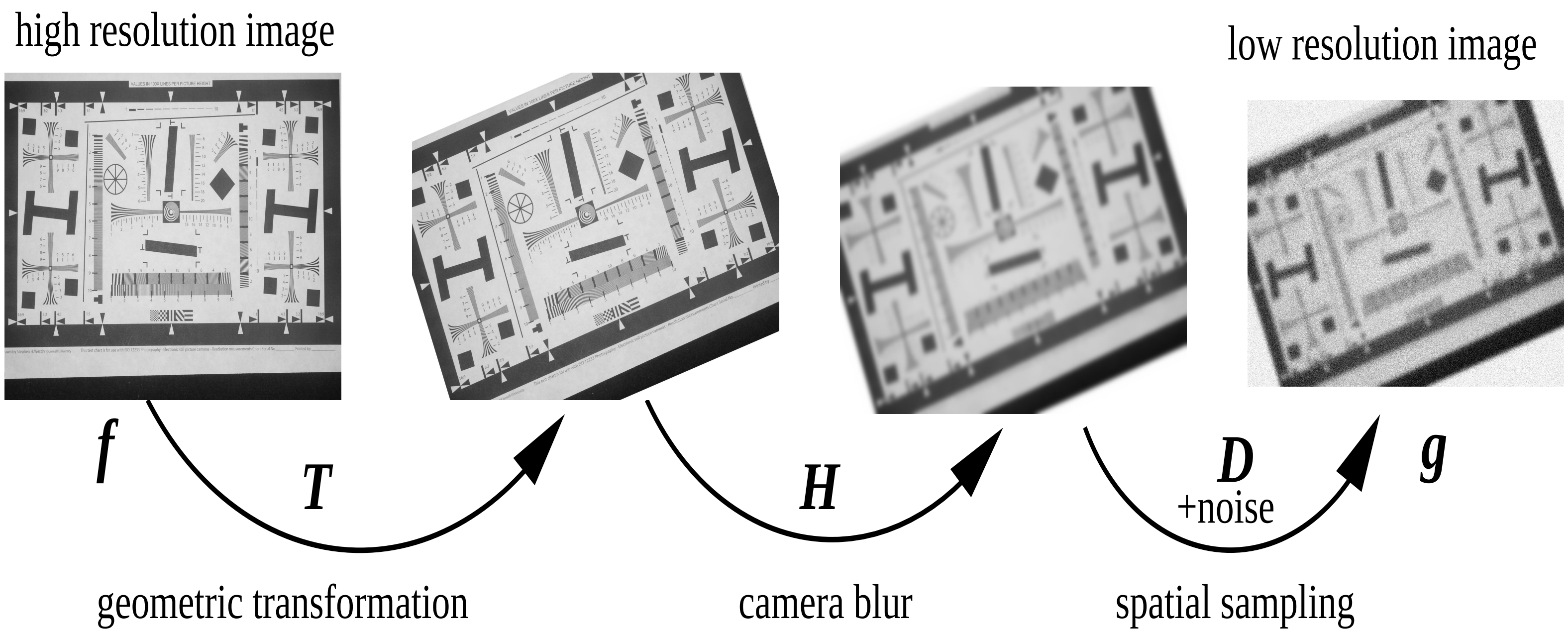
Super resolved image

What people have done?

- Motion based methods : use relative motion as the cue
  - frequency domain approach, nonuniform interpolation, DCT, wavelets etc.
- Motion free methods: use cues other than the relative motion like zoom, defocus blur and photometry.
- Learning based methods - no specific cue is used, priors are generic or learnt

Image formation model

$$\mathbf{g_k} = DH_kT_k\mathbf{f} + \eta_k \qquad 1 \leq k \leq n$$



Requires inversion process

Extensions

- High resolution expansion of single image: $\mathbf{g} = \mathbf{DHf} + \eta$
- Single image restoration and deblurring: $\mathbf{g} = \mathbf{Hf} + \eta$
- Single image noise removal: $\mathbf{g} = \mathbf{f} + \eta$

Registration

- Affine motion model
- Hierarchical model-based motion estimation by Bergen *et al.*

The PSF estimation

- Blur due to camera (lens and sensor) - no motion blur
- Assuming Gaussian PSF, the ESF for normalized edge is $s(x) = \frac{1}{2}(1 + erf(\frac{x}{\sigma\sqrt{2}}))$
- Estimate ESF by fitting equation to a normalized edge in least square sense
- PSF - approximately space invariant
- Estimates of σ - around 0.4 times the size of the LR pixel
- Gaussian PSF with 3×3 mask

Linear formulation

- Generative model: let $\mathbf{g}$ be the collection of all the low resolution pixels and $\mathbf{f}$ be the high resolution image then $\mathbf{g} = \mathbf{A}\mathbf{f}$
- 50×50 low resolution images, magnification = 4, number of unknowns = 40,000.
- Limits on SR
  - Baker and Kanade
  - Lin and Shum

Limits on SR (Baker and Kanade)

- For box PSF and registration as translation,

$$\mathbf{g}_i(m) = \sum_p \frac{\mathbf{f}(p)}{M^2} \int_p h_i \left(\frac{1}{M} z + \mathbf{c}_i - m \right) dz$$

  where m is a LR pixel, p is a HR pixel, M is a magnification factor, $\mathbf{c}_i$ is a translation (registration), $h_i(z)$ is the box PSF function ($1/S_i^2$ *iff* $z \in (-0.5S_i, 0.5S_i] \times (-0.5S_i, 0.5S_i]$, S_i is the width of the photosensitive area)
- Integration term is equivalent to model matrix $\mathbf{A}$
- ρ is a area of intersection of LR and HR pixel
  $\int_p h_i \left(\frac{1}{M} z + \mathbf{c}_i - m\right) dz = \frac{1}{S_i^2} \rho$

Limits on SR (Baker and Kanade)

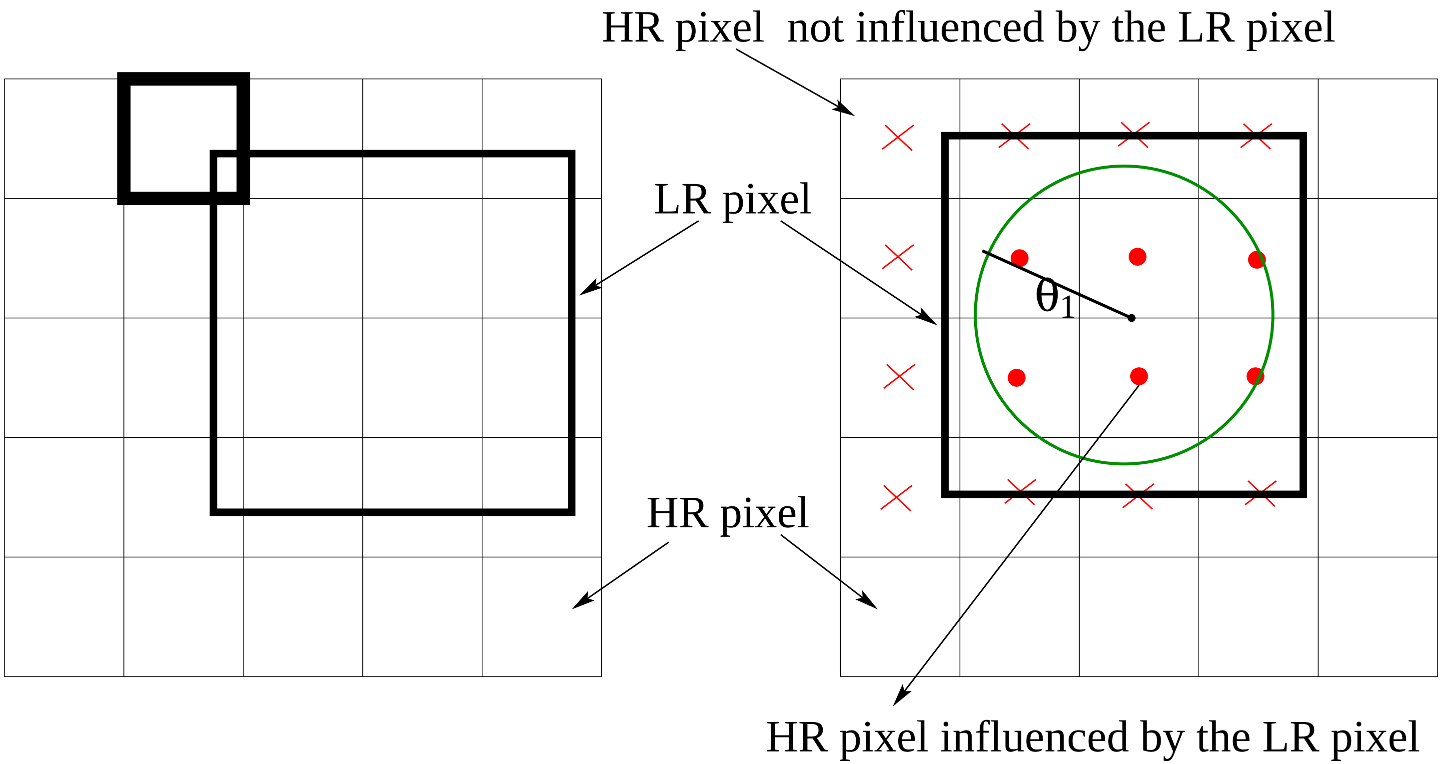
- The condition number of linear operator $\mathbf{A}$

$$Cond(A) = \frac{sup_{||x||_\infty=1}||Ax||_\infty}{inf_{||x||_\infty=1}||Ax||_\infty}$$

- If $\mathbf{f}(p) = 1 \forall p$ then $\mathbf{g}_i(m) = 1 \forall m$
- If $\mathbf{f}(p)$ is checker board then $|\mathbf{g}_i(m)| \leq 1/(M \cdot S_i)^2$
- Condition number is at least $(M \cdot S_i)^2$
- We show for an overlap of ρ the condition number increases by a fraction $1/$ρ
- For large condition number the inverse problem becomes ill-conditioned

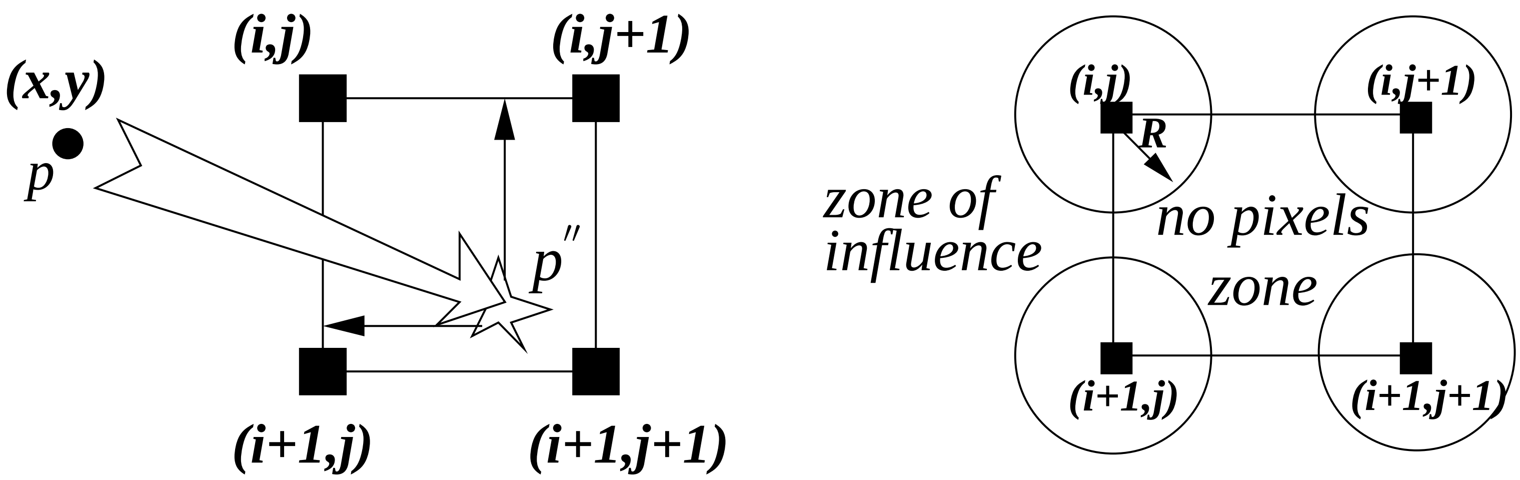
Well conditioning

- Effect of LR pixel



θ_1 = radius of the *zone of influence*

Well conditioning··



(a) (b)

- p is a HR pixel
- $p'' = DT_k p$, projection of p in k^{th} LR image
- $p' = R_k(p'')$, R_k is the rounding operator

Well conditioning··

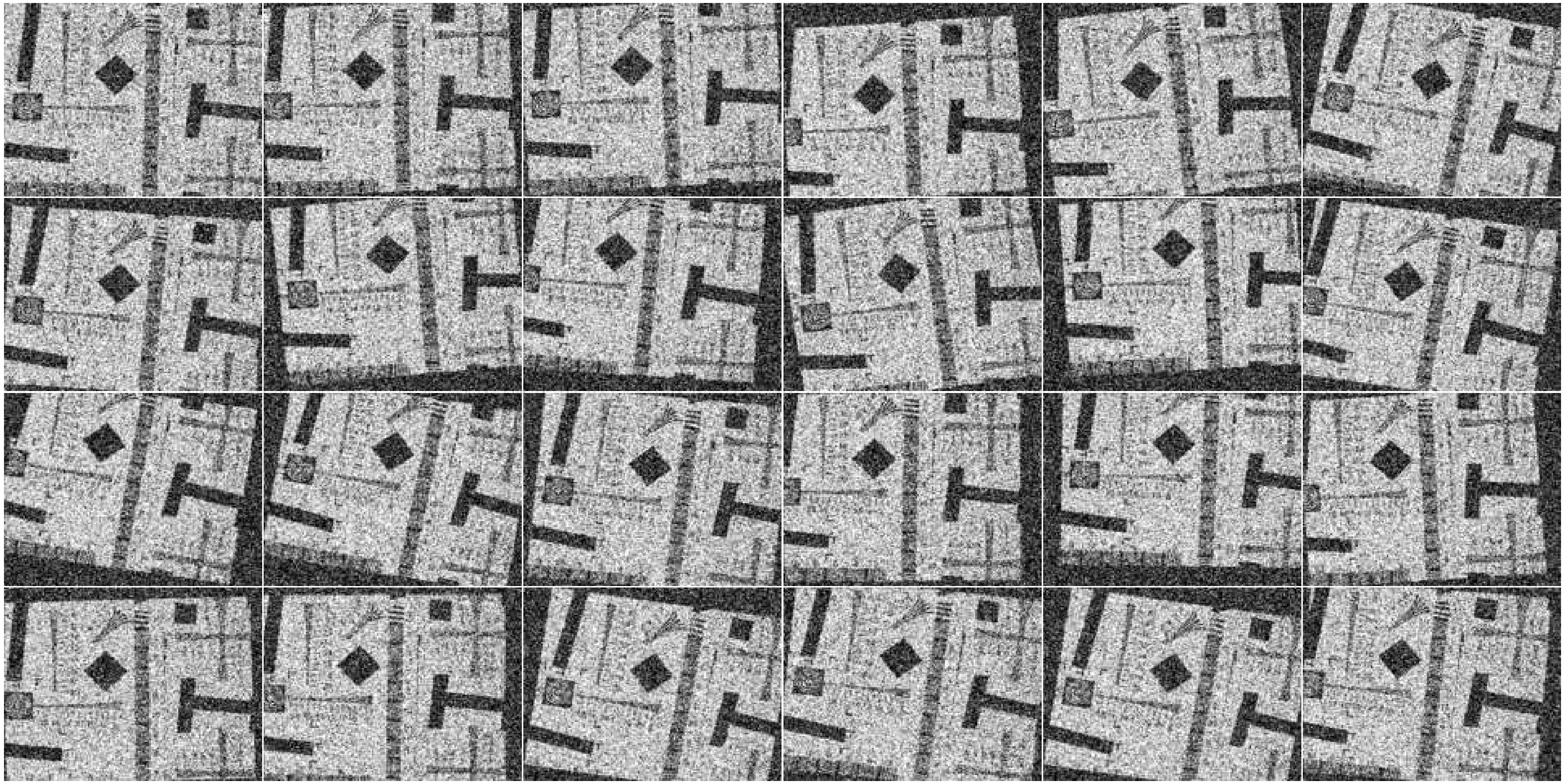
- ρ is a overlap of HR and LR pixels
- d is a Euclidean distance function
- We define the selector function $\alpha_k(p, p')$

$$\alpha_k(p,p') = \begin{cases} 1 & \text{if } d((p'),(p'')) < \theta_1 \text{ and } \rho \neq 0 \\ 0 & \text{otherwise} \end{cases}$$

- Selector function defines p' should contribute towards the reconstruction of p or not

Results: synthetic noisy images

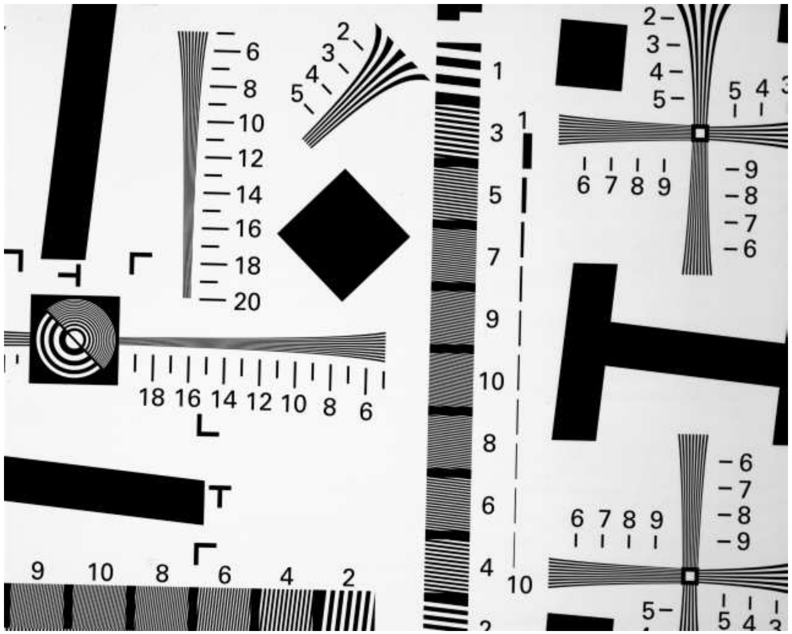
- size of a LR image = 160×128 , number of images = 24



Input images

Results: synthetic noisy images

- magnification = 16 $\times$ pixel
- size of HR image =640×512

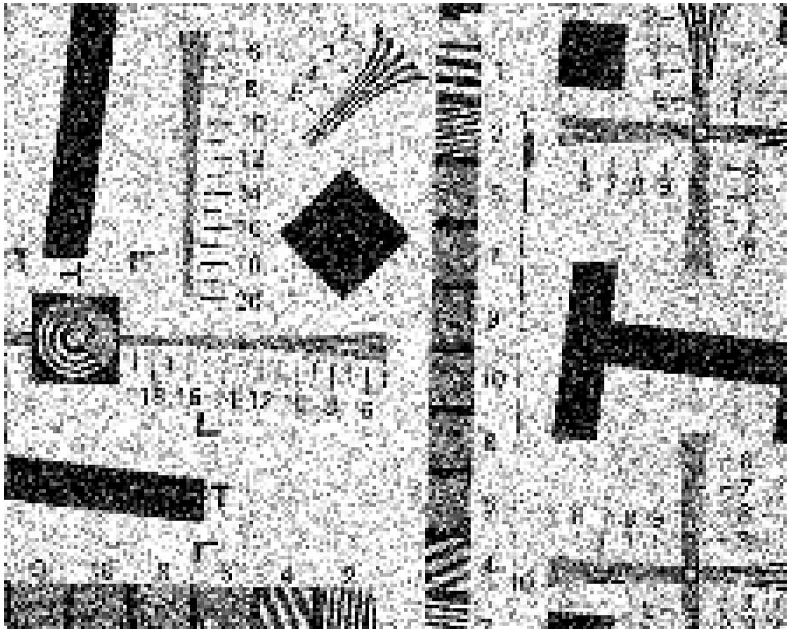


Ground truth

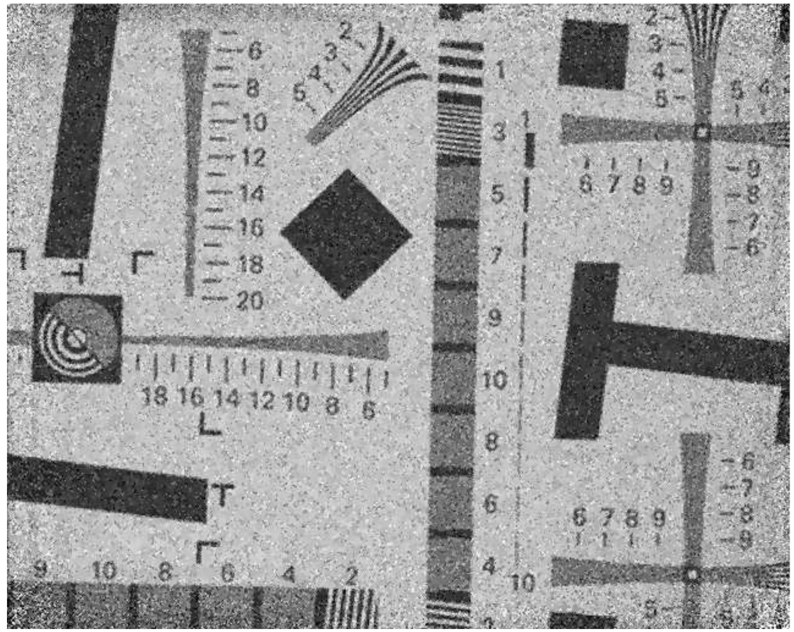
Bilinear interpolated

Results: synthetic noisy images

- Linear truncated prior $V_{p,q}(f_p - f_p, \Theta) = \lambda \ Min(20, |f_p - f_q|)$
- $\lambda = 6$



Single image expansion

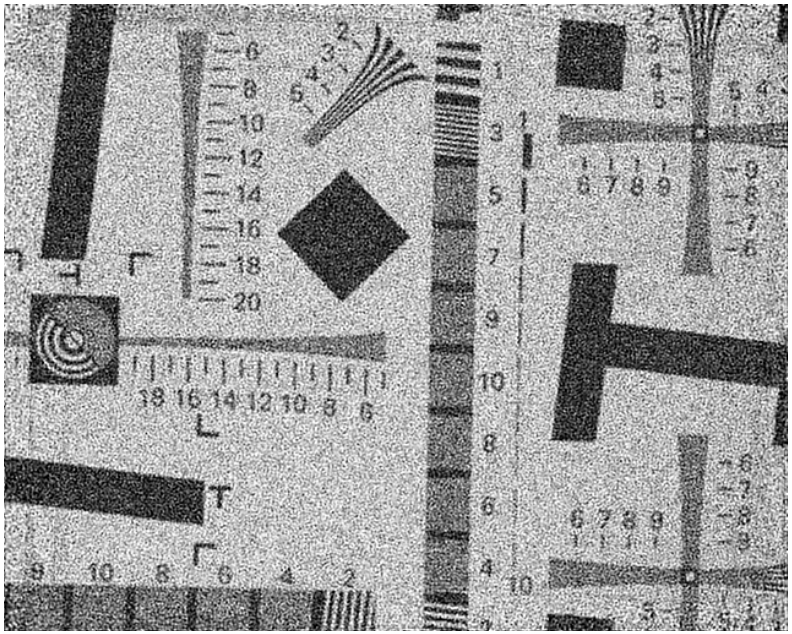


SR multiple images

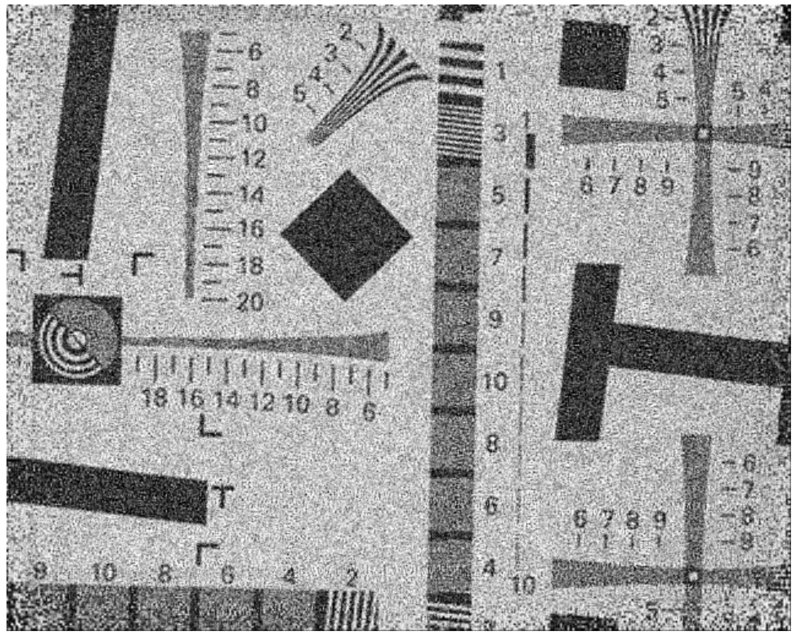
Conjugate gradient (CG) method

- Huber prior:

$$V_{p,q}(f_p - f_q, \Theta) = \begin{cases} (f_p - f_q)^2 & \text{if } |f_p - f_q| < \Theta \\ 2\,\Theta\,|f_p - f_q| - \Theta^2 & \text{otherwise} \end{cases}$$

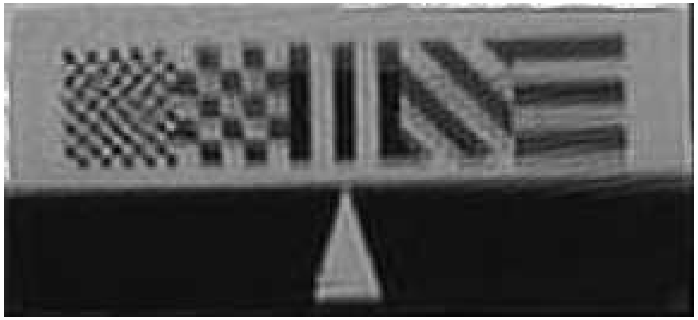


Without prior

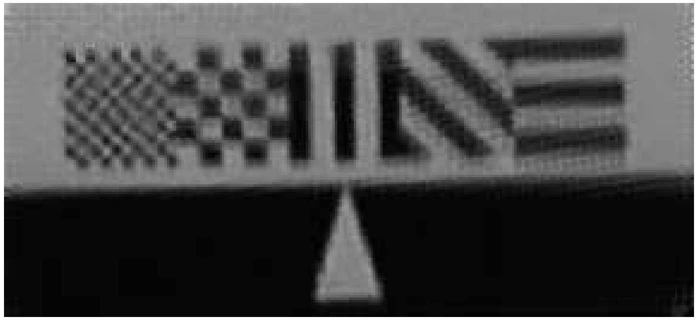


with Huber prior

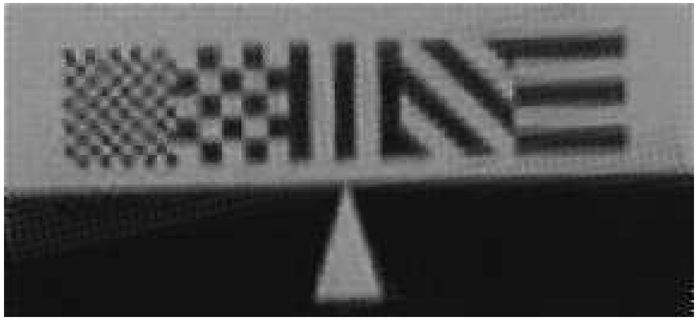
Real set



CG without prior



Graph cut without selector function



Graph cut with selector

Energy and *rms* error

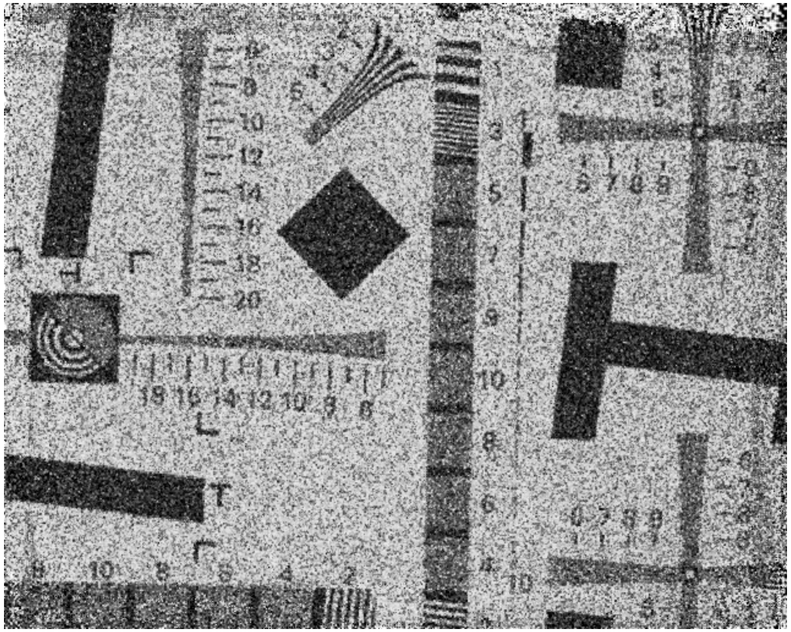
- Noisy synthetic images

Method	Energy	*rms* error
Conjugate gradient (CG)	$30.6e+08$	65.5
Graph cut	$29.8e+08$	60.9

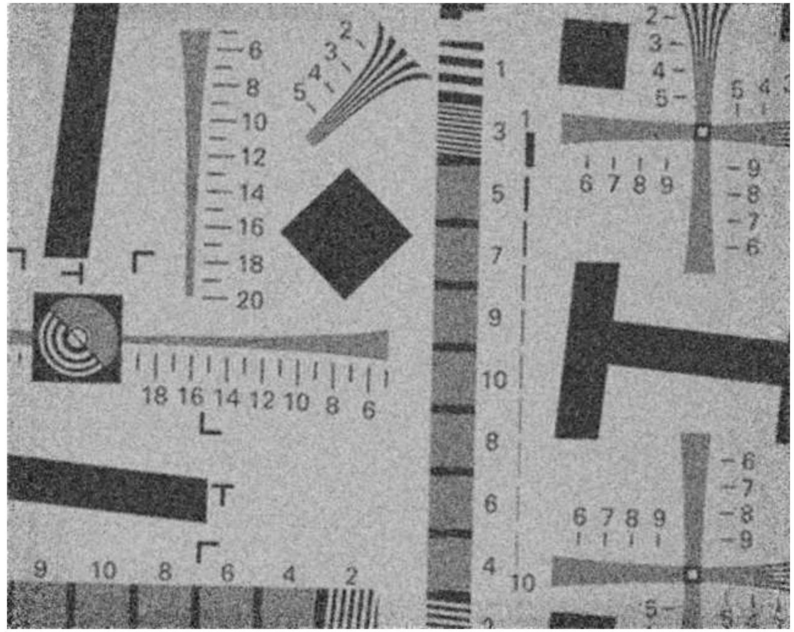
- Real images

Method	Energy
Conjugate gradient	$22.46e+06$
Graph cut without α_k	$22.81e+06$
Graph cut with α_k	$19.91e+06$

Effect of number of images

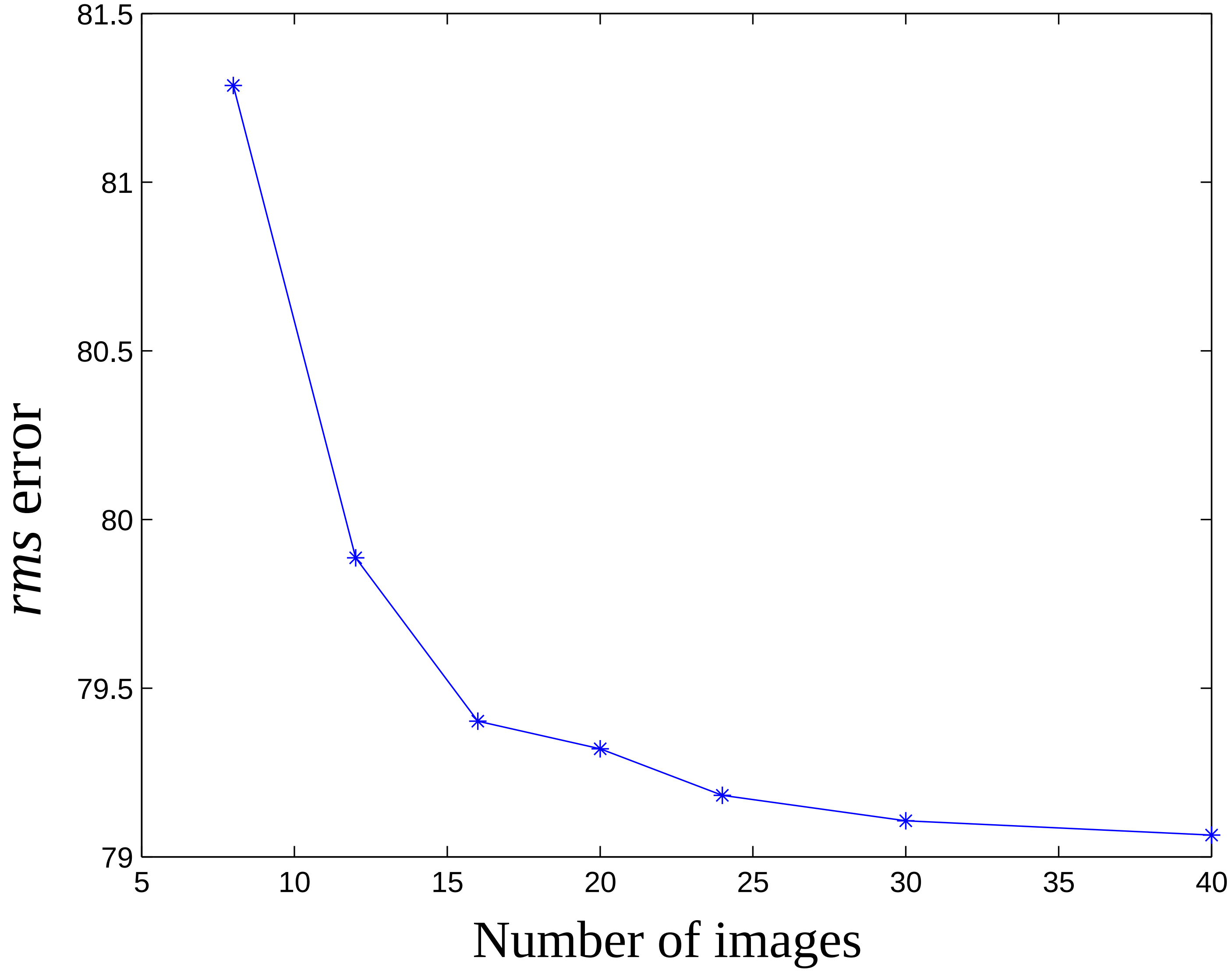


Using 8 input images



Using 40 input images

Effect of number of images



Effect of λ and Θ

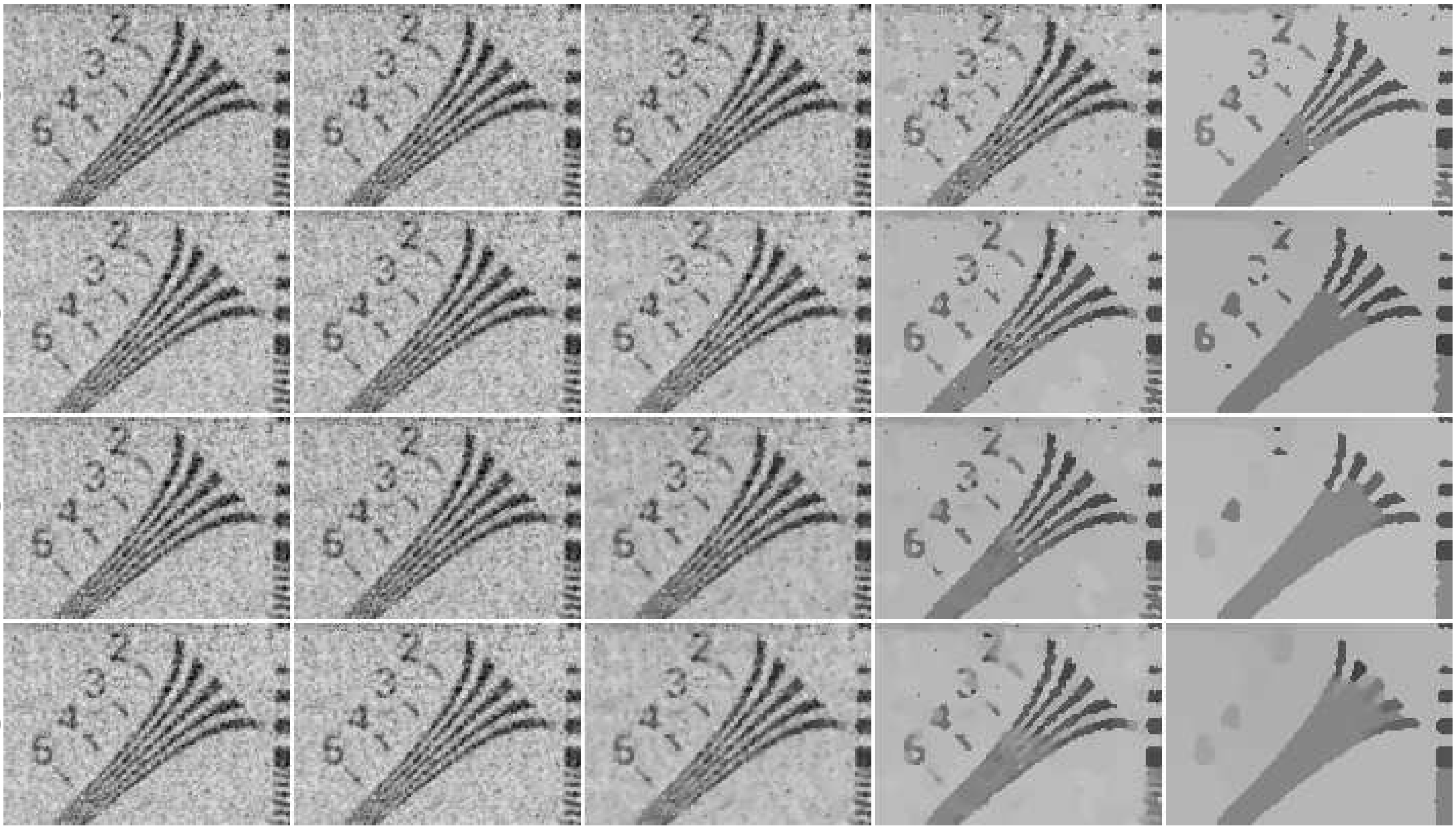
$\lambda = 0.06$ $\lambda = 0.6$ $\lambda = 6$ $\lambda = 20$ $\lambda = 60$

$\Theta = 10$

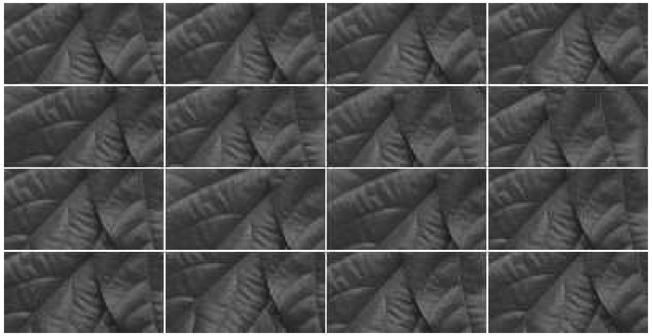
$\Theta = 20$

$\Theta = 30$

$\Theta = 40$



Real set 1



Input images (100×50 each)

Interpolated ($16\times$ pixel)

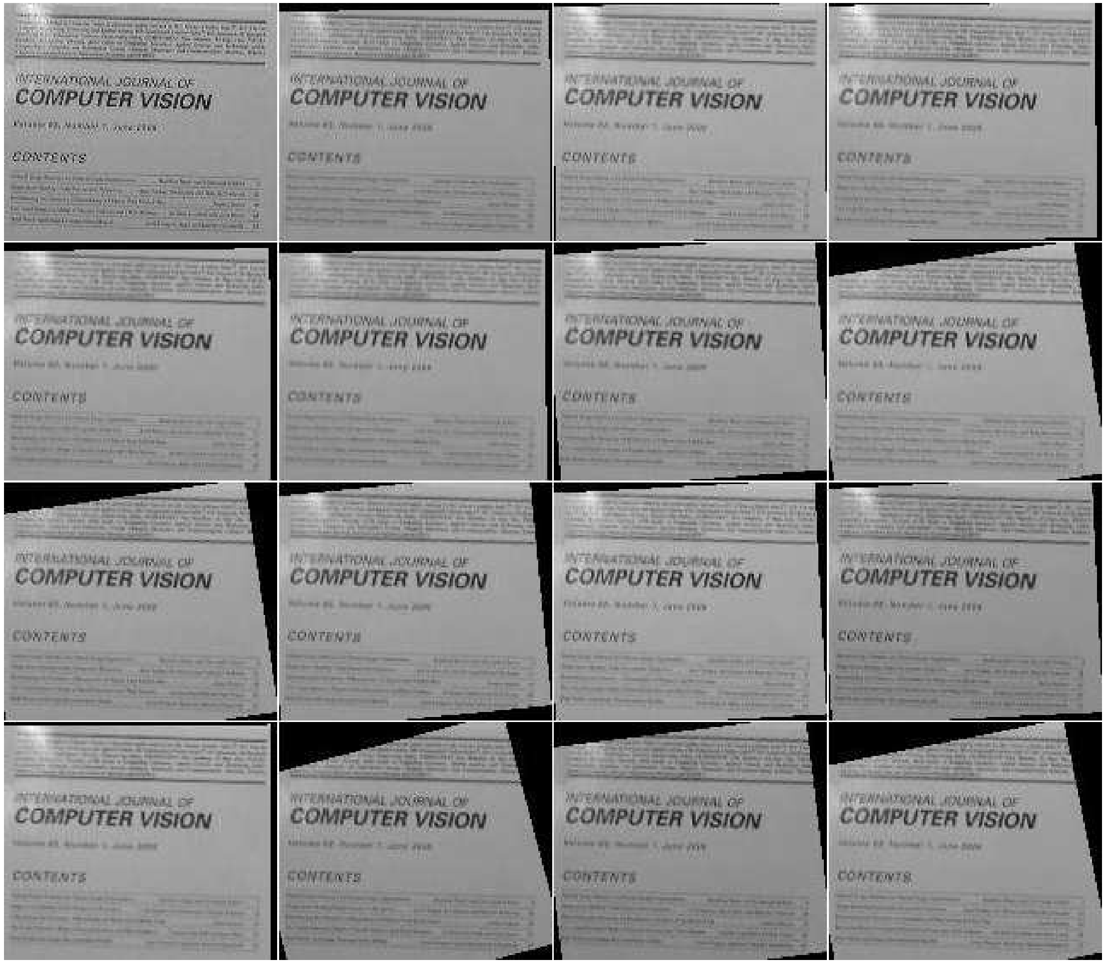
Single Image expansion

IBP method

SR without prior

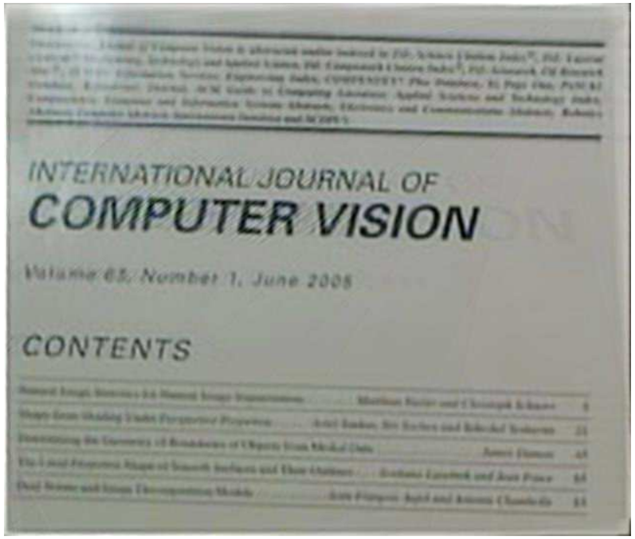
SR with prior

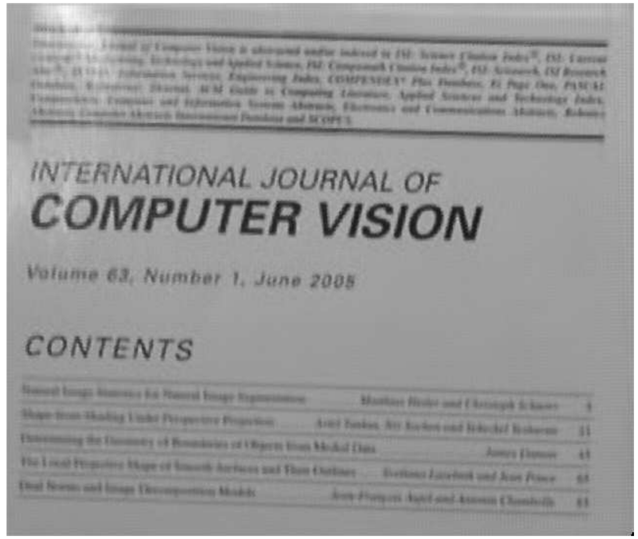
Real set 2



Registered input images(each of size 160×140)

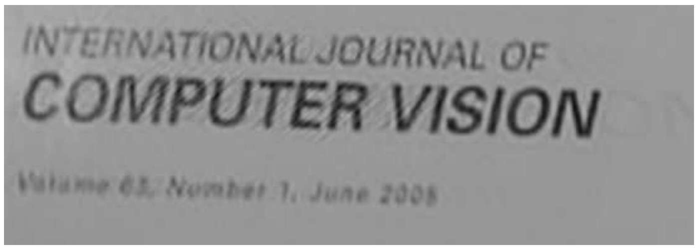
Real set 2

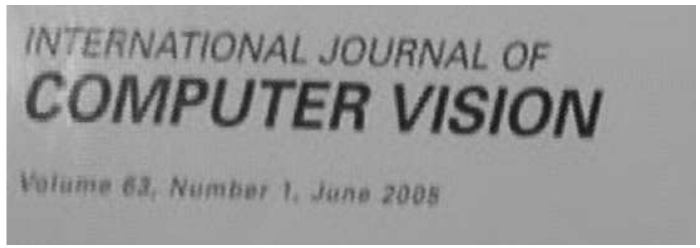




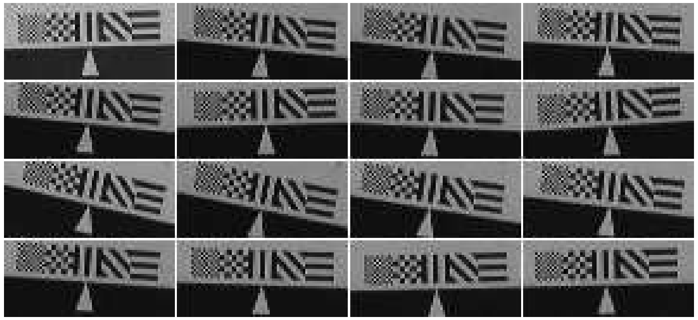
IBP method

SR multiple images

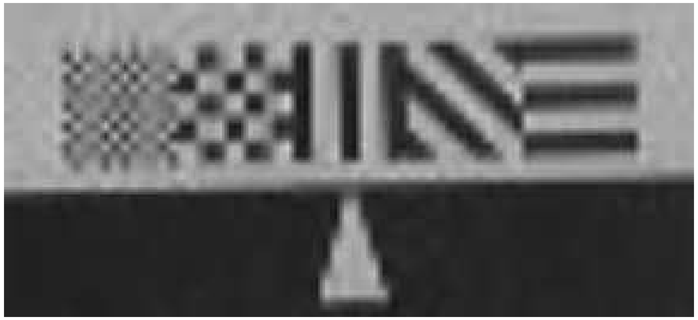




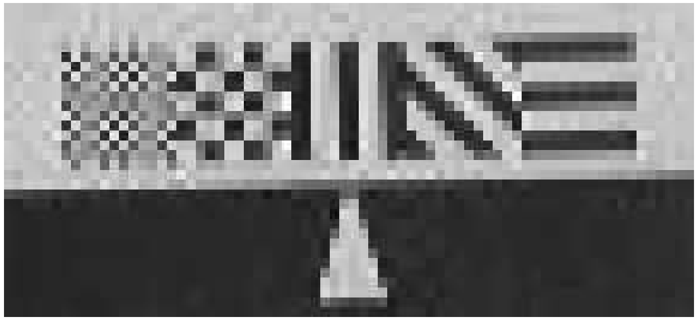
Real set 3



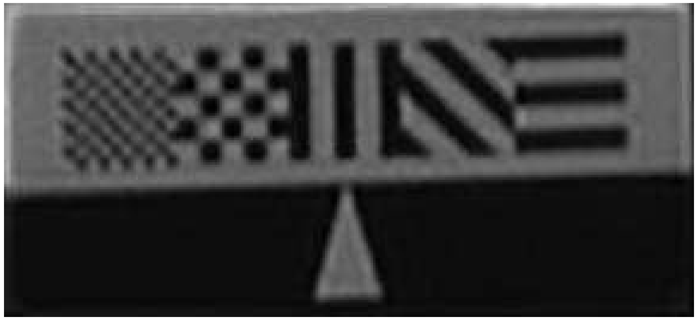
Input images (16 images each of 72×32)



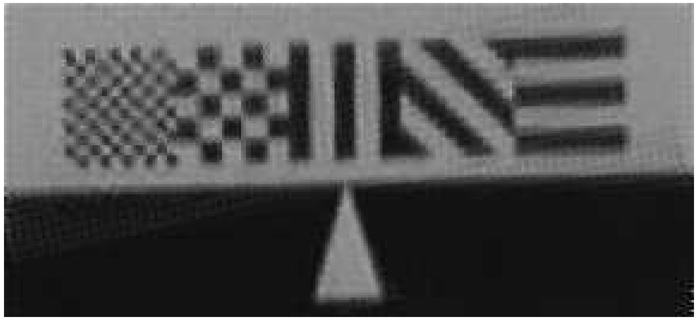
Interpolated



Single image expansion

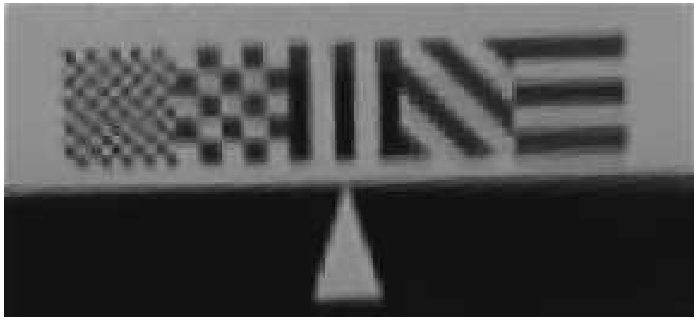


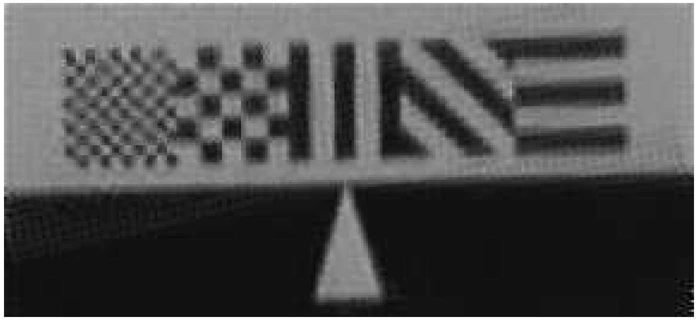
IBP method



SR multiple images

Effect of radius of *zone of influence*

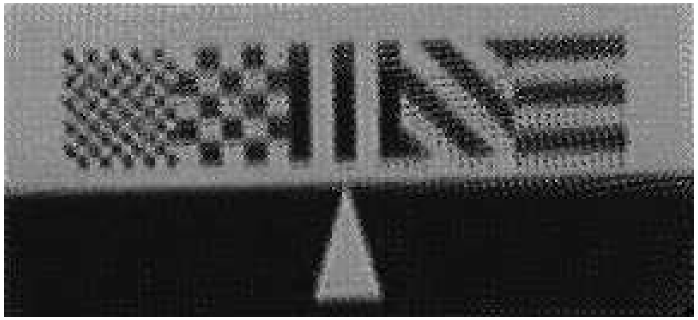




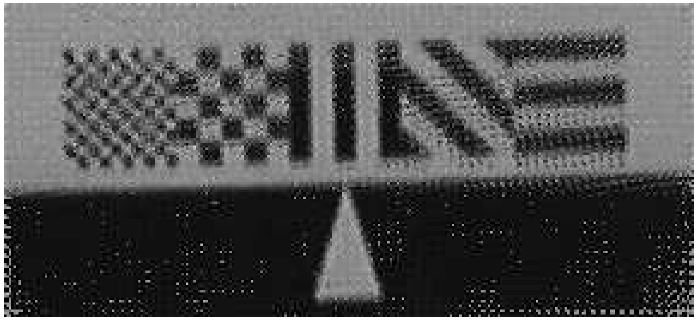
Radius

= 0.75

Radius = 0.50



Radius = 0.25



Radius = 0.125

Radius $\approx \sigma$ of the PSF function of camera

Super resolution in both space and time

Added complications with time

Motion blur

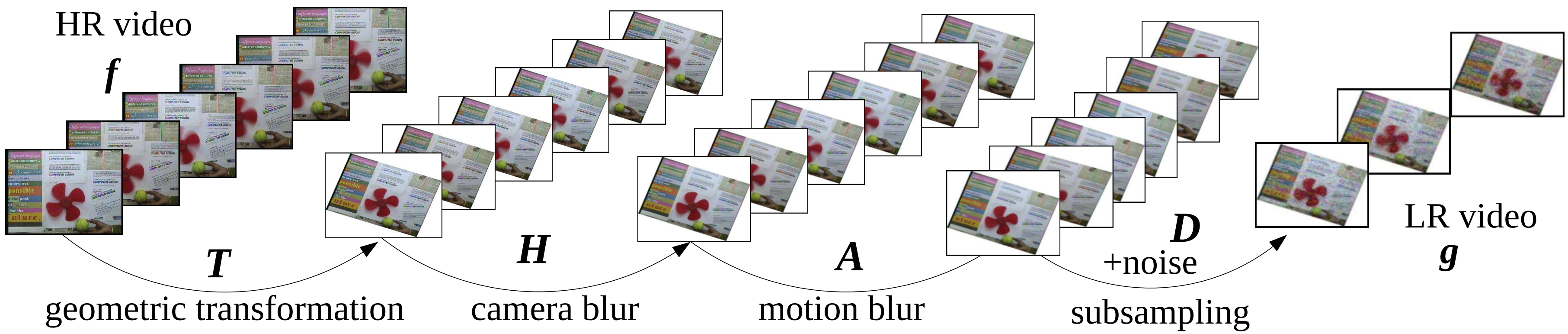
Motion aliasing

What people have done?

- Extensions of image SR - Stevenson (1999), Shah (1999)
- Dynamic SR - Elad and Milanfar *et al.* (1997, 2006)
- Video Epitome - Cheung *et al.* (2005)
- Reduction in motion blur - Raskar *et al.* (2006)
- Space-time SR - Shechtman *et al.* (2005, 2004)
  - use linear model
  - tradeoff in space time SR
  - input videos=8, space resolution= 2×2, temporal resolution=2

Space-time SR model

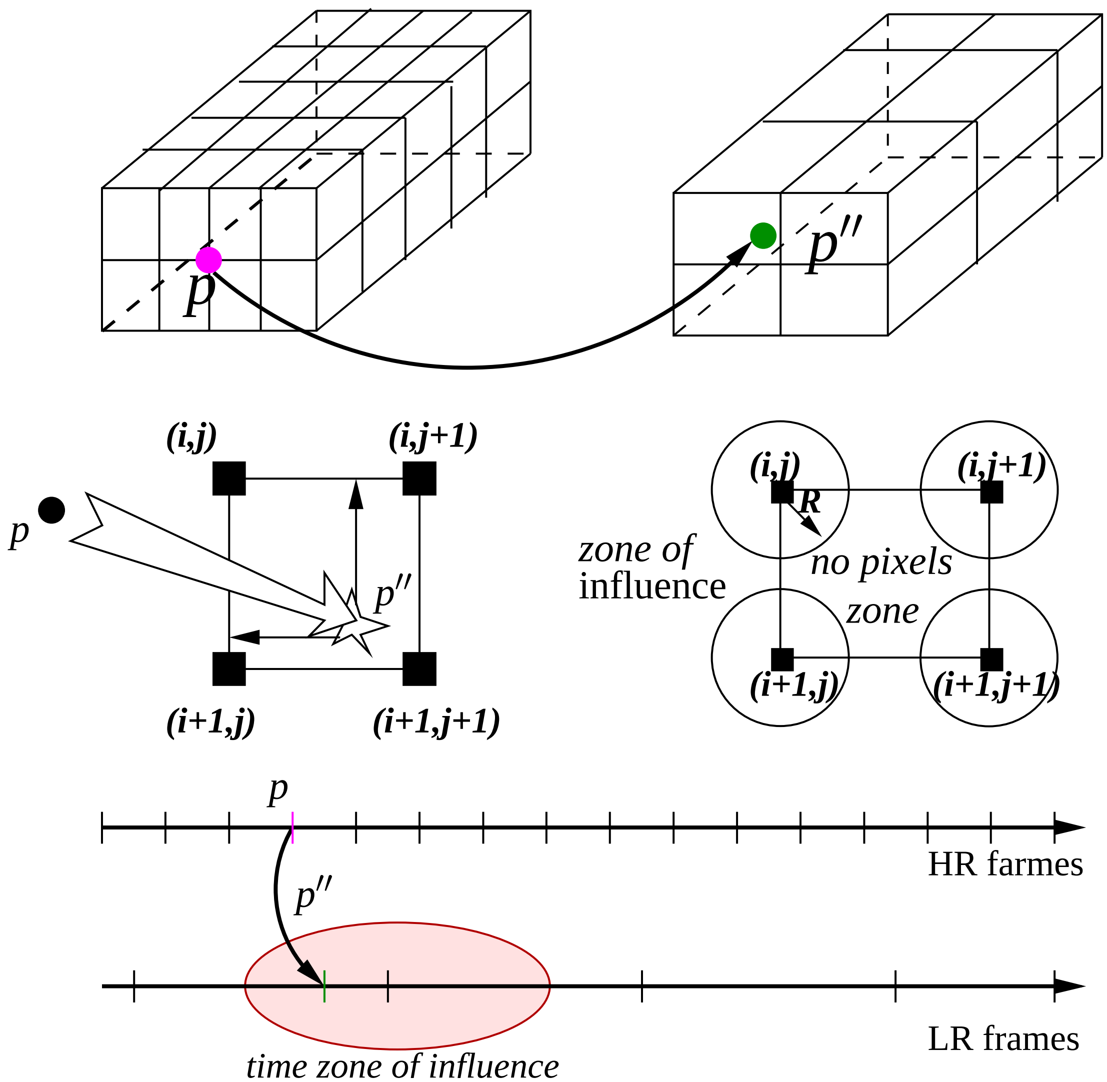
$$\mathbf{g_k} = DA_kH_kT_k\mathbf{f} + \eta_k \qquad 1 \le k \le n$$



Video registration

- T_1-scaling transformation between $\mathbf{f}$ and reference video ($\mathbf{g}_1$)
- $T_k = T_1\ T_{k\rightarrow 1}$
- Temporal misalignment - 1D affine
- Spatial transform - homography
- $T_{k\rightarrow 1}$ - using sequence-sequence alignment method proposed by Caspi and Irani

zone of influence



Space-time mapping

Selective SR

selector function for space

$$\alpha_k(p,p') = \begin{cases} 1 & \text{if } d((p'),(p'')) < \Theta_1 \text{ and } |t'-t''| \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

selector function for time

$$\beta_k(p,p') = \begin{cases} 1 & \text{if } d((p'),(p'')) < \Theta_1 \text{ and } |t'-t''| \leq \Theta_2 \\ 0 & \text{otherwise} \end{cases}$$

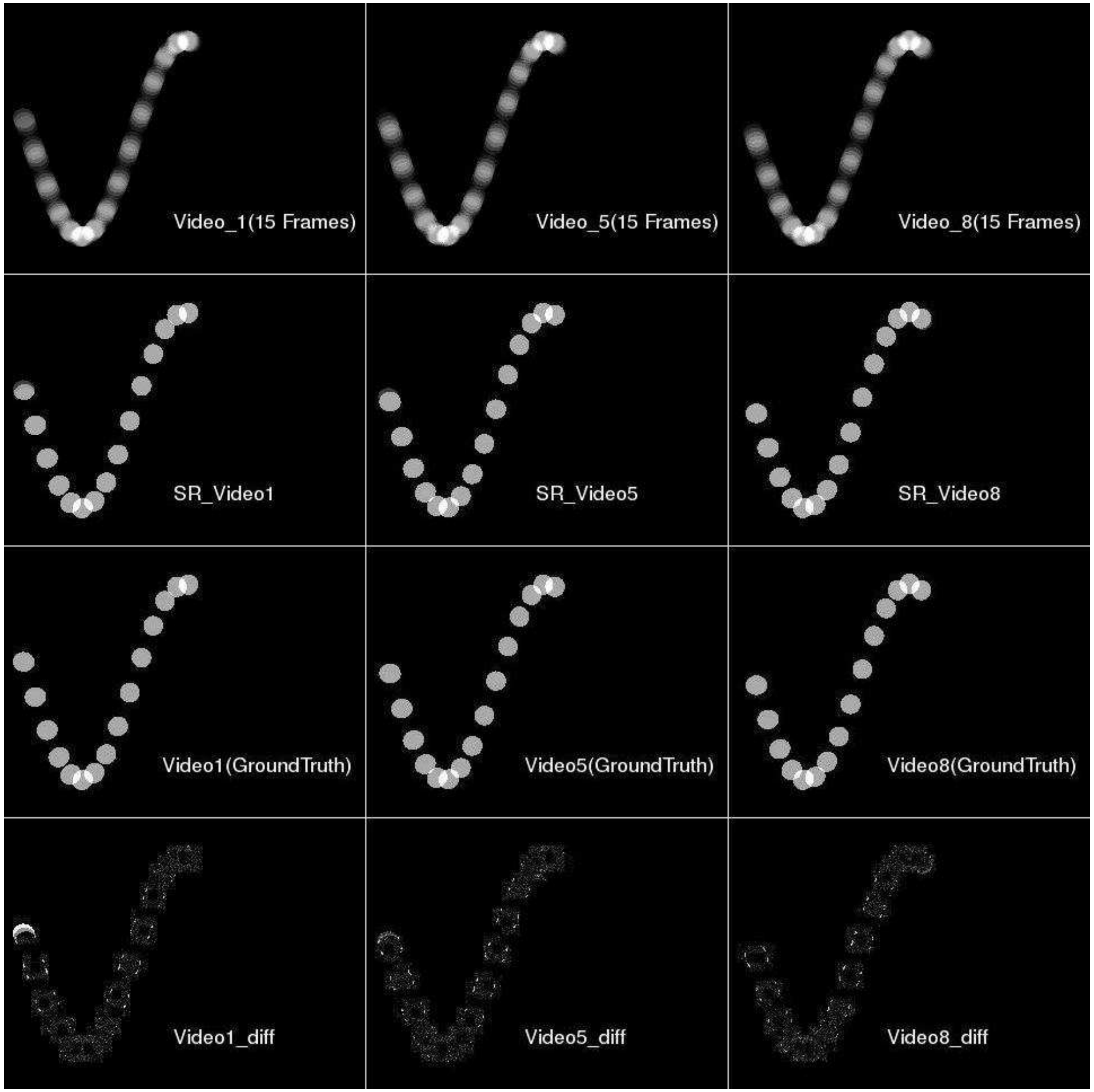
selector function for moving or static pixel

$$\gamma(x,y,t) = \begin{cases} 1 & \text{if } |g_1(x,y,t) - g_1(x,y,t-\Delta t)| > \Theta \text{ for } \Delta t = 1,2 \\ 0 & \text{otherwise} \end{cases}$$

selector function for space-time

$$\delta_k = \begin{cases} 1 & \text{if } \gamma(T_{k\to 1}p') = 1 \text{ and } \beta_k(p,p') = 1 \\ 1 & \text{if } \gamma(T_{k\to 1}p') = 0 \text{ and } \alpha_k(p,p') = 1 \\ 0 & \text{otherwise} \end{cases}$$

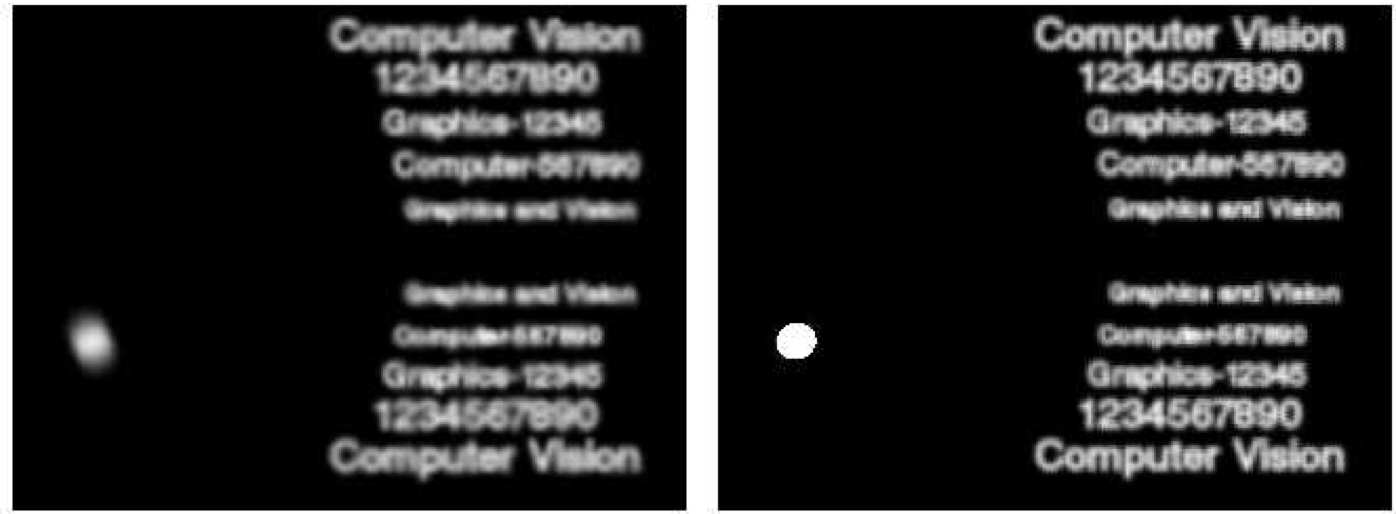
Ball example



Simultaneous super resolution

One frame from all the 9 input videos

Simultaneous super resolution



An interpolated and SR output frame

- number of input videos=9,
  magnification: space=2×2, time=9

Video SR - Different Inputs

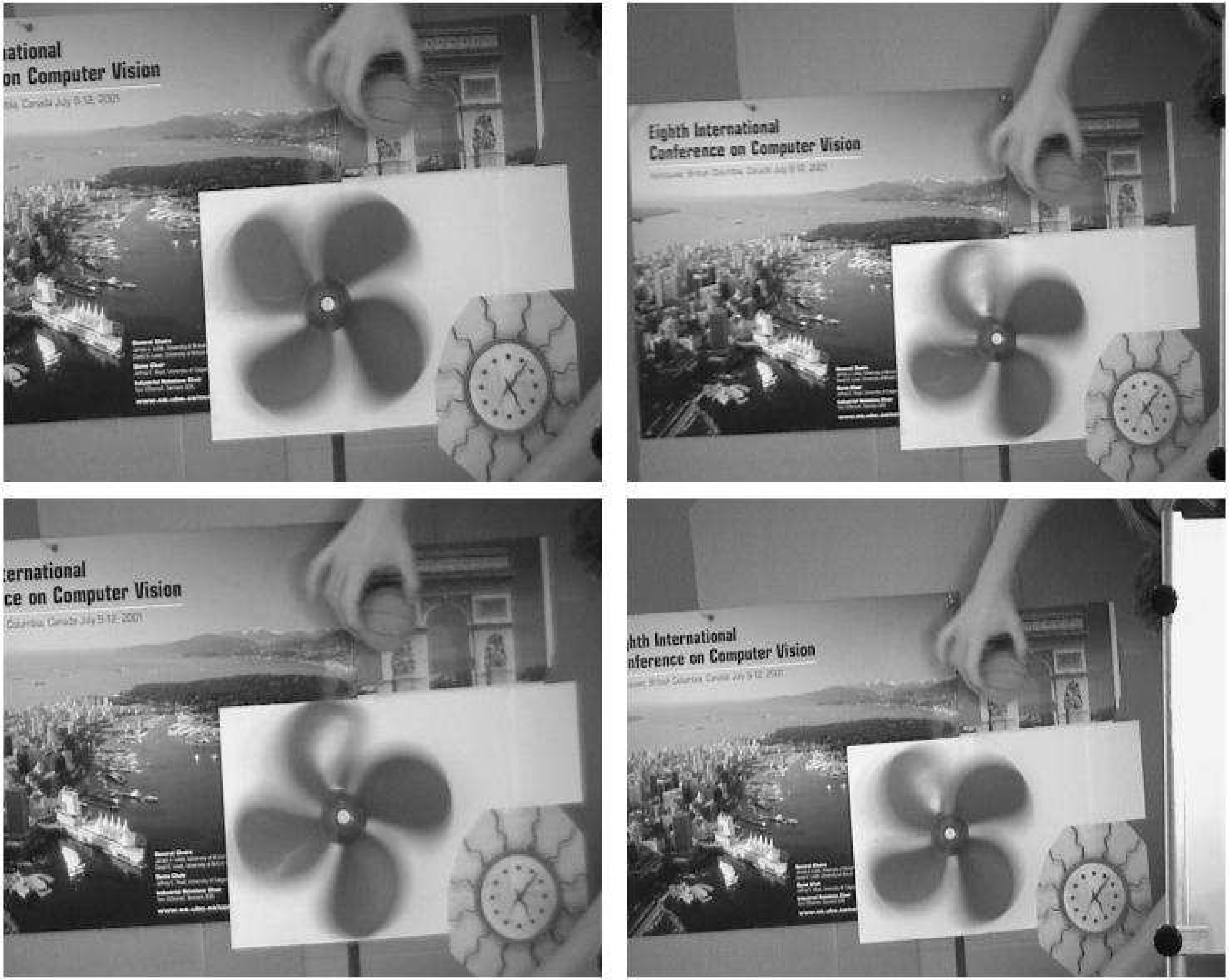
Video from camera -2   Interpolated video and VSR output

Video SR - Different Inputs

- Closeups from interpolated video and video SR output

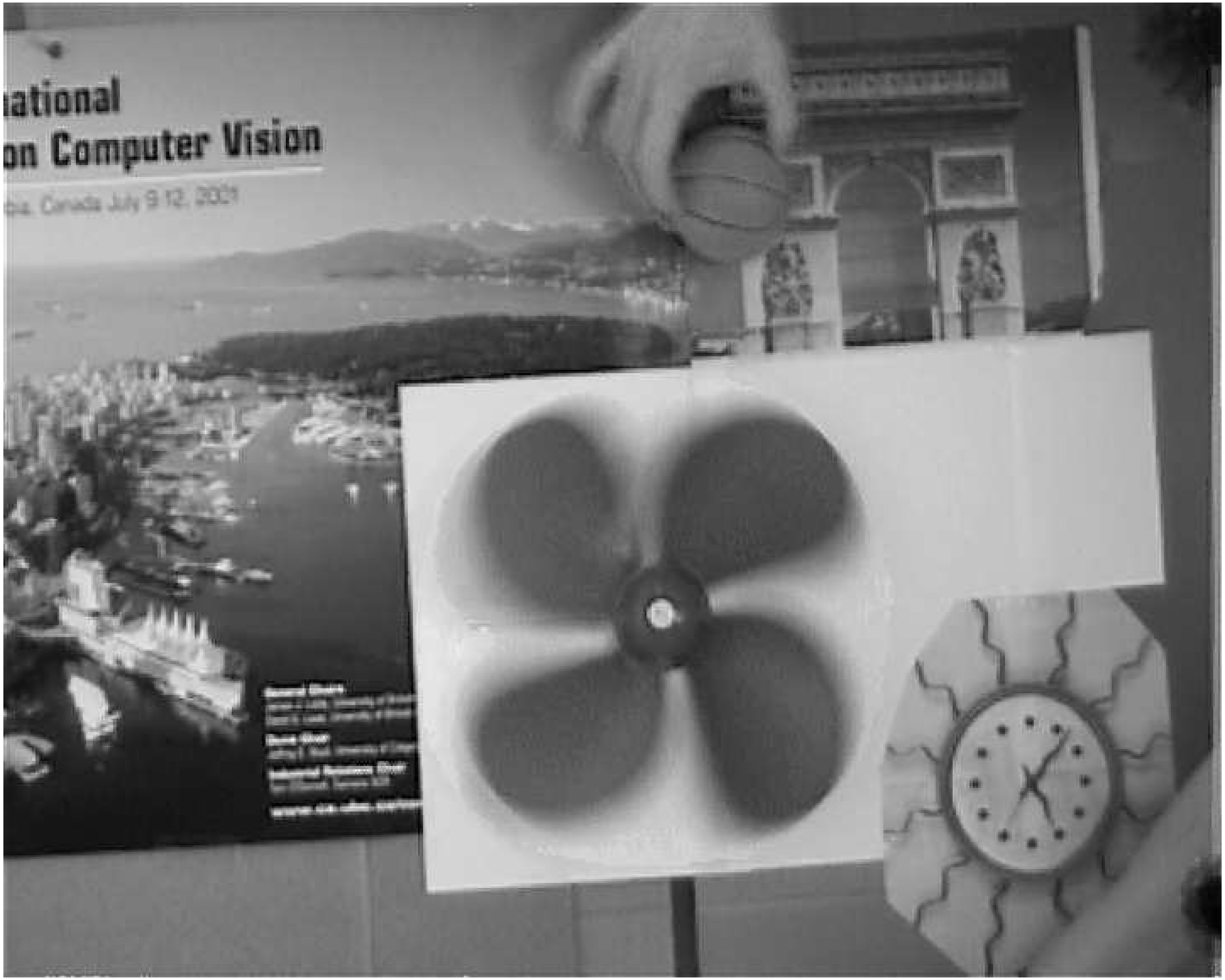
Motion aliasing and space SR

Input Videos



Input videos

Motion aliasing and space SR

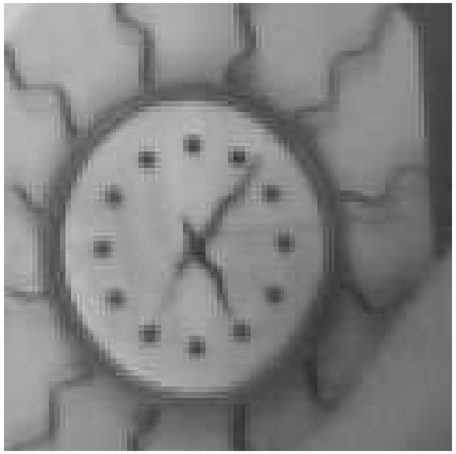
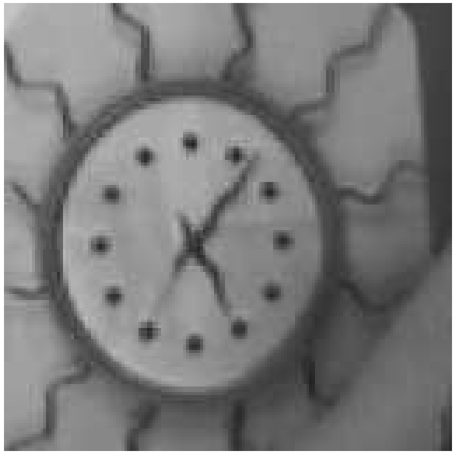


Space-time SR output

Close-ups




ational
on Computer Vision
Canada July 9-12, 2001

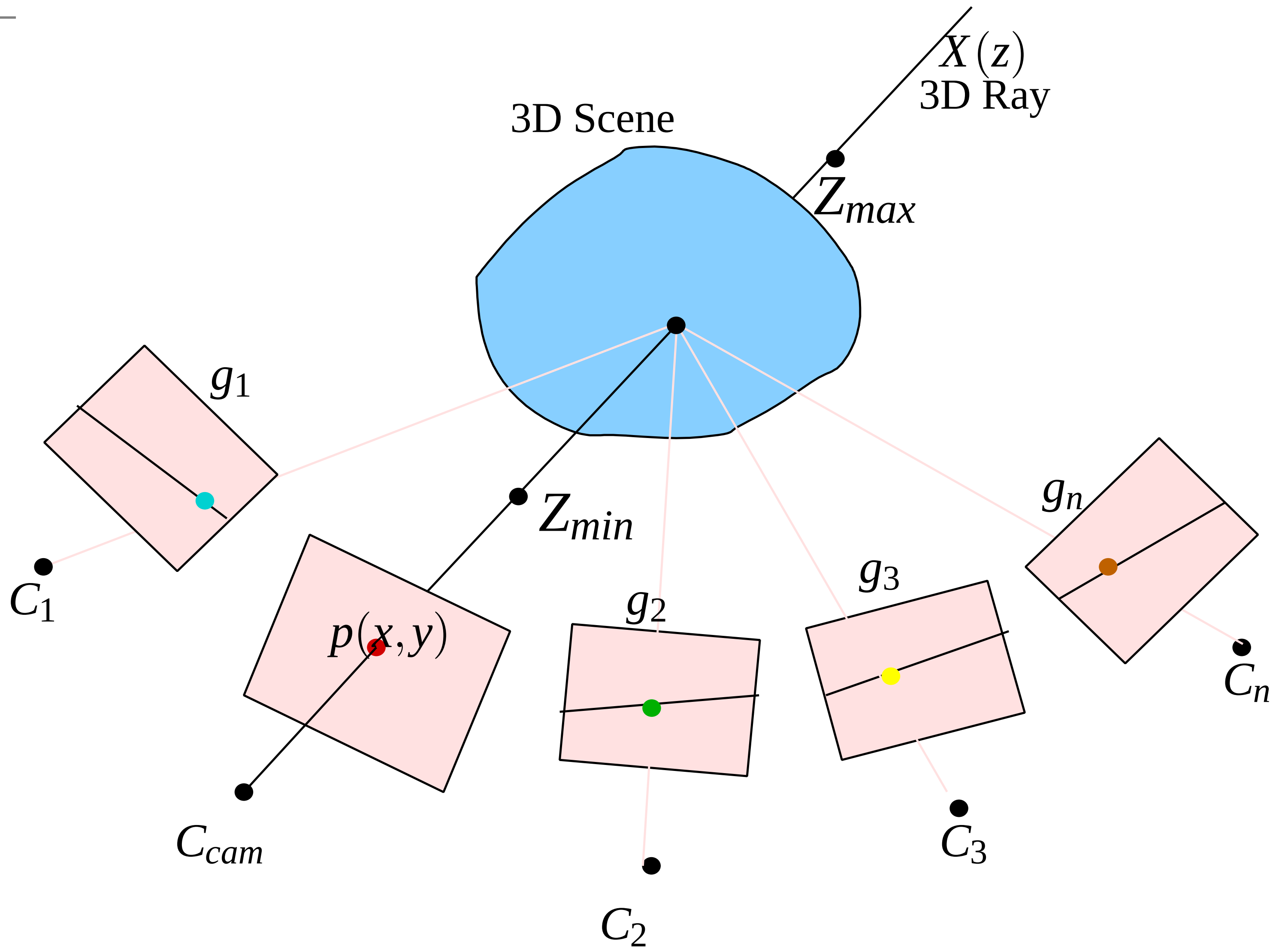
ational
on Computer Vision
Canada July 9-12, 2001



Super resolution of images of 3D scenes

Added complications

- Needs dense depth estimation
- Occlusion handling

Imaging geometry



MRF-MAP formulation

Energy

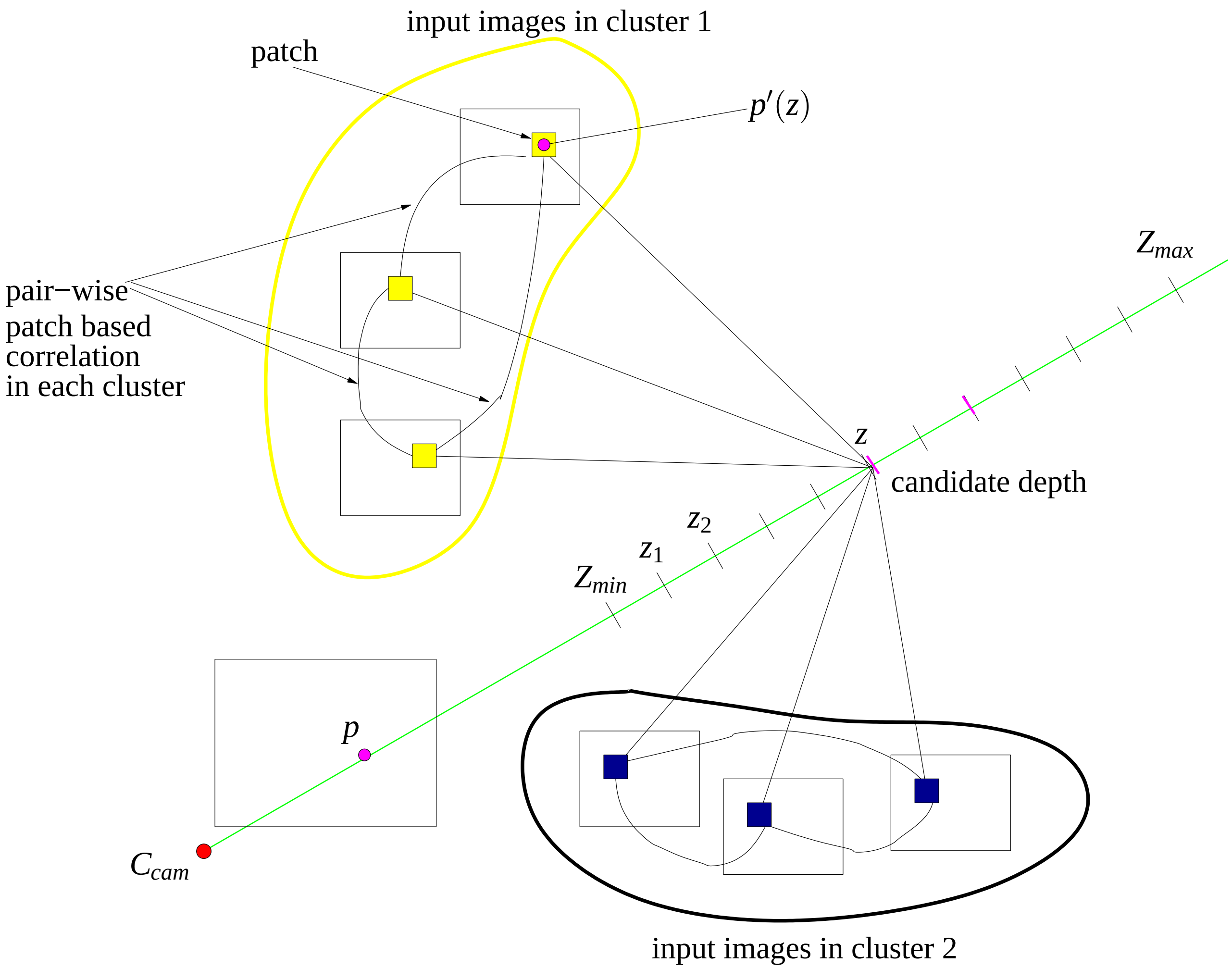
$$E(\mathbf{f},\mathbf{z}|\mathbf{g}) = \sum_{p\in\mathcal{S}}\sum_{k=1}^{n}\alpha_k(p,p'(z))(h(p)*\mathbf{f}(p,z)-\mathbf{g}_k(p'(z)))^2 + \lambda_s\sum_{p,q\in\mathcal{N}}V_{p,q}(\mathbf{f}(p),\mathbf{f}(q))$$

$\mathcal{S}$- set of sites (pixels) in the novel view, $\mathbf{f}(p,z)$ - label at site p and at depth z

Simultaneous minimization of $\mathbf{z}$ and $\mathbf{f}$ is intractable

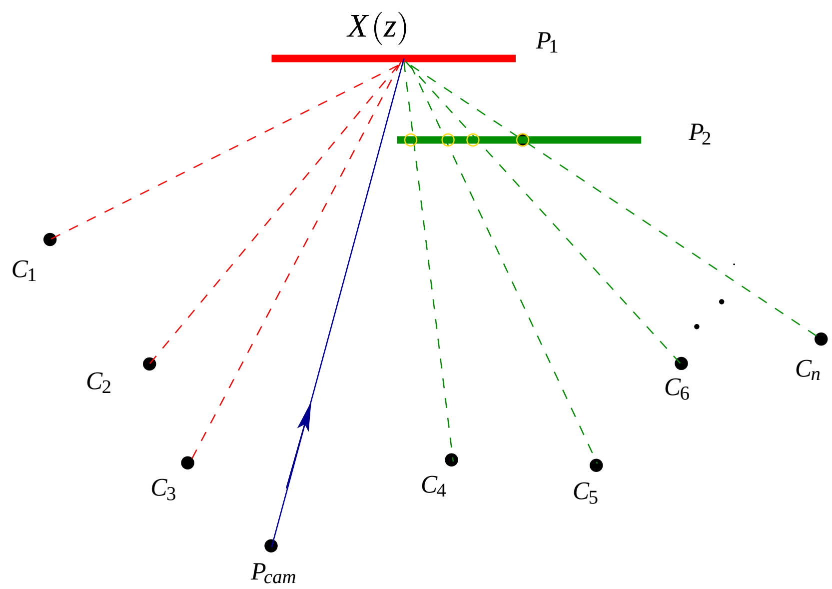
Pruning of colour labels

Clusters at a depth z



Clusters at candidate depths

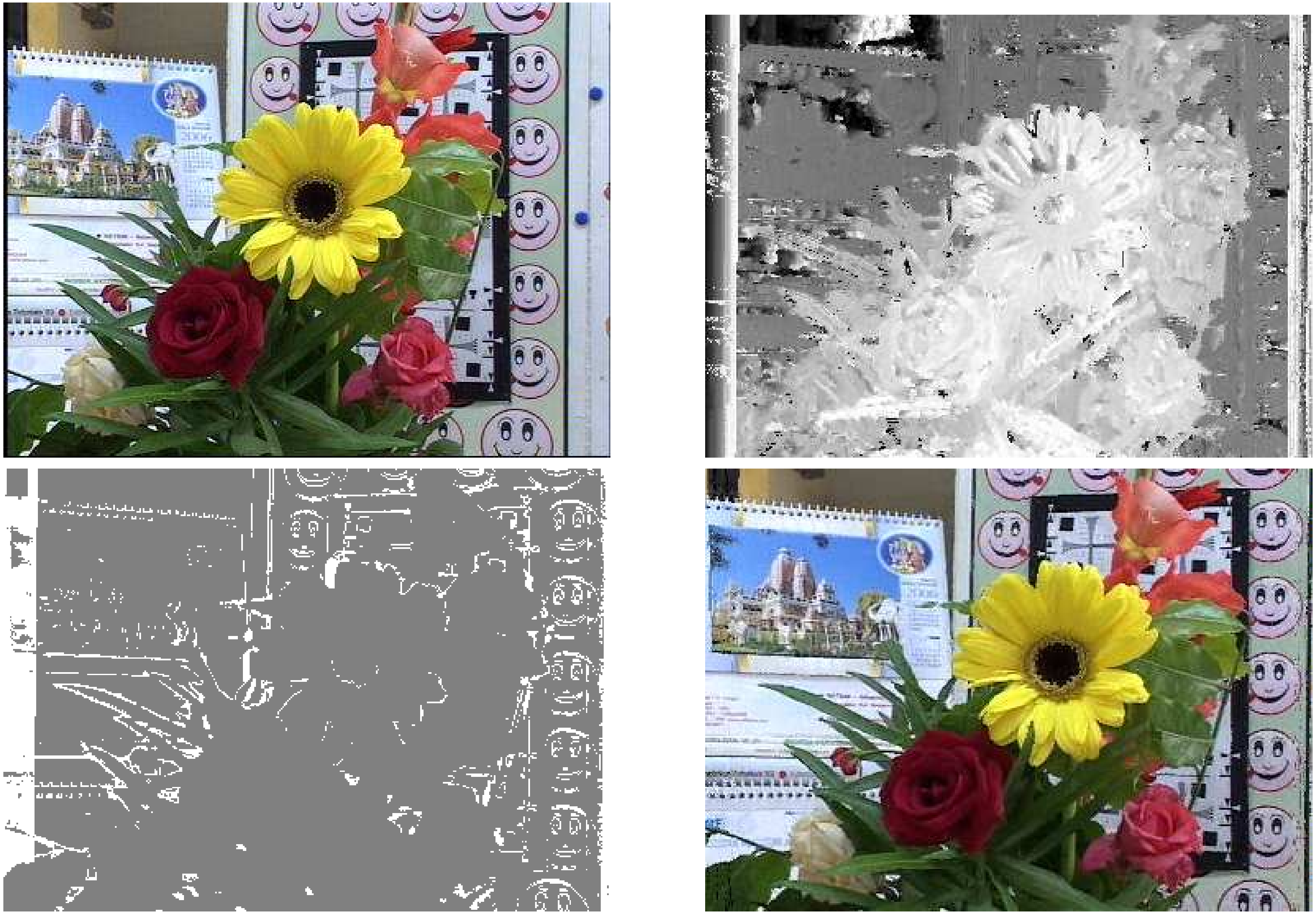
- One cluster: photo-consistent at this depth
- More than one: 3D point is occluded or HR pixel is on an edge



Occlusion handling

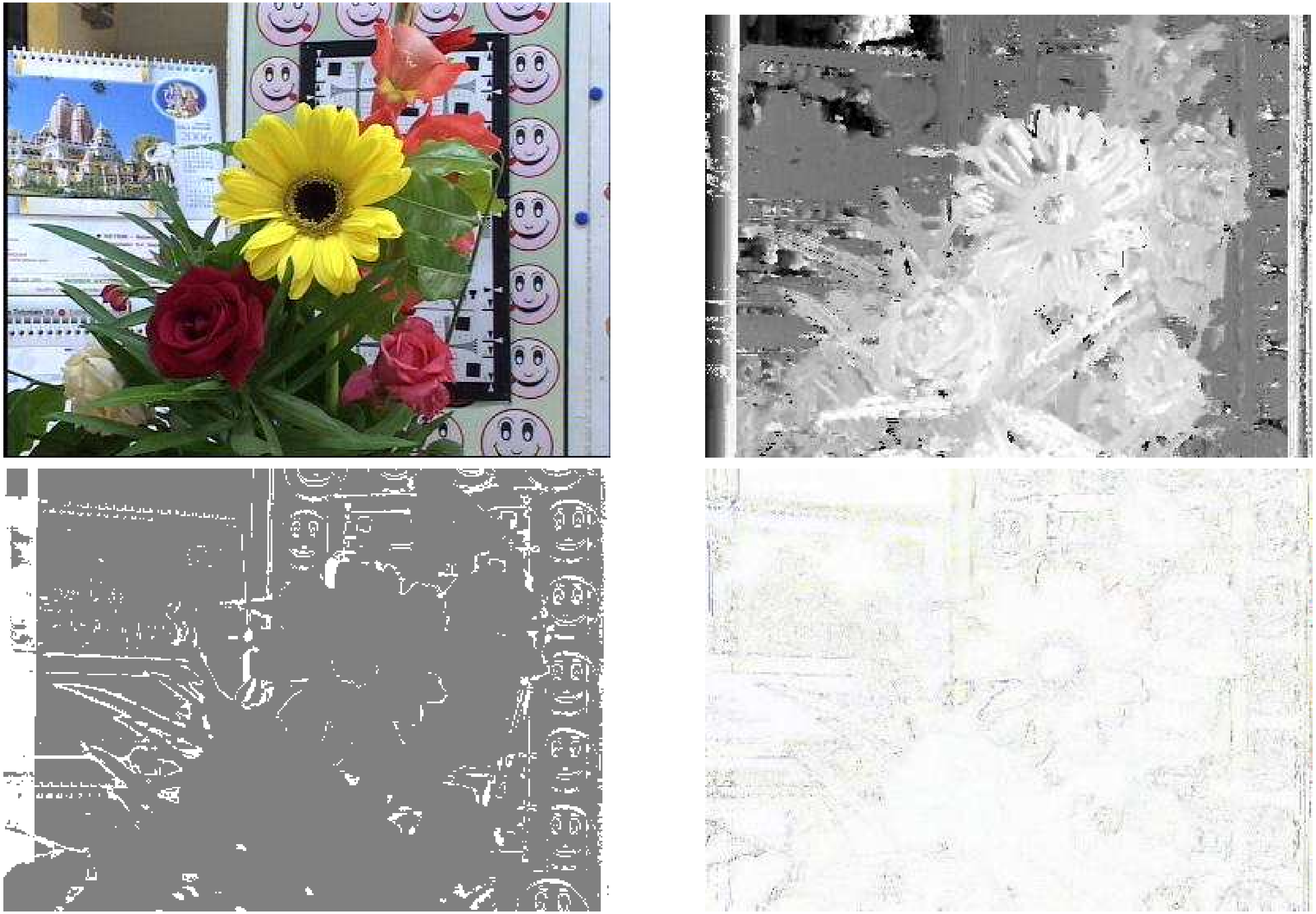
- Super resolution reconstruction
  - p is an edge point: we set weight to all clusters
  - p is not an edge point: look up the color in the LR image and set weight to 1 for the corresponding cluster and all others to zero
- Super resolved novel view
  - Ordinal visibility constraint: If $\mathbf{X}(z_1)$ and $\mathbf{X}(z_2)$ are any two points on two back projected rays then $\mathbf{X}(z_1)$ can occlude $\mathbf{X}(z_2)$ only if $z_1 < z_2$
  - We scan in the depth sorted order
  - Mark the pixels in LR images

Depth map and occlusion map



Row 1: one of the input images and depth map
Row 2: occlusion-map and the generated novel view

Difference with ground truth



Row 1: one of the input images and depth map
Row 2: occlusion-map and the error with ground truth

Novel view generation

Row 1: two input images out of 10

Row 2: novel view (using 10) and the difference with the ground

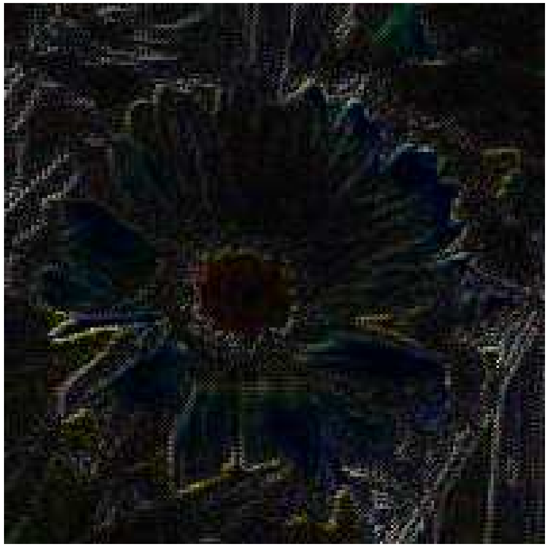
Interpolated missing view

Super resolved novel view

Close-ups

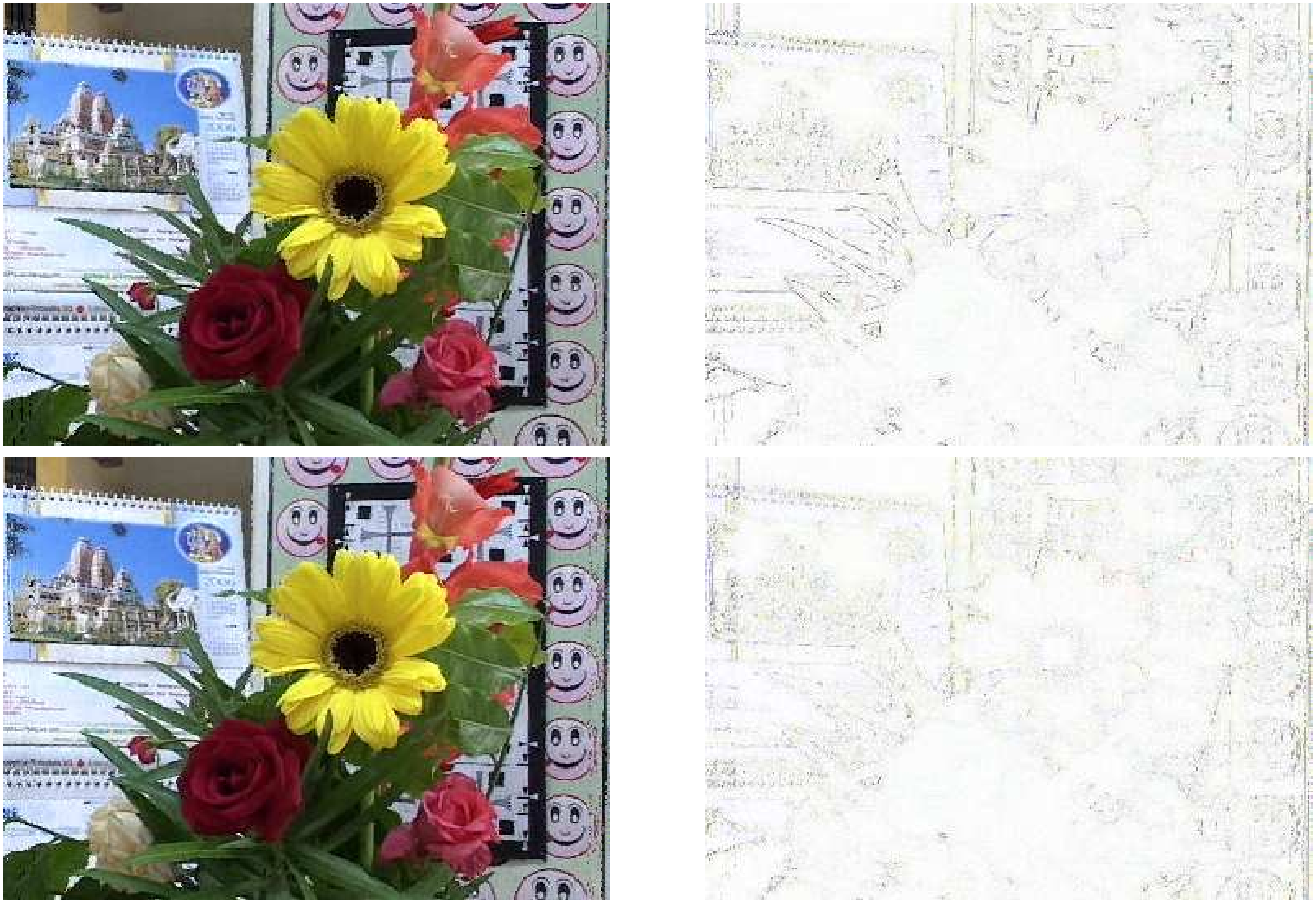
Interpolated

Super resolved



Difference

Effect of number of views

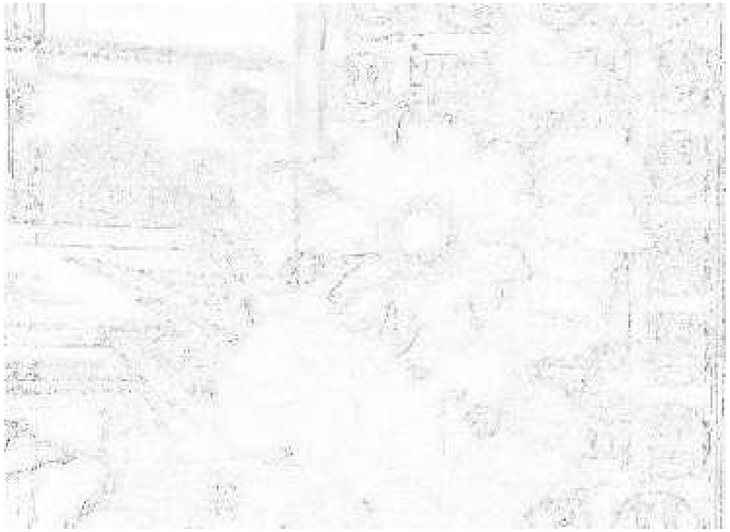


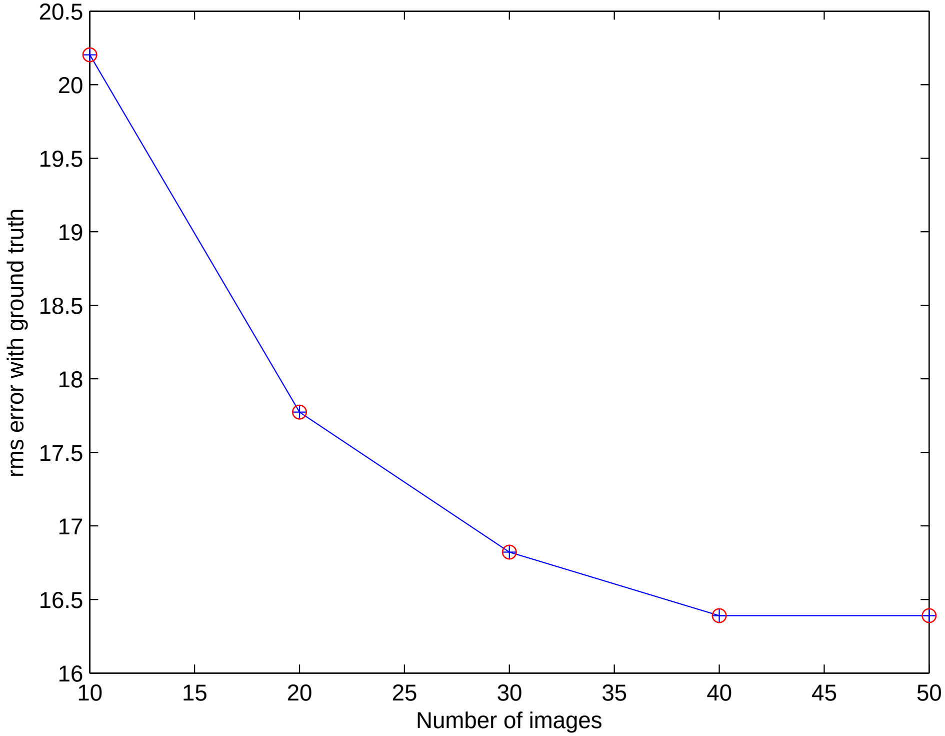
Novel view and the difference image using 10 and 20 images

Effect of number of views

Novel view and the difference image using 30 and 40 images

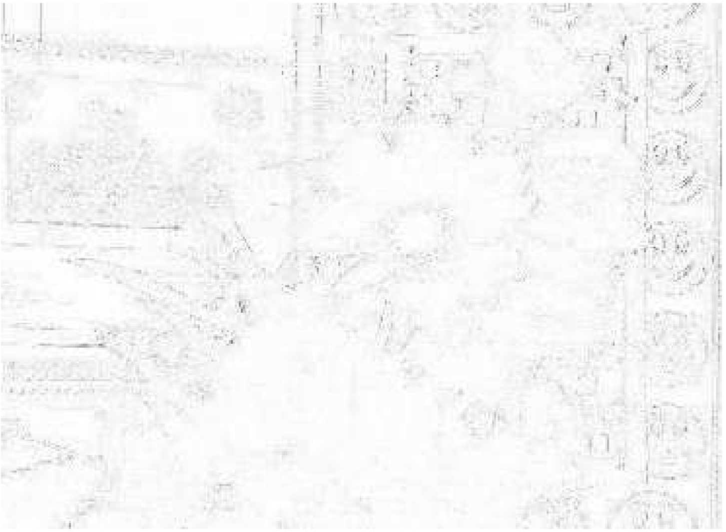
Effect of number of views





Effect of angle subtended

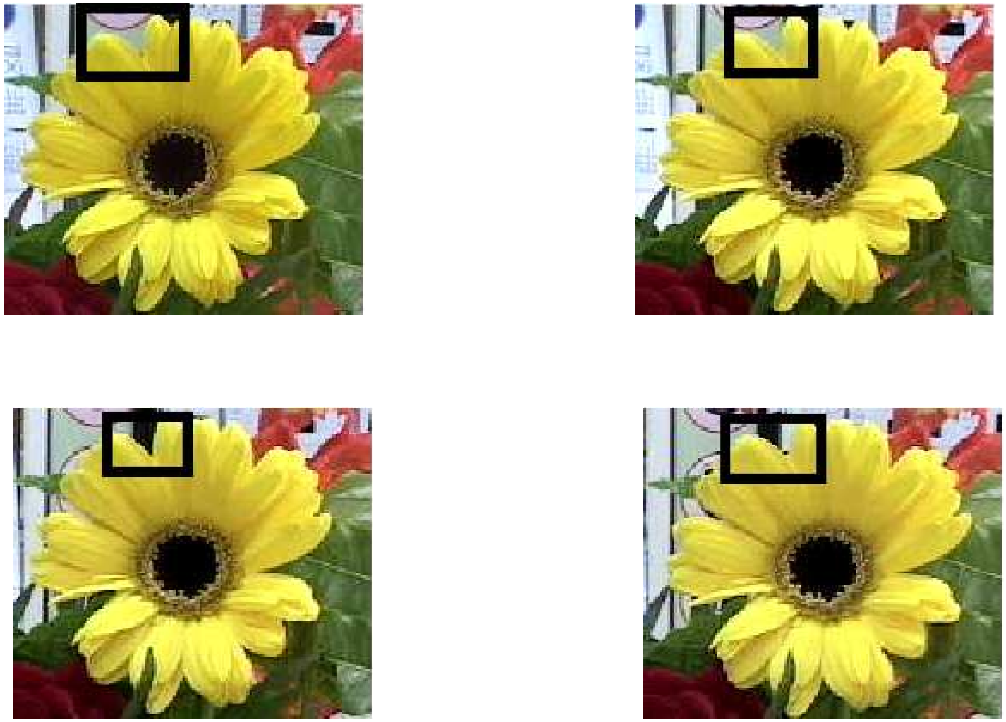


Novel view and the difference image with angles 11.5^o and 7.5^o

Occlusion handling

Occlusion handling



Close-ups of four input images

Occlusion handling

Super resolved novel view using 20 images