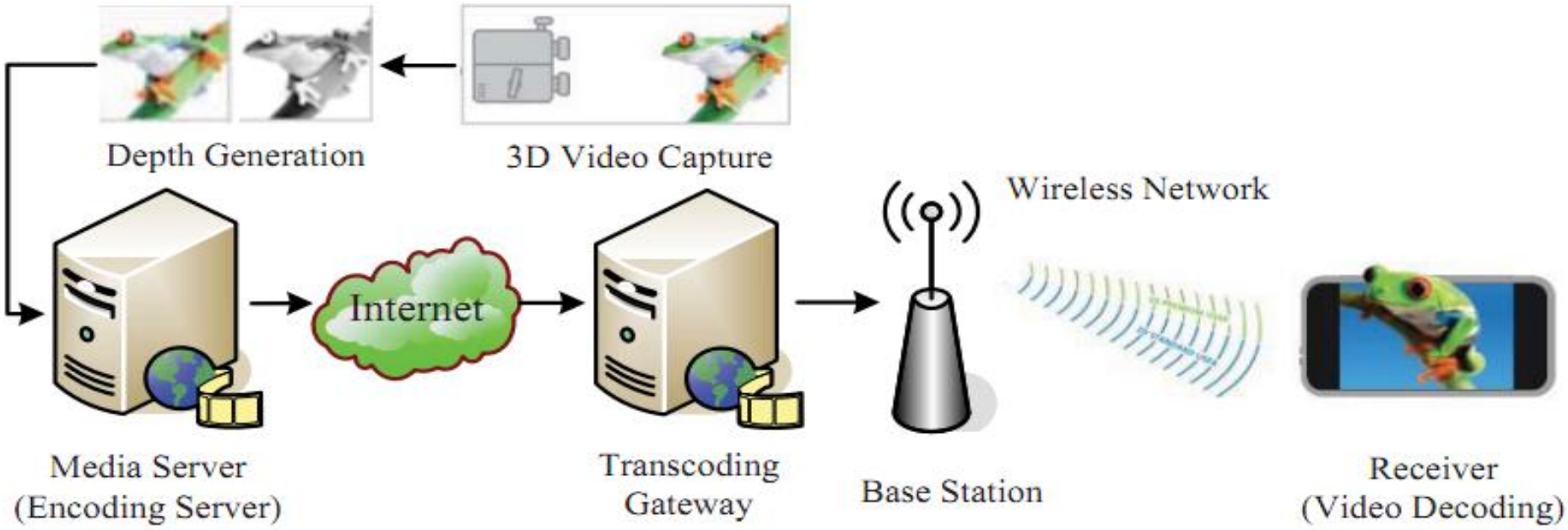
Scalable Disparity Estimation in 3D Multimedia Communications

Outline

- Stereoscopic image pairs
- Disparity estimation
- Local methods vs. global methods
- Computation time modeling
- Experimental results
- Applications

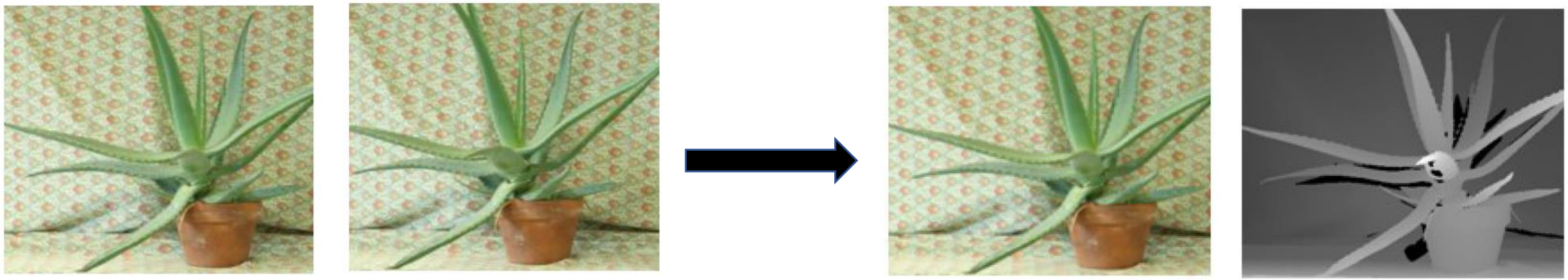
Stereoscopic image pairs

- 3D multimedia communications



Stereoscopic image pairs

- 3D multimedia communications

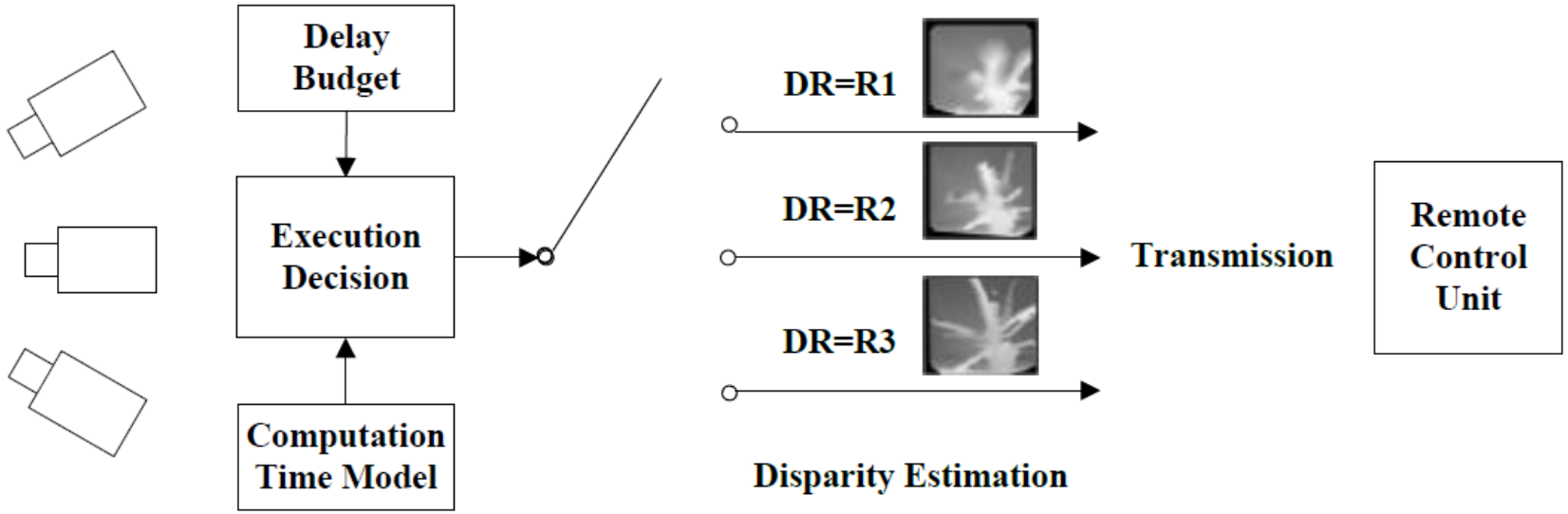


images from multiple cameras

color and disparity

Stereoscopic image pairs

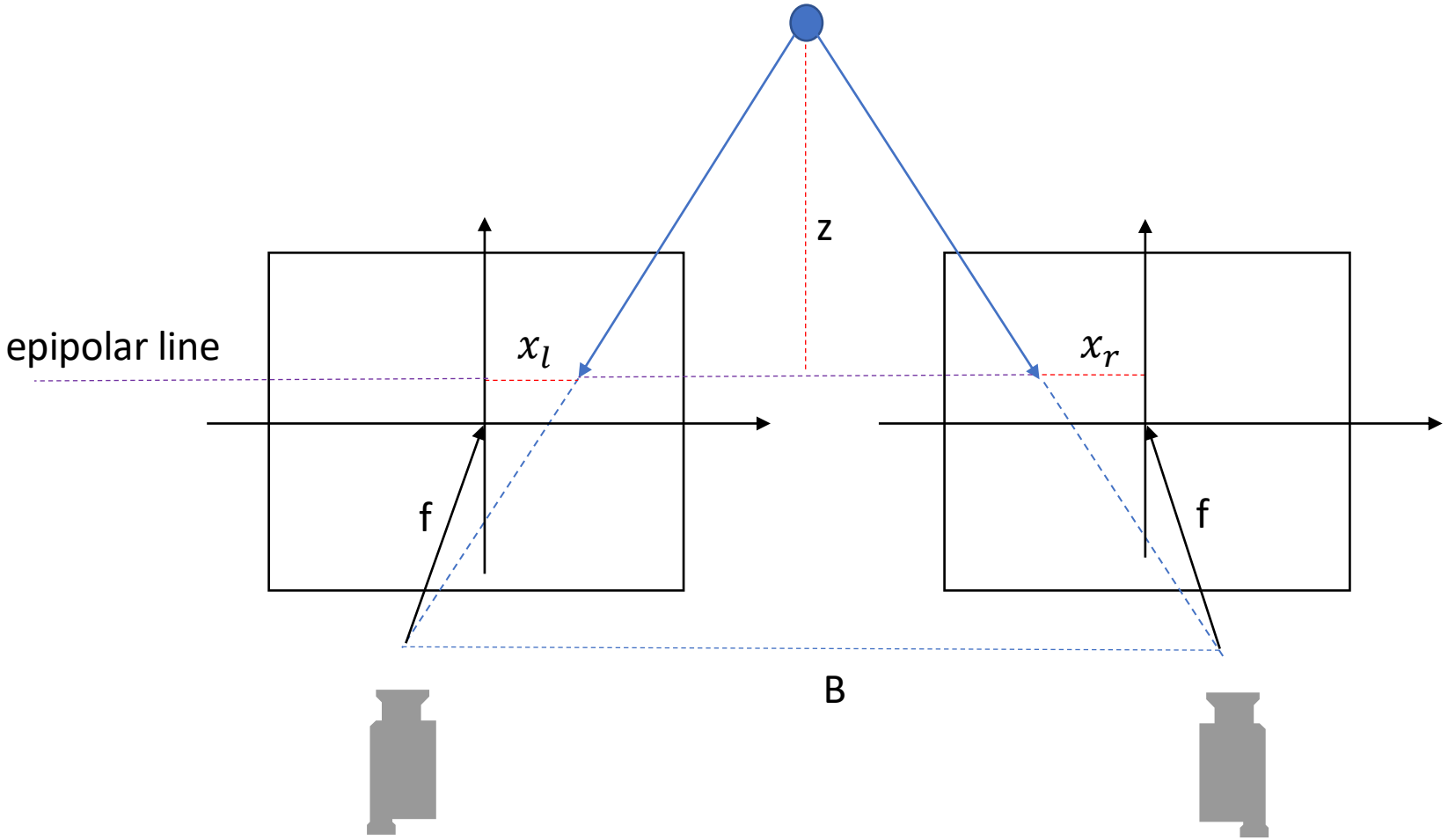
- Multiresolution strategy for real-time applications



Stereoscopic image pairs

- Depth and disparity

$$d = x_l - x_r = \frac{B \cdot f}{z}$$



Disparity estimation

- Find corresponding points in left and right images

$I_1(x_k, y_k)$ $I_2(x_k - d, y_k)$

Disparity estimation

- Local methods, feature matching
  - SSD (sum of squared difference), SIFT (scale invariant feature transform), Sobel operator, etc.

$$\text{SSD} = \sum_{(x_k,y_k)\epsilon\aleph_1((x,y))} \left(I_1(x_k, y_k) - I_2(x_k - d, y_k)\right)^2$$

- Pros & cons
  - low computational complexity
  - difficulty with similar texture, occlusion

Disparity estimation

- Global methods, MRF (Markov random field)
  - graph cut, belief propagation
  - ICM (iterated conditional modes), configuration of a local maximum of the joint probability of a MRF, by iteratively maximizing the probability of each variable conditioned on the rest.

- Pros & cons
  - data smoothness
  - high computational complexity

Disparity estimation

- Combined local and global method (CLG)
  - incorporate photo consistency and data smoothness
  - energy function

$$E\big(d(x,y)\big) = \iint_{\Omega} F(d)\,dx\,dy$$
$$= \iint_{\Omega} \big(\lambda_1 \cdot f_{data}(d) + \lambda_2 \cdot f_{smooth}(d)\big)\,dx\,dy$$

$$f_{data}\big(d(x,y)\big) = \frac{1}{2}\sum_{(x_k,y_k)\epsilon\aleph_1((x,y))}\big(I_1(x_k,y_k) - I_2(x_k - d, y_k)\big)^2$$

$$f_{smooth}(d) = \frac{1}{2}|\nabla d|^2 \ ,$$

$d(x,y)$	the estimated disparity at pixel (x,y)
λ_1, λ_2	the scaling factors
I_1, I_2	the left, right image

Disparity estimation

- Multiresolutional approach
  - supporting points at lower resolution
  - interpolation and refinement at higher resolution

- Challenges
  - spurious points at lower resolution
  - occlusion detection

Implementation

- Supporting points at lower resolution
  - down sampling
  - feature point selection
  - robust estimation, ICM

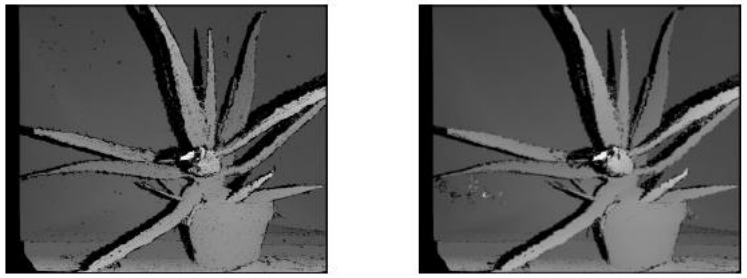
$$(d_1, d_2, \dots, d_n) = arg\ min(f_{data}(d))$$

$$s.t.\ \ f_{data}(d_1) \le f_{data}(d_2) \le \cdots \le f_{data}(d_n),$$

$$O(d_1) = \text{false}$$

$$O(d) = \begin{cases} \text{false}, & \text{if } f_{data} \le T_1 \\ \text{true}, & \text{otherwise} \end{cases}$$

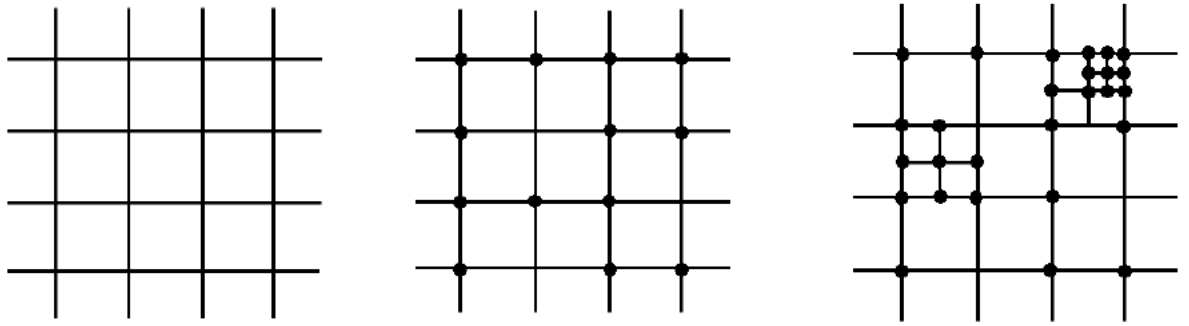
$$d = arg\ \ min_{(d_1, d_2, \dots, d_n)} F(d)$$



Aloe (1282x1110)

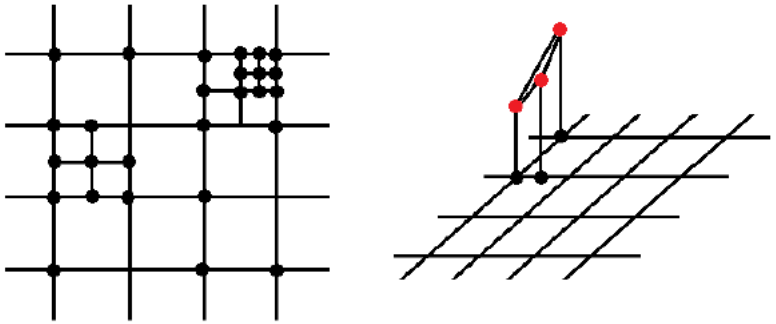
Implementation

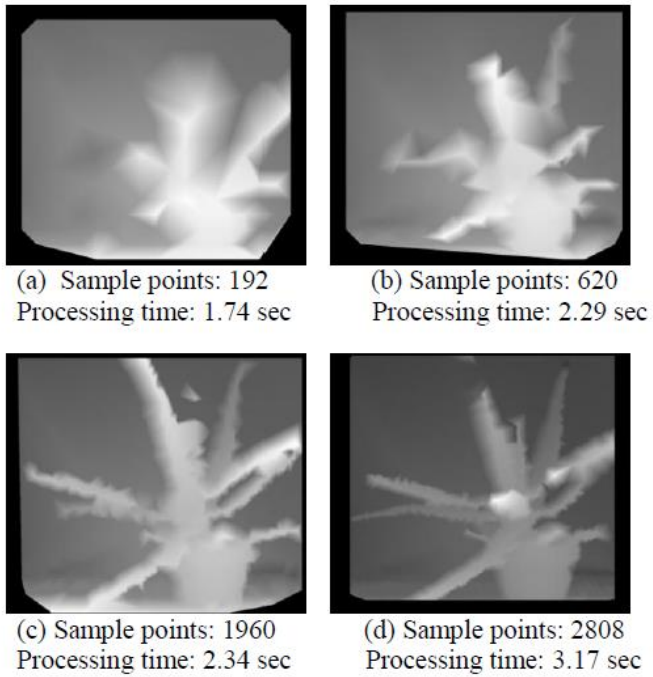
- Supporting points at lower resolution
  - adaptive estimation, non-uniform sampling



Implementation

- Interpolation at higher resolution
  - Delaunay triangulation





(a) Sample points: 192
Processing time: 1.74 sec

(b) Sample points: 620
Processing time: 2.29 sec

(c) Sample points: 1960
Processing time: 2.34 sec

(d) Sample points: 2808
Processing time: 3.17 sec

Figure 3: Sampling and disparity interpolation on left image, with different cell size, using uniform sampling with a (a) 64x64 cell size, (b) 32x32 cell size, (c) 16x16 cell size, and (d) nonuniform sampling with adaptive cell size ranging from 64x64 to 8x8.

Implementation

- Refinement at higher resolution
  - energy minimization, or MAP (maximum a posteriori) MRF problem

$$
\begin{aligned}
(D_l^*, O_l^*) &= arg\,max\big(p(I_r|I_l, D_l, O_l) \cdot p(I_l, D_l, O_l)\big) \\
&= arg\,max\Big(exp\big(-\rho_1(I_r|I_l, D_l, O_l)\big) \cdot exp\big(-\rho_2(I_l, D_l, O_l)\big)\Big) \\
&= arg\,min\Big(-log\big(exp\big(-\rho_1(I_r|I_l, D_l, O_l)\big)\big) - \\
&\qquad log\Big(exp\big(-\rho_2(I_l, D_l, O_l)\big)\Big)\Big) \\
&= arg\,min\ \big(\textstyle\sum_{x\in I_l} \rho_1\big(I_r\big(x - D_l(x)\big), I_l(x), O_l\big) + \\
&\qquad \textstyle\sum_{x\in I_l} \sum_{x'\in \aleph_x} \rho_2(D_l(x), D_l(x'), O_l)\big)
\end{aligned}
$$

$$\rho_1(D) = \lambda_1 \cdot f_{data}\left(|I_r(x - D) - I_l(x)|\right)$$

$$\rho_2(D) = \lambda_2 \cdot f_{smoot\,h}\big(D, D(x')\big)$$

Implementation

- Refinement at higher resolution
  - iterative Gauss-Seidel method
  - bilateral filter

$$\frac{\partial F}{\partial d} = \lambda_1 \cdot \sum_{(x_k,y_k)\epsilon\aleph_1((x,y))} I_{2x}(x_k - d, y_k)\big(I_1(x_k, y_k) - I_2(x_k - d, y_k)\big) - \lambda_2 \cdot \Delta d = 0$$

$$\lambda_1 \cdot \sum_{(x_k,y_k)\epsilon\aleph_1((x,y))}\big(I_2(x_k - d, y_k) - I_2(x_k - 1 - d, y_k)\big)\big(I_1(x_k, y_k) - I_2(x_k - d, y_k)\big) - \lambda_2 \cdot \sum_{d_j\epsilon\aleph_2(d)}\big(d_j - d\big) = 0$$

$$d_i^{t+1} = \frac{\lambda_1}{|\aleph_2|}\sum_{(x_k,y_k)\epsilon\aleph_1((x,y))}\big(I_2(x_k - 1 - d_i^t, y_k) - I_2(x_k - d_i^t, y_k)\big)\big(I_1(x_k, y_k) - I_2(x_k - d_i^t, y_k)\big) + \frac{\lambda_2}{|\aleph_2|}\Big(\sum_{d_j\epsilon{\aleph_2}^-(d_i)} d_j^{t+1} + \sum_{d_j\epsilon{\aleph_2}^+(d_i)} d_j^t\Big)$$

I_{2x}	the derivative of the feature response in I_2
Δ	the Laplacian operator
${\aleph_2}^-(d_i)$, ${\aleph_2}^+(d_i)$	the neighboring pixels processed before and after pixel i

Computation time modeling

- Performance influential parameters

$$T(S_{max}, S_{min}, I_{no}, DR) = \frac{c_0}{DR^2} \cdot \left(\frac{e^{-c1} \cdot c_2}{{S_{min}}^2} + \frac{(1-e^{-c1}) \cdot c_3}{{S_{max}}^2} + I_{no} \right)$$

S_{max} : maximum cell size
S_{min} : minimum cell size
I_{no} : iteration number
DR : down-sampling rate

Experimental results

- Mismatch rate (MR) and computation time
  - computer configuration, 2.66GHz CPU, 4GB RAM
  - Middlebury dataset

	Cones (900x750)		*Teddy* (900x750)		*Aloe* (1282x1110)	
	M.R.	Points	M.R.	Points	M.R.	Points
8x8	6.2	1497	7.6	1247	9.9	2503
16x16	6.7	541	7.8	405	10.1	864

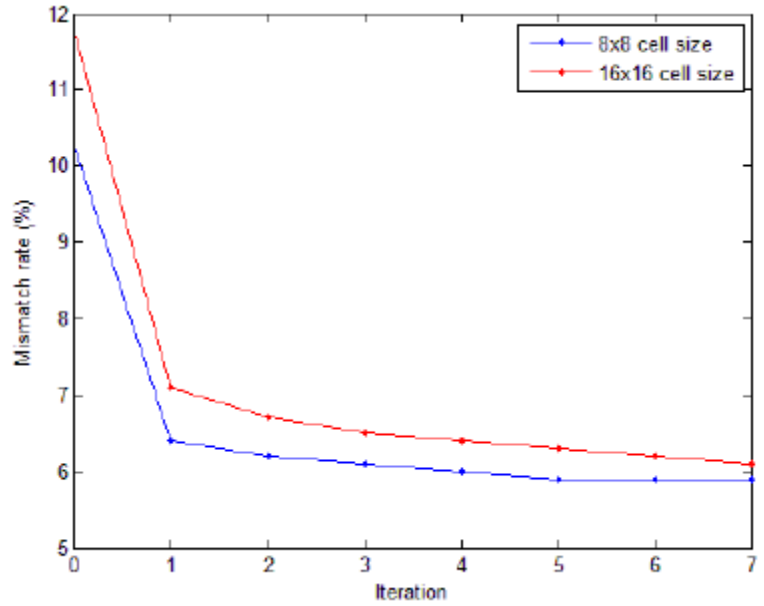
Table 1 Mismatch rate (%) and the number of supporting points.

	Supp.	Tri.	Iter.1	Iter.2	Iter.3	Iter.4	Iter.5
8x8	489	158	57	52	52	52	52
16x16	383	124	62	53	52	52	53

Table 2 Processing time (ms) for different phases: computing supporting points, triangulation interpolation, 1st iteration, 2nd iteration, 3rd iteration, 4th iteration, and 5th iteration.

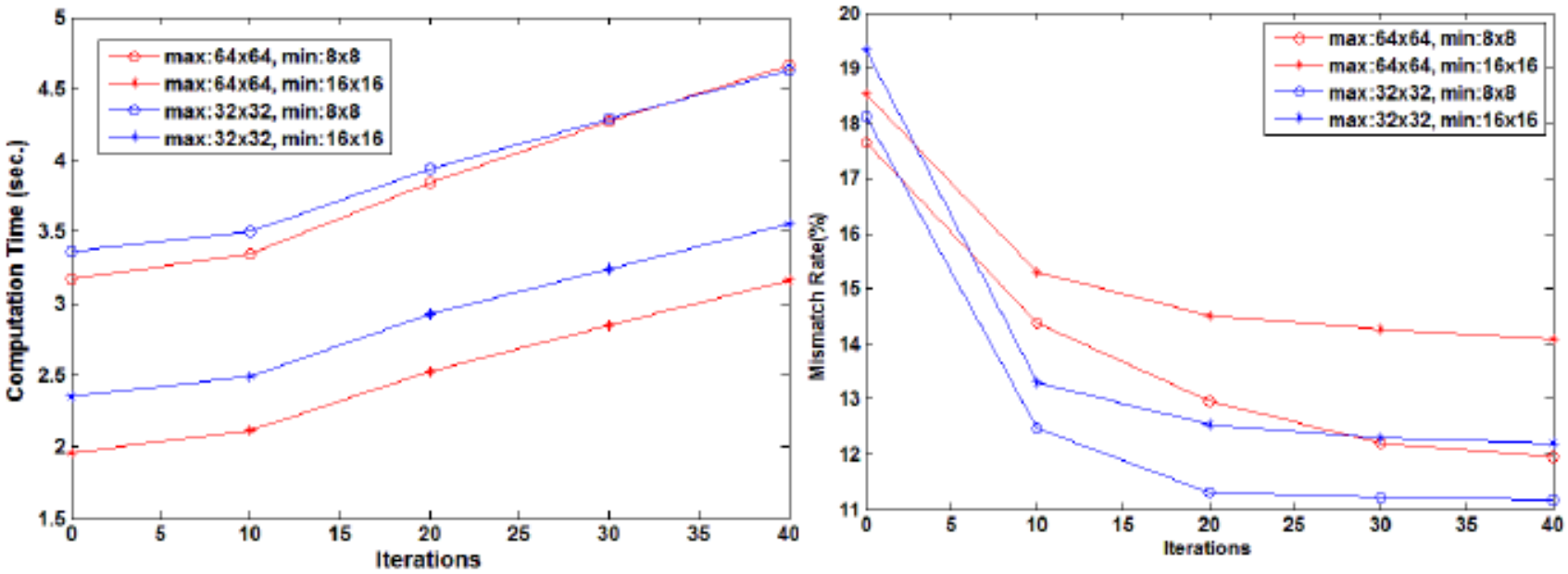
Experimental results

- Mismatch rate and iteration number



Experimental results

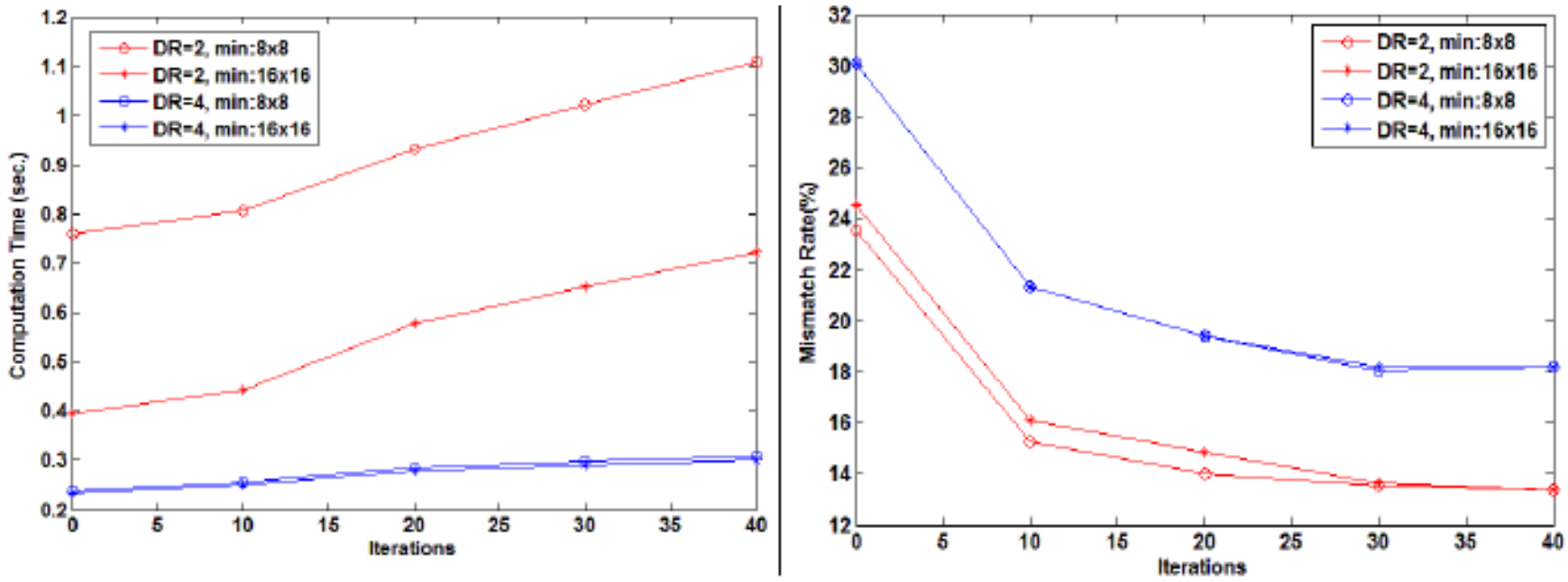
- Cell size and iteration number



(a) Computation time and mismatch rate with different cell size range (DR=1).

Experimental results

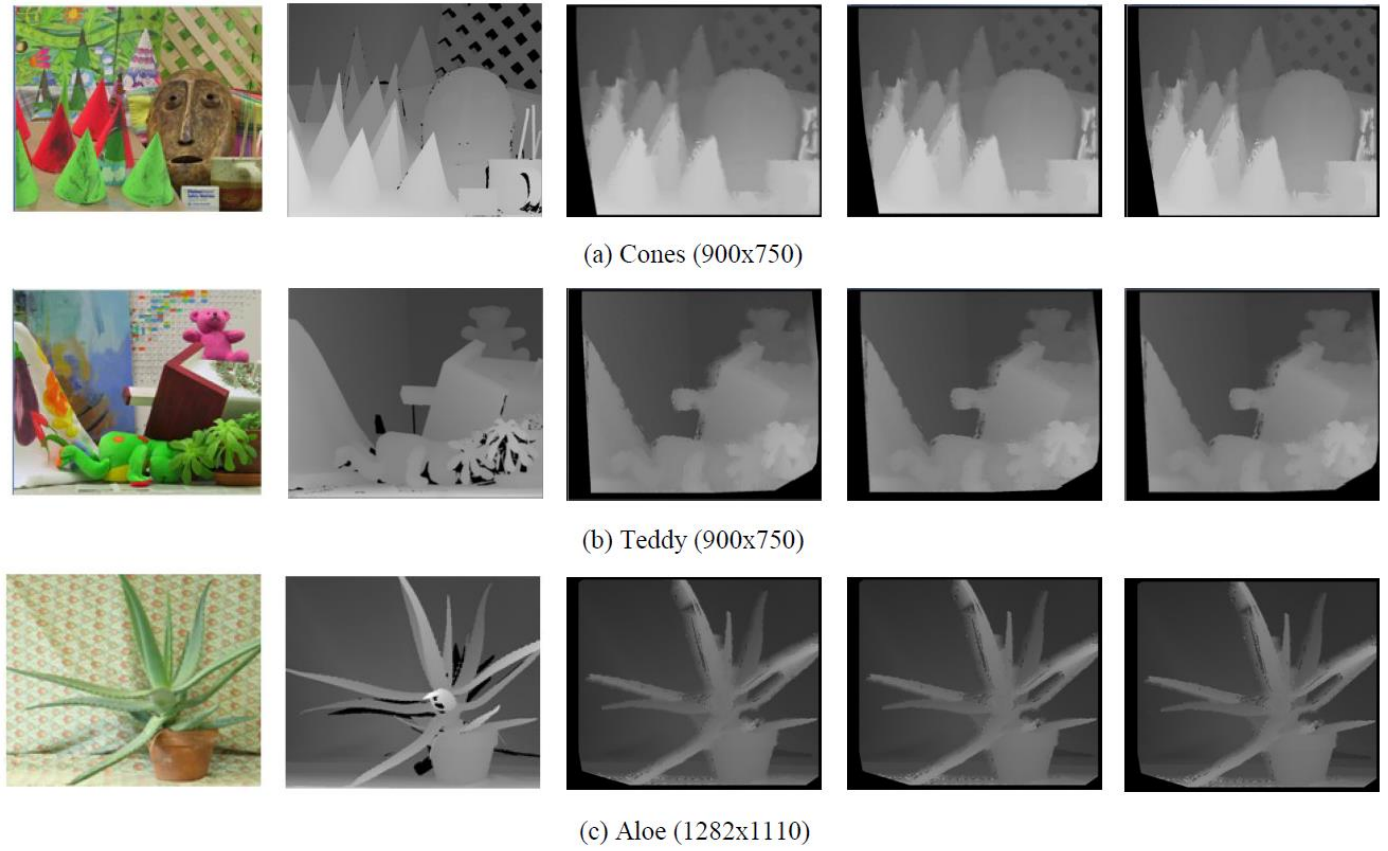
- Down-sampling rate and iteration number



(b) Computation time and mismatch rate with different down-sampling rate (maximum cell size: 32x32).

Experimental results

- Visual effect with different iteration numbers



(a) Cones (900x750)

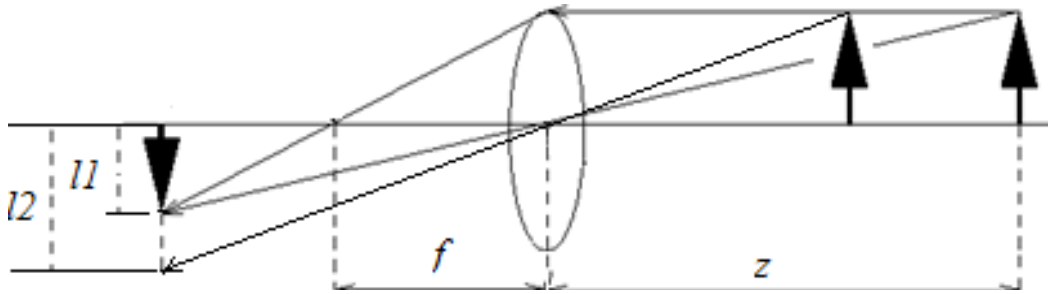
(b) Teddy (900x750)

(c) Aloe (1282x1110)

Figure 4: Disparity estimation. From left to right: the left image, ground truth, initial estimation, 1st iteration, 2nd iteration.

Application in real-time object tracking

- PTU (pan-tilt unit) camera tracking
  - mean-shift algorithm, kernel filtering based on color distribution
  - tracking window

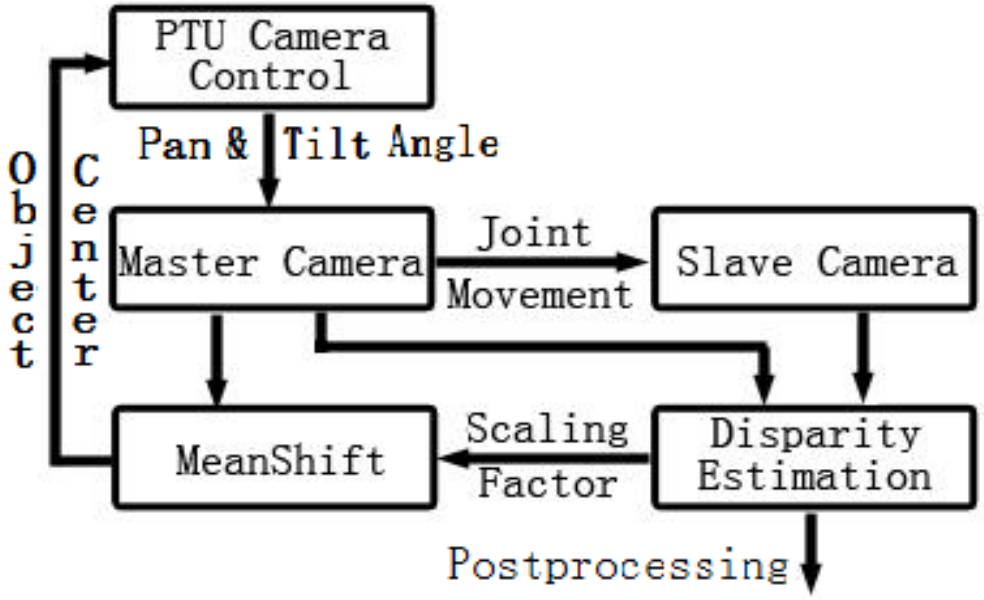


$$l1/l2 = z2/z1 = d1/d2$$

$l1, l2$	the edge length of the tracking window at two consecutive updates
z	the depth of the object
d	the average estimated disparity for the object region

Application in real-time object tracking

- PTU (pan-tilt unit) camera tracking
  - adjustable window size according to disparity information



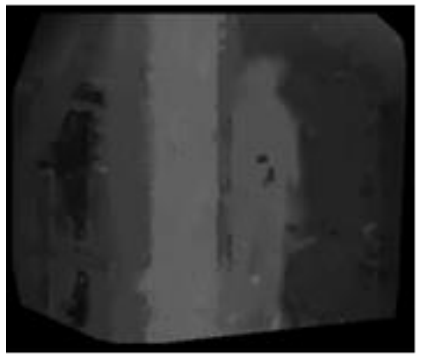
Application in real-time object tracking

- Experiment setup
  - two PointGrey Firefly MV CMOS cameras mounted on PTU, connected to a computer via a 1394 firewire USB2.0 hub
  - 640x480 video at a frame rate of 15 fps
  - average 178 ms per frame on disparity estimation

Application in real-time object tracking

- PTU (pan-tilt unit) camera tracking with adjustable window size






Conclusions

- Combined local and global method
- Multiresolution processing
- Accuracy at lower resolution
- Time-MR performance influential parameters
- Other clues

References

1. Y. Ye, S. Ci, A. K. Katsaggelos and Y. Liu, "A Multi-camera Motion Capture System for Remote Healthcare Monitoring," The IEEE International Conference on Multimedia and Expo, July 2013.
2. Y. Ye, S. Ci, Y. Liu, H. Wang, and A. K. Katsaggelos, "Binocular Video Object Tracking with Fast Disparity Estimation," the IEEE International Conference on Advanced Video and Signal-Based Surveillance, 2013.
3. Y. Liu, S. Ci, H. Tang, Y. Ye, and J. Liu, "QoE-oriented 3D Video Transcoding for mobile streaming," ACM Transactions on Multimedia Computing, Communications and Applications, Volume 8, Issue 3s, article 42, 2012.
4. Y. Ye, S. Ci, Y. Liu, and H. Tang, "Dynamic Video Object Detection with Single PTU Camera," Visual Communications and Image Processing, Nov 2011.
5. A. Geiger, M. Roser and R. Urtasun, "Efficient Large-Scale Stereo Matching", ACCV2010, pp. 25-38.
6. B.M. Smith, Li Zhang, and Hailin Jin, "Stereo matching with nonparametric smoothness priors in feature space," CVPR2009, pp.485-492.
7. S. Kosov, T. Thormahlen, and H. P. Seidel, "Accurate Real-Time Disparity Estimation with Variational Methods," ISVC '09 Proceedings of the 5th International Symposium on Advances in Visual Computing: Part I, pp. 796-807.
8. D. Comaniciu, V. Ramesh, and P. Meer, "Kernel-based object tracking," TPAMI, vol.25, no.5, pp. 564- 577, May 2003.
9. J. Besag, "On the statistical analysis of dirty pictures," J. R. Stat. Soc. B, vol. 48, pp. 259-502, 1986.
10. D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," IJCV, 47(1-3):7–42, 2002.