the City College of New York
OF THE CITY UNIVERSITY OF NEW YORK

cuny.edu
THE CITY UNIVERSITY OF NEW YORK

# 3D SCENE MODELING AND UNDERSTANDING FROM IMAGE SEQUENCES

Thesis defense for the Degree of Doctor of Philosophy at CUNY

**Hao Tang**

**Committee members**

Professor Zhigang Zhu (mentor), City College of New York
Professor Jizhong Xiao, City College of New York
Professor Ioannis Stamos, Hunter College
Dr. Rakesh (Teddy) Kumar, SRI International - Sarnoff

# Outline of the Presentation

- **Introduction**
- **Related Work**
- **Research Topics and Methodology**
- **Summary and Future Work**

# Outline of the Presentation

- **Introduction**
  - Problem statement
  - Thesis statement and main contributions
- **Related Work**
- **Research Topics and Methodology**
- **Summary and Future Work**

# Thesis statement

- **Reconstructing**, **representing** and **labeling** large-scale 3D (**urban**) scenes from image sequences

- Some major challenges
  - **Computational issue**: large amount of data
  - **Performance issue**: accurate 3D modeling
  - **Representation issue**: Integration of many images with small FOV

# Problem statement

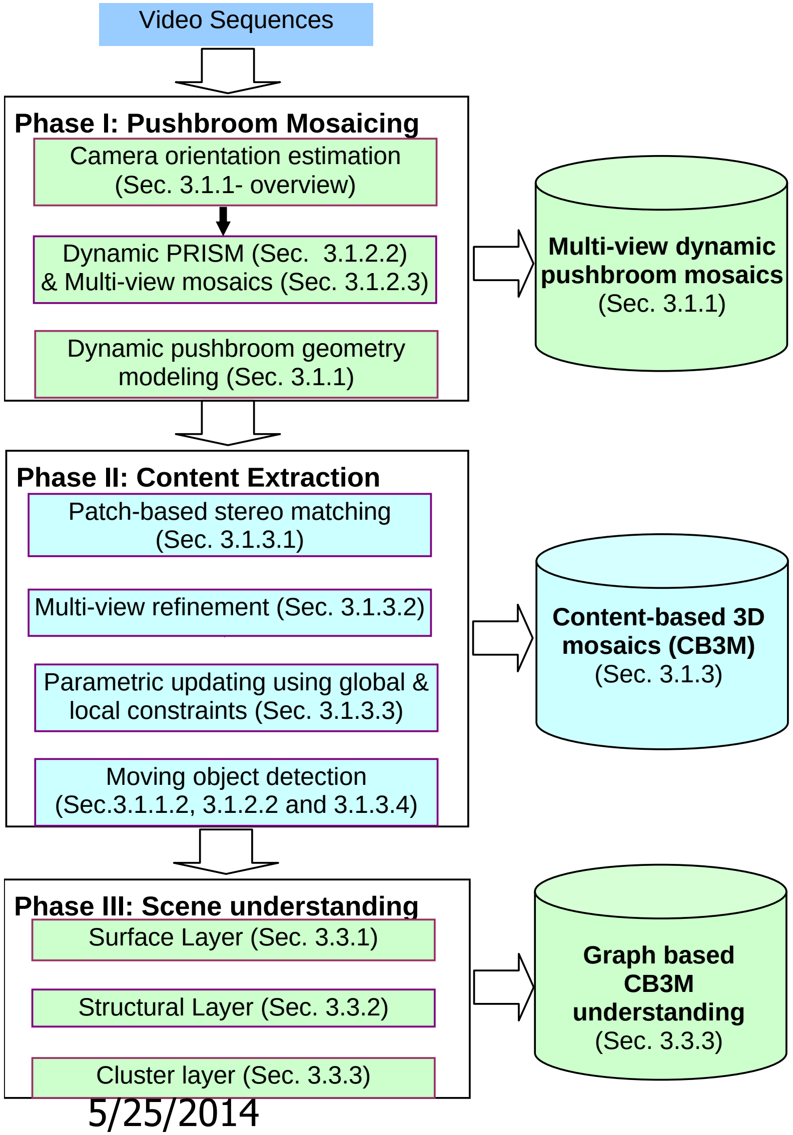
- A mosaic based method is proposed (Zhu et al. 2004)
  - Multi-view pushbroom mosaics
    - large FOV stereo viewing
  - Depth error is linearly proportional to depth
    - Theoretical foundation for 3D from stereo mosaics

- Many open problems remain:
  - Robust and efficient stereo matching for urban scenes
    - large textureless regions => feature detection/matching
    - Cluttered scenes => occlusion handling
    - Moving objects => 3D background alignment
  - Higher-level scene understanding
    - compact and content-based data representation
    - Higher-level scene labeling into meaningful regions/clusters

# Thesis contributions

- Model:
  - Extend the previous work on stereo mosaics from static 3D scenes to **dynamic** 3D scenes
- Algorithm:
  - An effective and efficient **patch-based stereo matching** algorithm
- Experimental analysis:
  - Thorough **experimental analysis** of the robustness and accuracy of parallel-perspective stereo mosaics
- 3D understanding:
  - A **graph-based higher-level scene labeling** approach
- patch-based method in different geometry:
  - Used to **transduce and highlight** the important/highlighted information to visually impaired

# Work Flow



Video Sequences

**Phase I: Pushbroom Mosaicing**

Camera orientation estimation (Sec. 3.1.1- overview)

Dynamic PRISM (Sec. 3.1.2.2) & Multi-view mosaics (Sec. 3.1.2.3)

Dynamic pushbroom geometry modeling (Sec. 3.1.1)

**Multi-view dynamic pushbroom mosaics** (Sec. 3.1.1)

**Phase II: Content Extraction**

Patch-based stereo matching (Sec. 3.1.3.1)

Multi-view refinement (Sec. 3.1.3.2)

Parametric updating using global & local constraints (Sec. 3.1.3.3)

Moving object detection (Sec.3.1.1.2, 3.1.2.2 and 3.1.3.4)

**Content-based 3D mosaics (CB3M)** (Sec. 3.1.3)

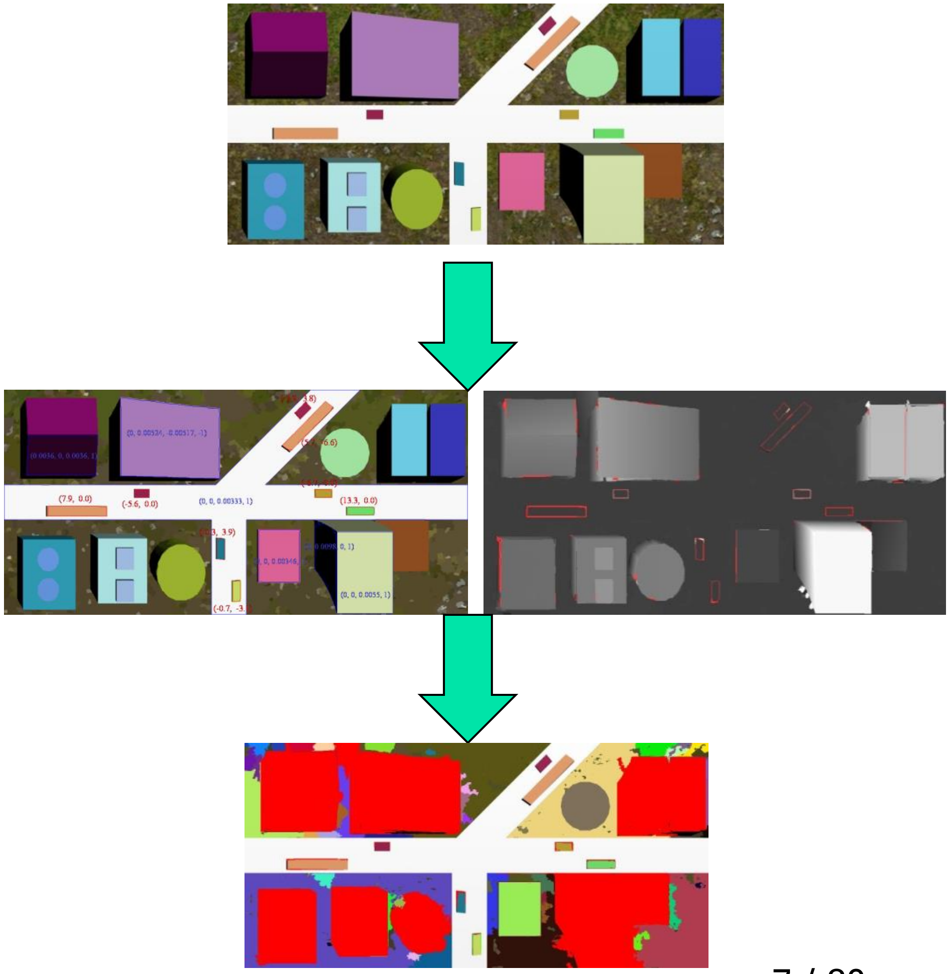
**Phase III: Scene understanding**

Surface Layer (Sec. 3.3.1)

Structural Layer (Sec. 3.3.2)

Cluster layer (Sec. 3.3.3)

**Graph based CB3M understanding** (Sec. 3.3.3)



# Outline of the Presentation

- **Introduction**
- **Related Work**
  - Mosaics representations
  - Stereo matching
  - 3D modeling from video mosaics
  - Simultaneous localization and mapping (SLAM)
- **Research Topics and Methodology**
- **Summary and Future Work**

# Related Work

- Mosaics representations
  - Video Mosaics
    - Irani, et al, 1996; Hsu & Anandan, 1996; Sawhney et al, 1998; Odone, et al, 2000; Leung & Chen, 2000, Cai et al, 2010, Vivet et al. 2011
  - Pushbroom mosaic
    - Gupta & Hartley, 1997, Chai & Shum, 2000; Zhu, et al, 2004

# Related Work

- Stereo matching (Scharstein &. Szeliski, 2002)
  - Adaptive window method
    - Kanade & Okutomi, 1991, Fusiello, et al, 1997
  - Color segmentation
    - Tao, et al, 2001
  - Layer based segmentation
    - Ke & Kanade, 2001, Xiao & Shah 2004
  - Energy minimization (graph cuts & belief propagation)
    - Boykov, et al, 2001; Kolmogorov & Zabih, 2001, Sun et al. 2005
  - Large amount of real data
    - Pollefeys, et al 2008, Cornelis et al 2008

# Related Work

- Large scale 3D modeling from video mosaics
  - Rav-Acha et al. 2008
  - Zheng and Shi 2008

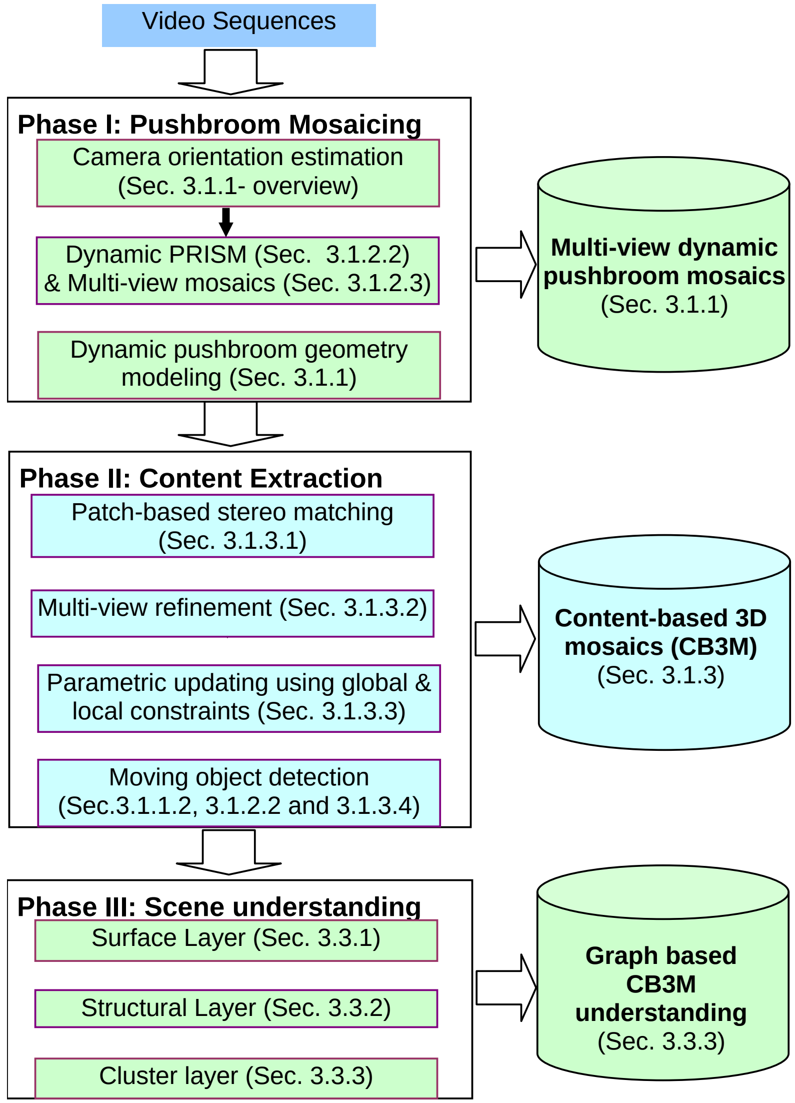
# Related Work

- Simultaneous localization and mapping (SLAM)
  - Davison and Murray 2001; Davison et al. 2007, Oskiper et al. 2007, Doucet and Johansen 2008
  - Nistér et al. 2004, Mouragnon et al. 2006, Klein and Murray 2007, Newcombe and Davision 2010

# Outline of the Presentation

- **Introduction**
- **Related Work**
- **Research Topics and Methodology**
  - 3D modeling from video – **a mosaic based approach and the core algorithms**
  - **Scene understanding /labeling** from Content Based 3D Mosaics
  - 3D modeling–the core algorithms **extended to perspective stereo** images
- **Summary and Future Work**

# Work Flow



Video Sequences

**Phase I: Pushbroom Mosaicing**

Camera orientation estimation (Sec. 3.1.1- overview)

Dynamic PRISM (Sec. 3.1.2.2) & Multi-view mosaics (Sec. 3.1.2.3)

Dynamic pushbroom geometry modeling (Sec. 3.1.1)

**Multi-view dynamic pushbroom mosaics** (Sec. 3.1.1)

**Phase II: Content Extraction**

Patch-based stereo matching (Sec. 3.1.3.1)

Multi-view refinement (Sec. 3.1.3.2)

Parametric updating using global & local constraints (Sec. 3.1.3.3)

Moving object detection (Sec.3.1.1.2, 3.1.2.2 and 3.1.3.4)

**Content-based 3D mosaics (CB3M)** (Sec. 3.1.3)

**Phase III: Scene understanding**

Surface Layer (Sec. 3.3.1)

Structural Layer (Sec. 3.3.2)

Cluster layer (Sec. 3.3.3)

**Graph based CB3M understanding** (Sec. 3.3.3)

# Outline of the Presentation

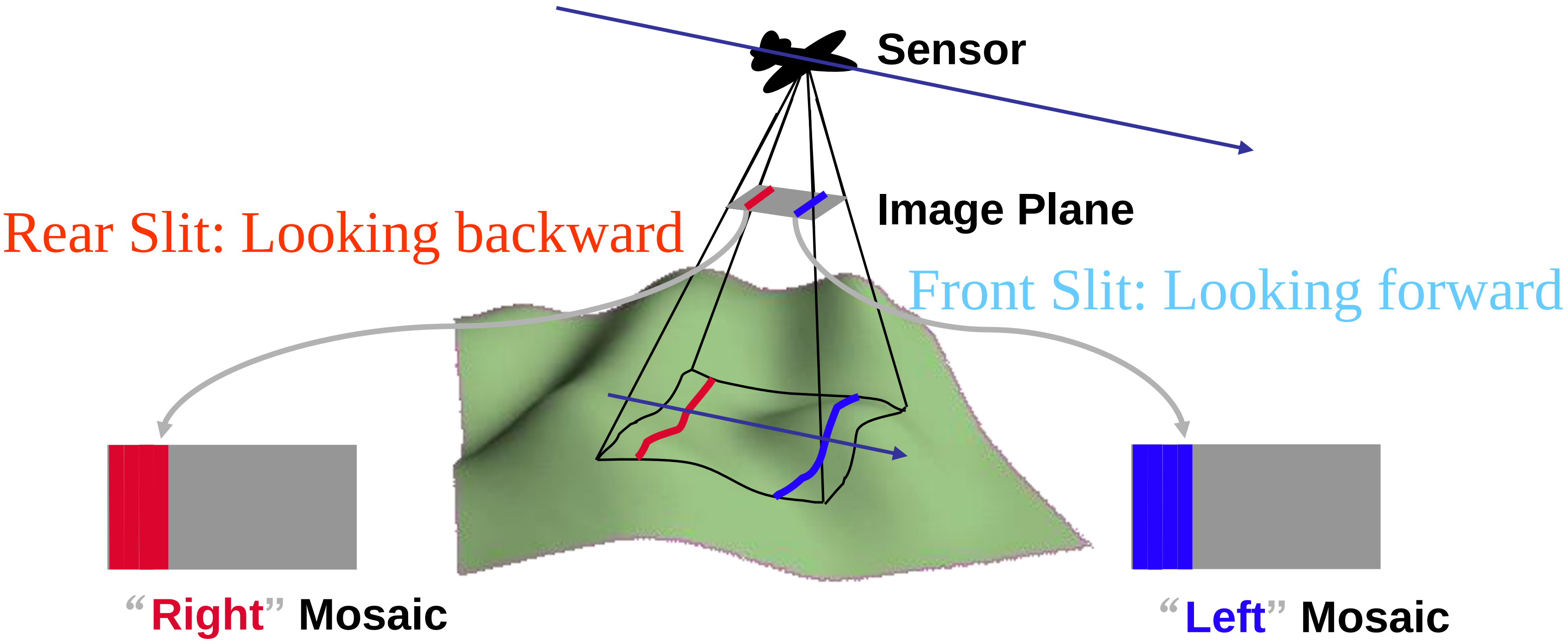
- **Introduction**
- **Related Work**
- **Research Topics and Methodology**
  - 3D modeling from video – **a mosaic based approach and the core algorithms**
  - **Scene understanding /labeling** from Content Based 3D Mosaics
  - 3D modeling from images – the core algorithms **extended to perspective stereo** images
- **Summary and Future Work**

# 3D modeling from video – a mosaic based method using pushbroom geometry

- ➡ Dynamic Pushbroom Stereo Mosaic Geometry
- 3D and Motion Content Extraction
- Content-Based 3D Mosaics (CB3M)
- Experimental Results and analysis

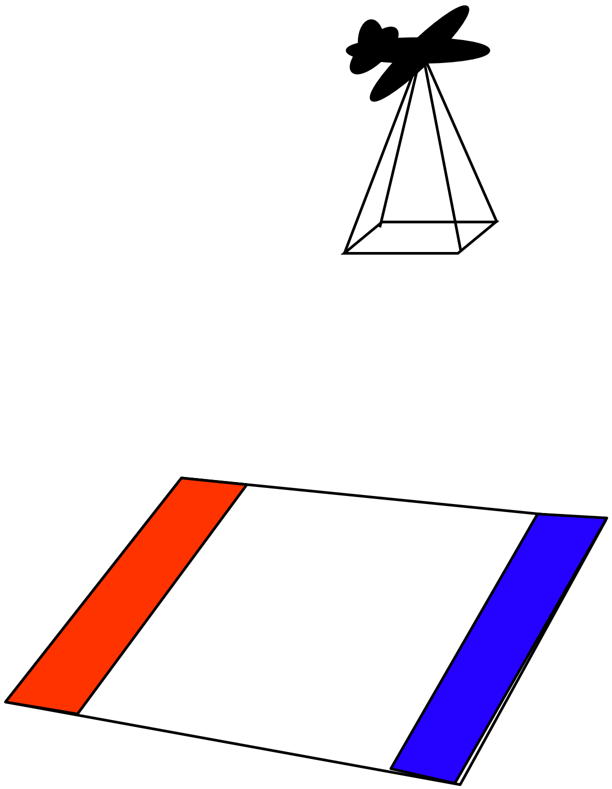
# Dynamic Pushbroom Stereo Mosaic Geometry

- Ideal case: Sensor motion is pure translation



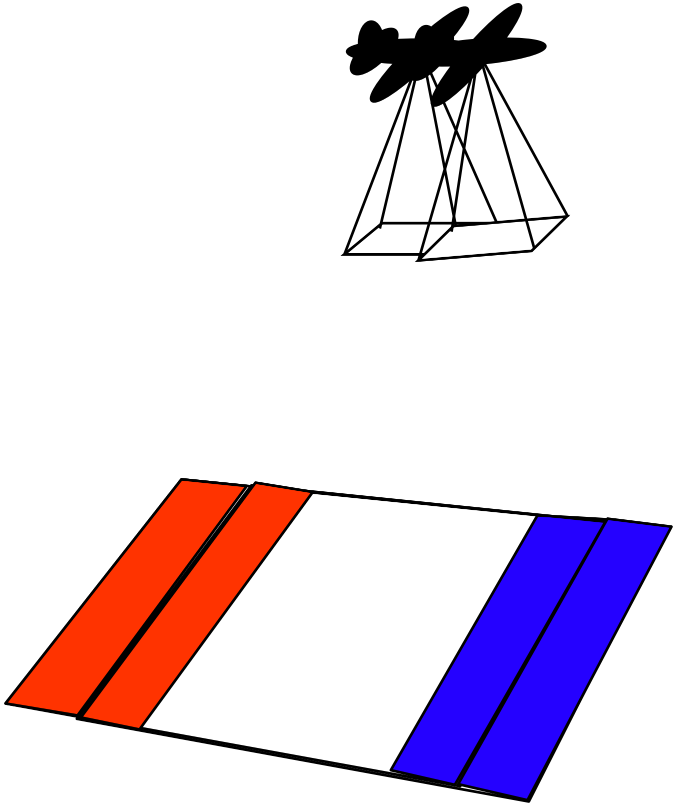
# Dynamic Pushbroom Stereo Mosaic Geometry

Stereo pair of two sets of parallel rays with large FOVs and adaptive baselines



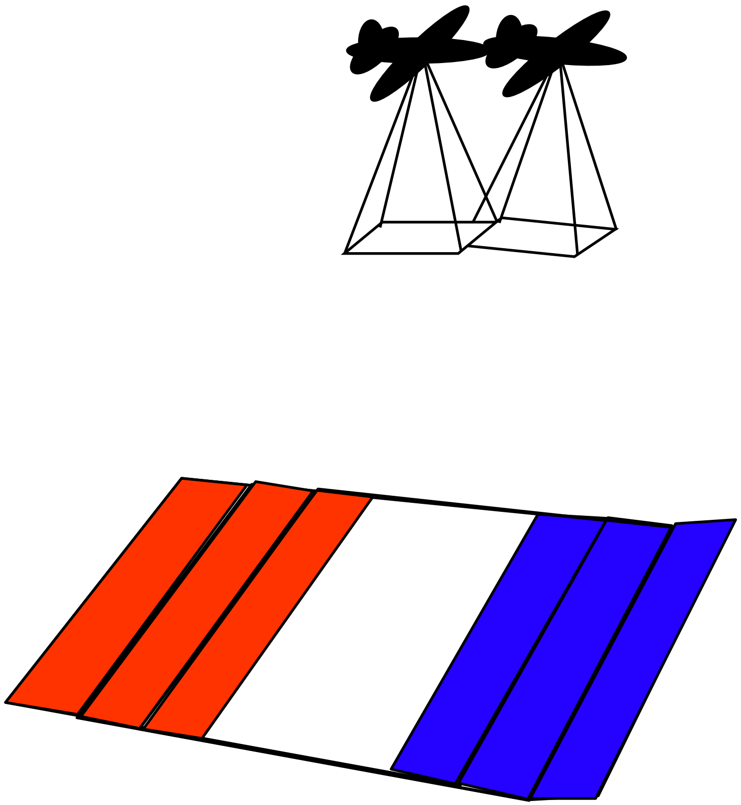
# Dynamic Pushbroom Stereo Mosaic Geometry

Stereo pair of two sets of parallel rays with large FOVs and adaptive baselines



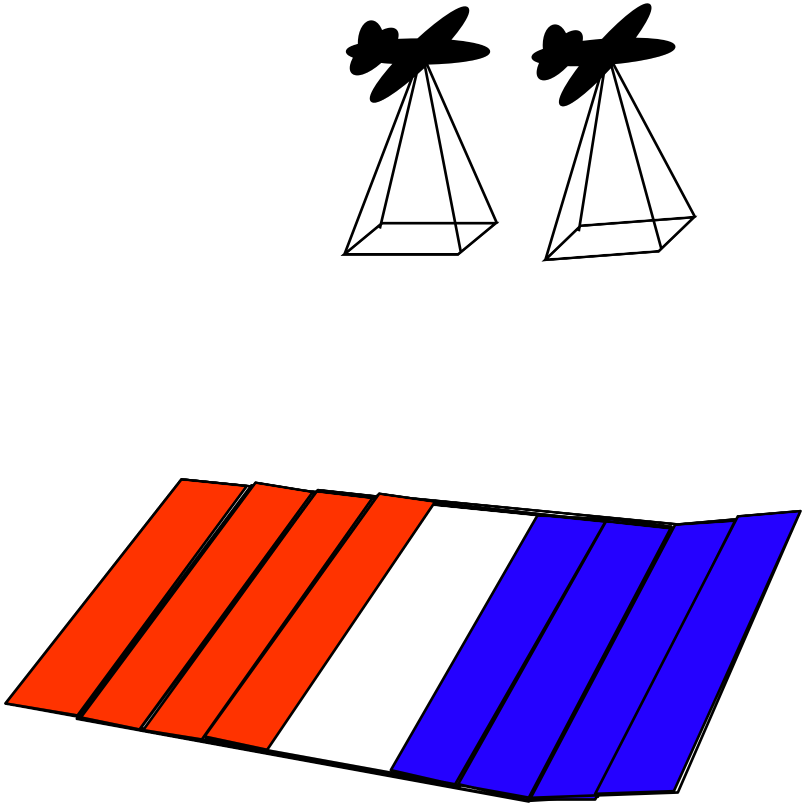
# Dynamic Pushbroom Stereo Mosaic Geometry

Stereo pair of two sets of parallel rays with large FOVs and adaptive baselines



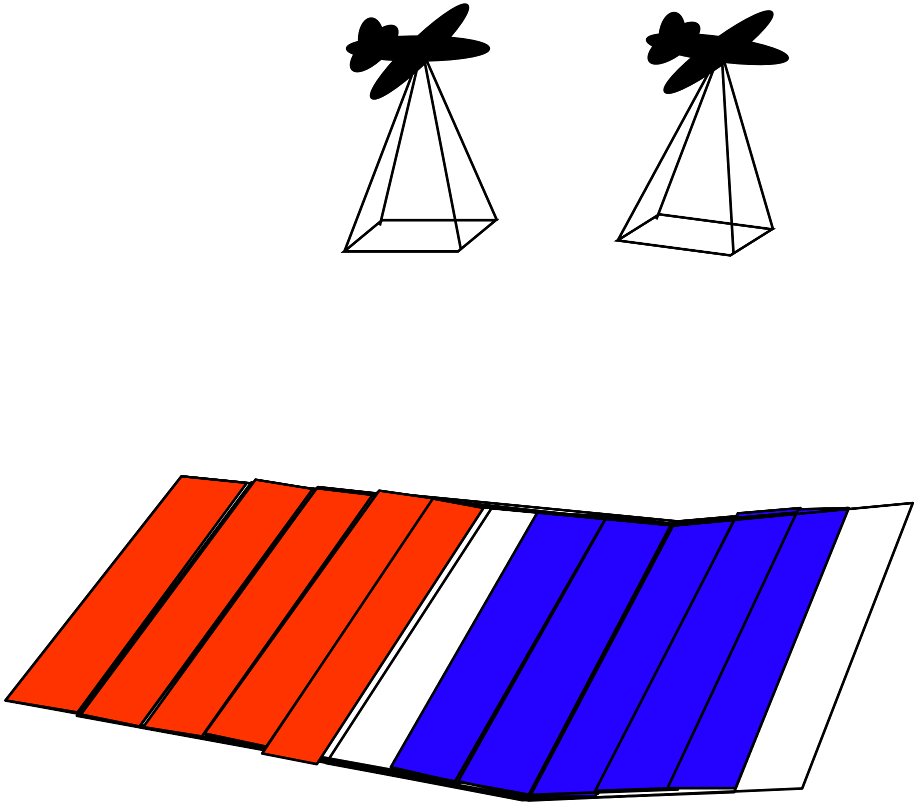
# Dynamic Pushbroom Stereo Mosaic Geometry

Stereo pair of two sets of parallel rays with large FOVs and adaptive baselines



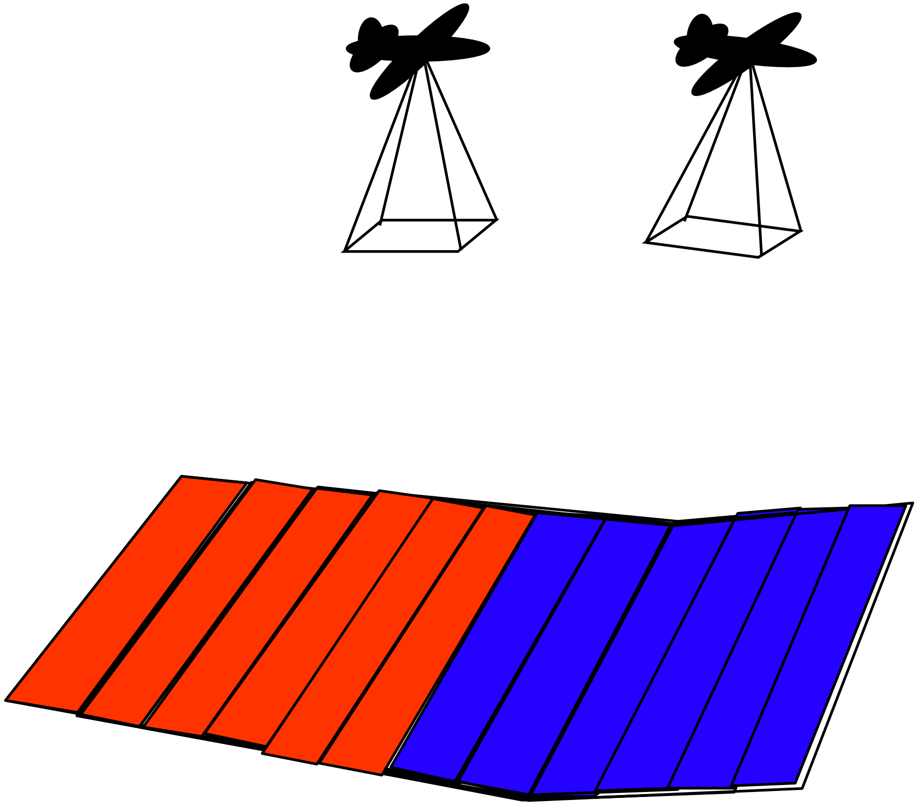
# Dynamic Pushbroom Stereo Mosaic Geometry

Stereo pair of two sets of parallel rays with large FOVs and adaptive baselines



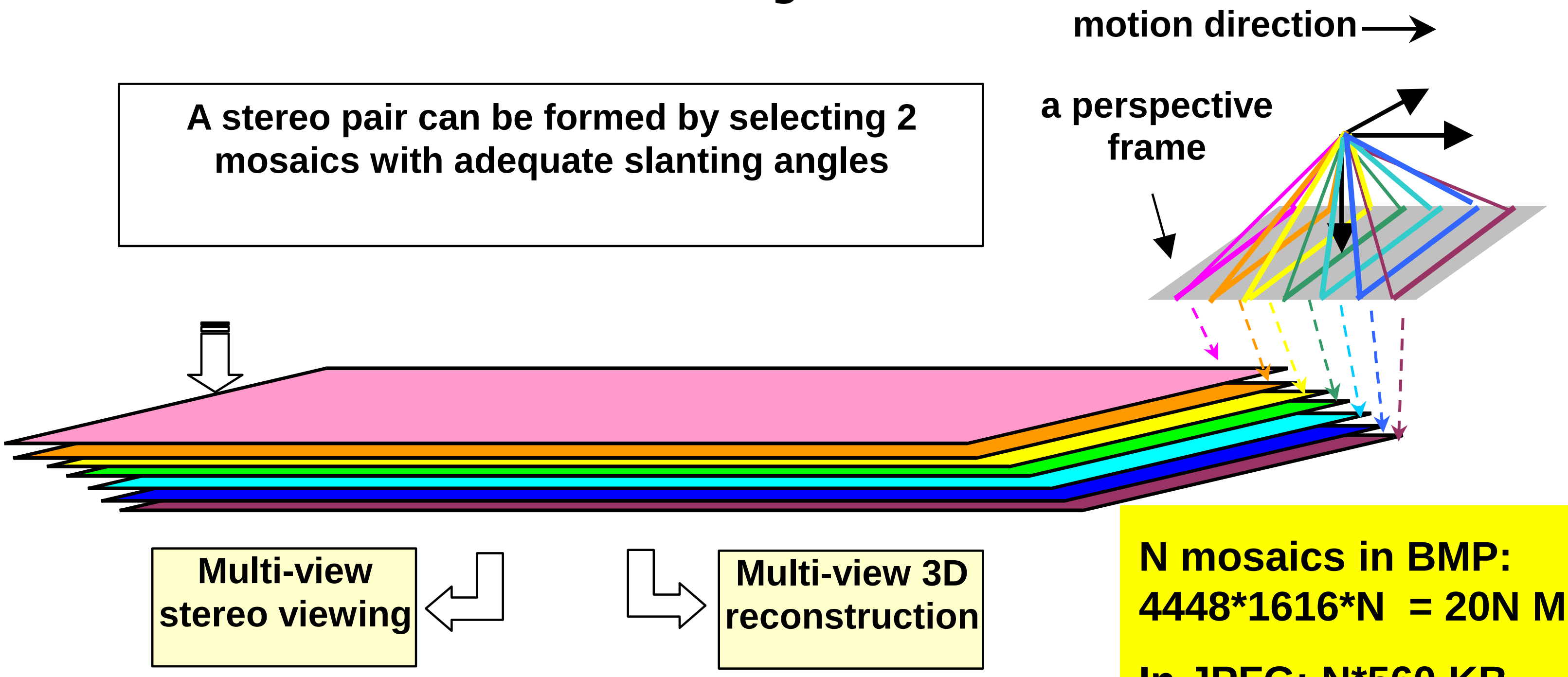
# Dynamic Pushbroom Stereo Mosaic Geometry

Stereo pair of two sets of parallel rays with large FOVs and adaptive baselines



# Multiple View Pushbroom Mosaics

- Multi-disparity stereo:
  - Correspondence for 3D reconstruction
- Mosaic-based rendering without 3D
  - View selection and rendering



motion direction

a perspective frame

A stereo pair can be formed by selecting 2 mosaics with adequate slanting angles

Multi-view stereo viewing

Multi-view 3D reconstruction

**N mosaics in BMP:**
**4448*1616*N = 20N MB**

**In JPFG: N*560 KB**

# Recovering depth from mosaics

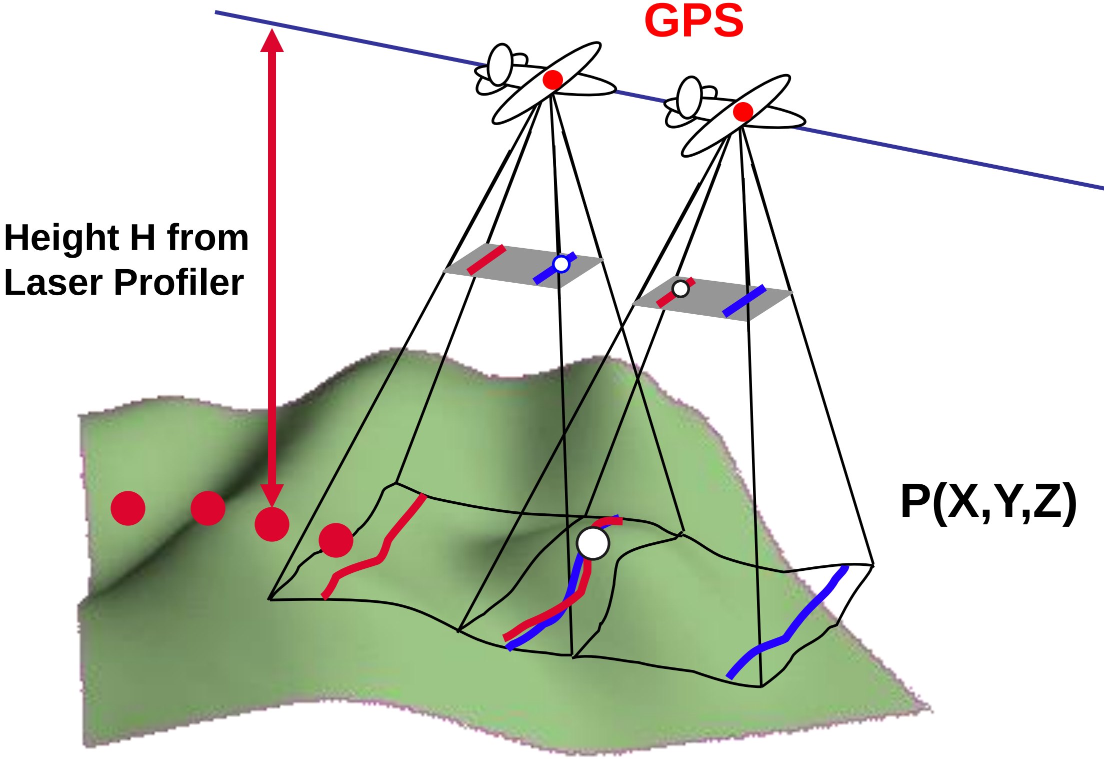
- **Pushbroom stereo mosaics**
  - **parallel-perspective**

**Depth accuracy (relative error) independent of depth**

baseline displacement

$$Z = F\frac{B_y}{d_y} = H + H\left(\frac{\Delta y}{d_y}\right)$$

disparity Fixed !



Two views from different perspective ⇒ stere

# Dynamic pushbroom stereo geometry

- Static scene
  - Δy visual motion

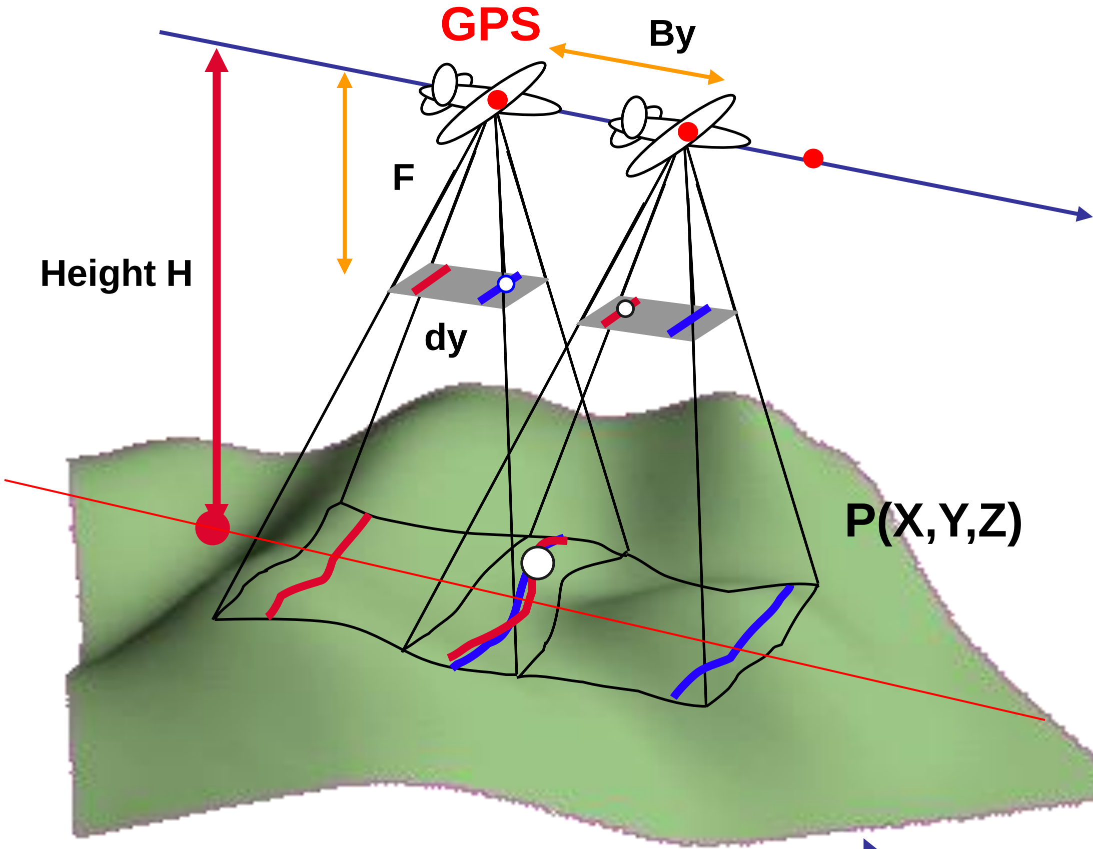
baseline

displacement

$$Z = F\frac{B_y}{d_y} = H(\frac{d_y + \Delta y}{d_y})$$

disparity

Fixed !



Two views from different perspective ➡ stereo

# Dynamic pushbroom stereo geometry

- Dynamic scene
  - $\Delta y$ visual motion in stereo
  - $s_y$ target motion in y

baseline

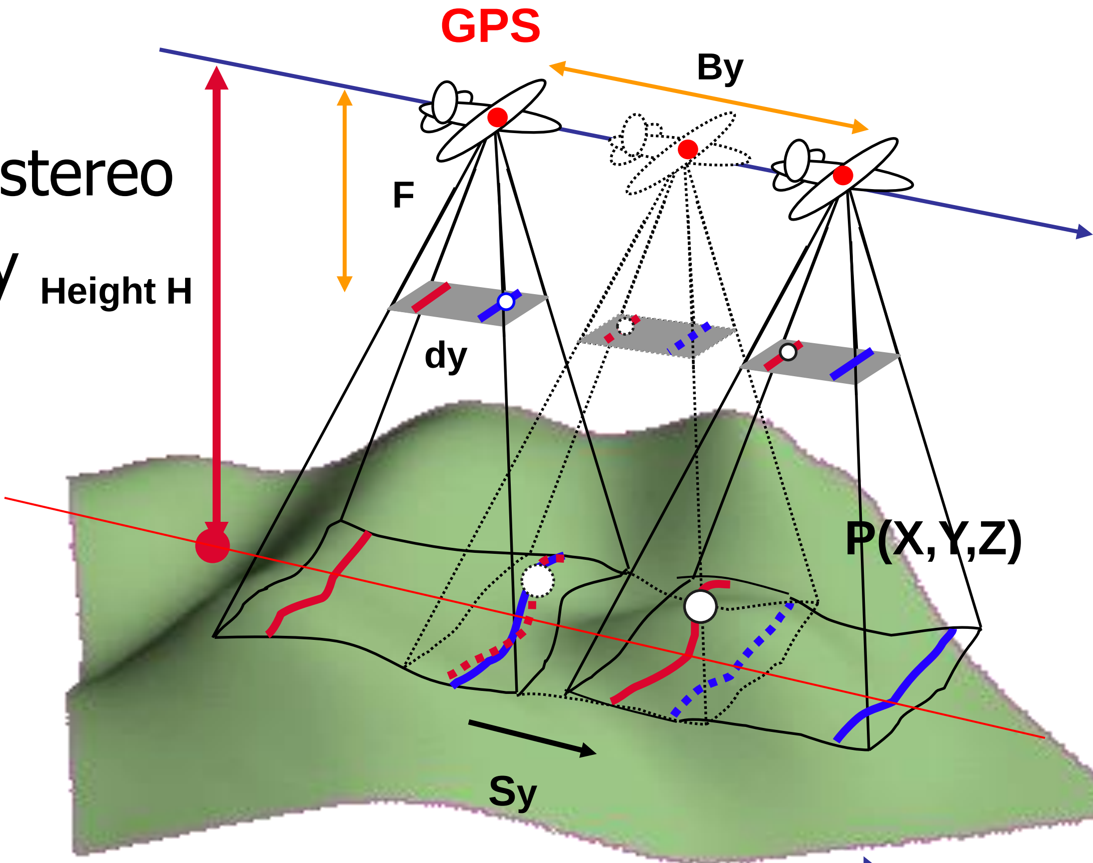
displacement

$$Z = F\frac{B_y - S_y}{d_y} = H\left(\frac{d_y + \Delta y - s_y}{d_y}\right)$$

disparity

Fixed !

$s_y = F\,S_y / H$



Two views from different perspective ⇒ stereo

# Dynamic pushbroom stereo geometry

- Dynamic scene
  - $\Delta y$ visual motion in stereo
  - $s_y$ target motion in y

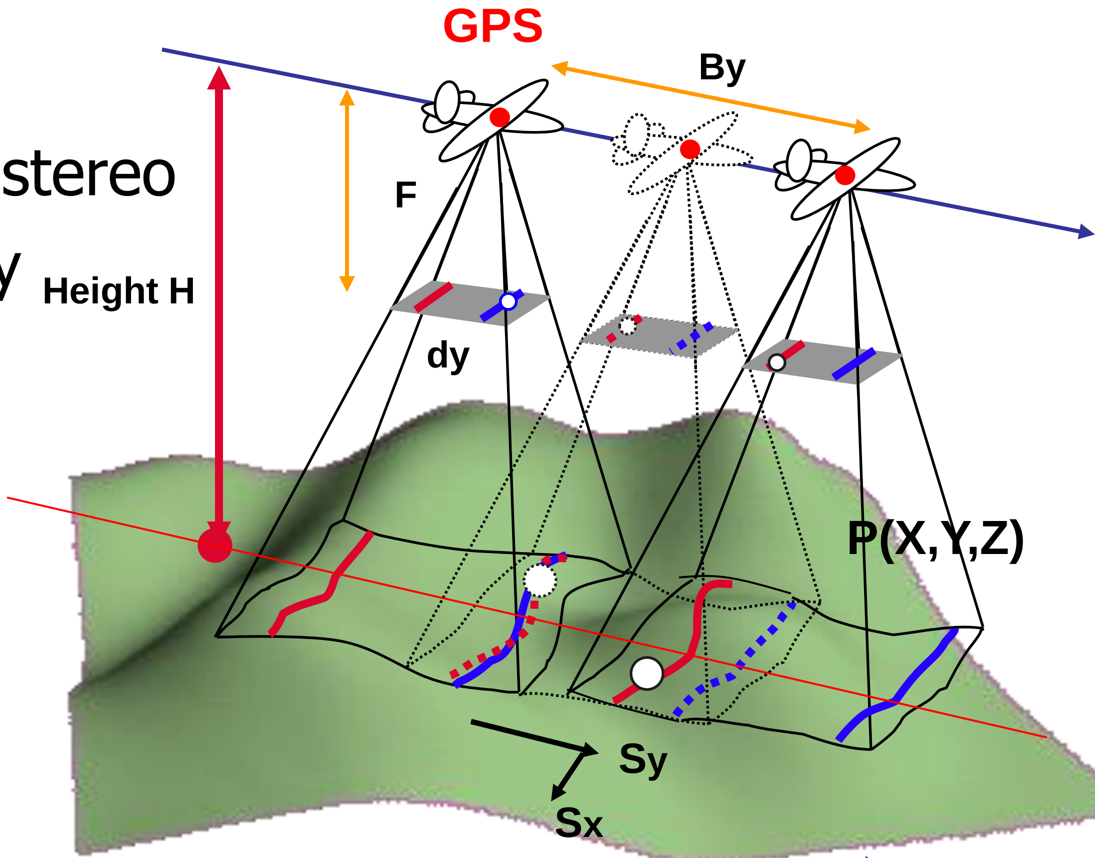
baseline

displacement

$$Z = F\frac{B_y - S_y}{d_y} = H\left(\frac{d_y + \Delta y - s_y}{d_y}\right)$$

disparity

Fixed !



Two views from different perspective ⇒ stereo

$s_y = F\,S_y / H$, $\mathbf{s_x = F\,S_x / H}$

# Moving targets are “out-of-place”

- General Cases – Epipolar Constraints
  - Violation of epipolar geometry
    - if $s_x$ <> 0
- Singularity Cases – 3D Constraints
  - Hanging above the road ($Z_{object} < Z_{road}$)
    - if sy < 0 (opposite direction)
  - Hiding below the road ($Z_{object} > Z_{road}$)
    - if sy > 0 (same direction)
- Special cases – Matching Constraints
  - Never seen – faster than sensor, same speed
  - No matches – same speed, too fast
  - …

dynamic pushbroom

$$Z = H\left(\frac{d_y + \Delta y - s_y}{d_y}\right)$$

object motion

$s_y = F\ S_y / H$

$\mathbf{s_x = F\ S_x / H}$

# 3D modeling from images – a mosaic based method using pushbroom geometry

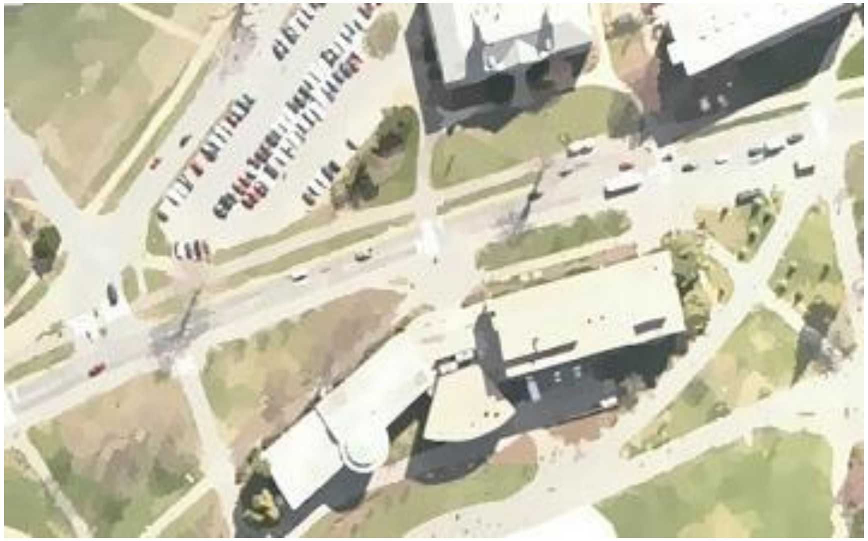
- Dynamic Pushbroom Stereo Mosaic Geometry
- 3D and Motion Content Extraction
- CB3M: Content-Based 3D Mosaics
- Experimental Results and analysis

# Summary of the core algorithms

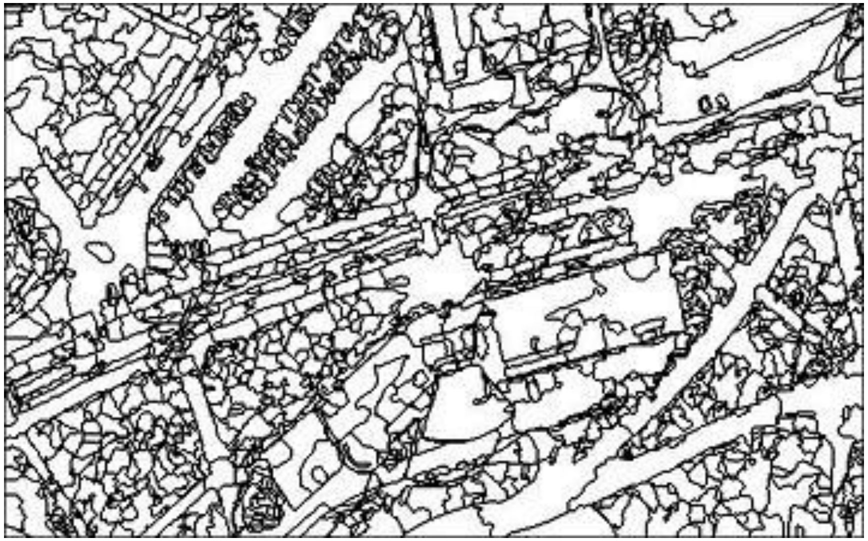
- **Mean-Shift method** for matching feature extraction
  - *Natural matching primitives* extracted for handling sharp depth boundaries and textureless regions
- **Patch-based stereo** matching – along epipolar lines
  - *Local matching on feature points + plane fitting* for 3D reconstruction
- **3D region refinement** from multiple views
  - *Multi-view integration* to obtain the best results for each patch
  - $N$ pairs of stereo mosaics, $(a_k, b_k, c_k, d_k)$, $k=1,\ldots,N$, **=>**$aX + bY + cZ = d$
- **3D region update** using local and global scene constraints
  - *Local Support**:** merge neighborhood regions*
  - *Global support:* a few dominant planes (urban scene)
  - *Classify reliable match*: classified into reliable matches and unreliable matches
- **Moving object extraction**
  - *Outlier detection*: Search the matches for outliers (unreliable matches) in 2D
  - *3D anomaly detection*: Using 3D information of surrounding static regions

# Extracting Nature Matching Primitives

- Mean-Shift method for color image segmentation
  - *Natural matching primitives* extracted for handling sharp depth boundaries and textureless regions



Color label image

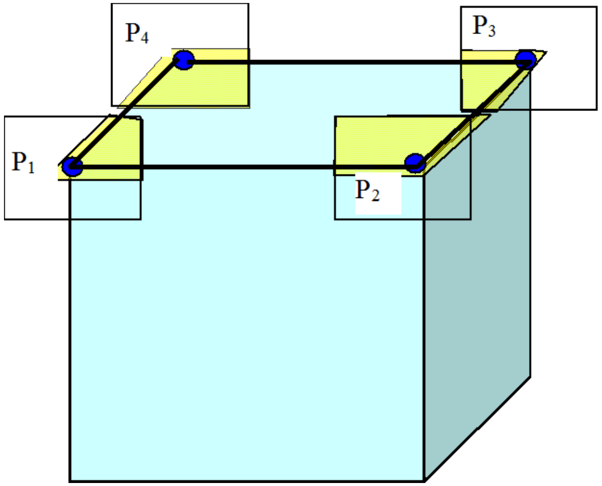


Boundary image

# Natural Matching Primitives (NMP)

- Square windows with varying sizes
- Detect features on patch’s boundary with large curvature
- Window based match with NMP **mask** :

$$W_l(u,v) = \begin{cases} 1, \text{if } (\mathrm{x}+\mathrm{u}, \mathrm{y}+\mathrm{v}) \in P_i \\ 0, \text{otherwise} \end{cases}$$

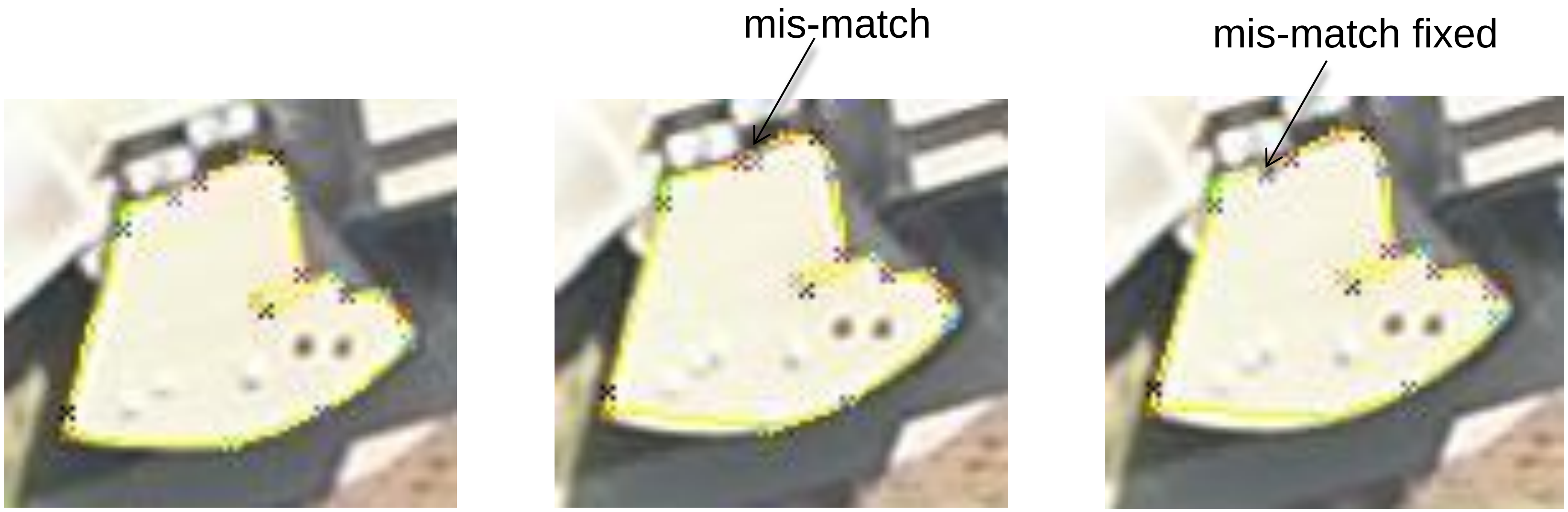


# Feature correspondence

- Cross correlation of **masked window of only interest points on boundaries**(to favor sharp depth boundaries)
- Search correspondence along **epipolar curve** of pushbroom stereo
- Search using **multiple scale** (adaptive window size and search steps) for speedup
- Filter false alarm using **Cross-Check**

# NMP: Surface Fitting and Refinement

- Use RANSAC to fit the plane.
  - Each time three interest points are used to fit a plane.
  - The best set of plane parameters is extracted with best fitting most of the interest points in the patch.

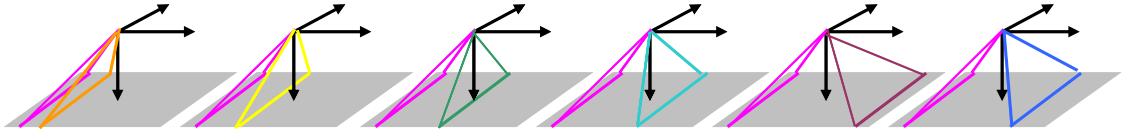


mis-match

mis-match fixed

# 3D Region Refinement from Multiple Views (multi-baseline method)

- Match reference mosaic with other mosaics.



$$\partial Z = \frac{H}{d_y} \partial \Delta y$$

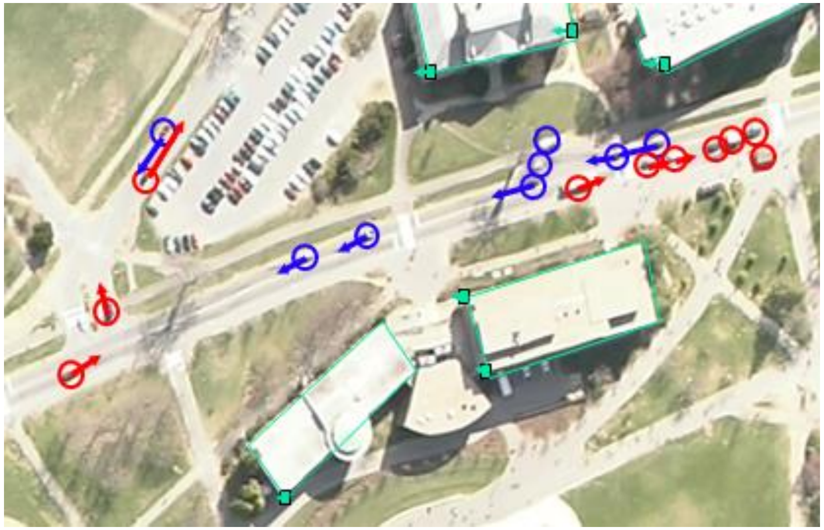
- Multiple sets of plane parameters are extracted from multi-pair stereo mosaics.
  - To facilitate both correspondence and reconstruction
- The best set of plane parameters is obtained.
  - Warping the patch from the reference mosaic to the target mosaic using each set of plane parameters.
  - Find the set of parameters with the best match score.

# Plane updating using local and global scene constraints

- Local scene constraints
  - Refine the estimate using **neighborhood support**
- Global scene constraints
  - For urban scenes with a number of **dominant surface directions**
  - Compute dominant directions by clustering plane norms
  - Apply the dominant plane parameters on each patch to refine plane estimates

# Moving Object Extraction

- Classify reliable category and unreliable category
- Outlier detection (*unreliable matches*)
  - Search the matches for outliers in 2D region in the multiple mosaics.
- 3D anomaly detection (*reliable match*)
  - Using 3D information of surrounding static regions to detect the 3D anomalies, e.g. height of object is 50 meters and all surrounding objects are just 1-2 meter high.



# 3D modeling from video – a mosaic based method using pushbroom geometry

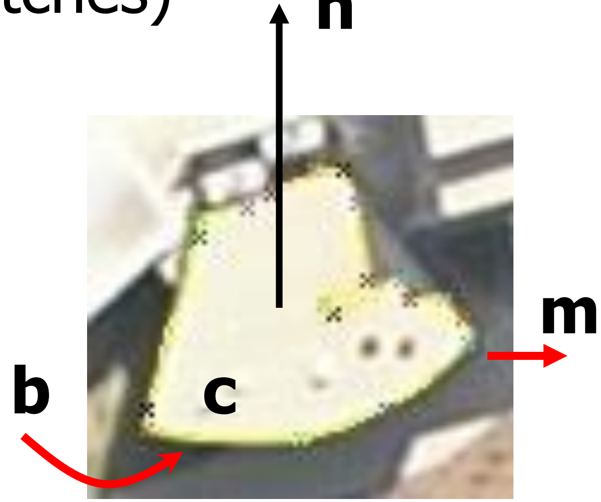
- Dynamic Pushbroom Stereo Mosaic Geometry
- 3D and Motion Content Extraction
- ➡ Content-Based 3D Mosaics (CB3M)
- Experimental Results and analysis

# CB3M: **C**ontent-**B**ased **3D** **M**osaic

- a set of video object (VO) primitives (patches)

$$\mathbf{CB3M} = \{VO_i\} = \{ (\mathbf{c}_i, \mathbf{b}_i, \mathbf{n}_i, \mathbf{m}_i)\}$$

i =1, …, N (number of patches)



- $\mathbf{c}_i$: color
- $\mathbf{b}_i$ : boundary points in chain codes
- $\mathbf{n}_i = (n_x, n_y, n_z, d)$  planar surface parameters
- $\mathbf{m}_i = (S_1, S_2, \ldots, S_L)$ motion para, e.g. (Sx, Sy) if L = 2

# CB3M: Content-Based 3D Mosaic

- a set of video object (VO) primitives (patches)

$$\mathbf{CB3M} = \{VO_i\} = \{\, (\mathbf{c}_i, \mathbf{b}_i, \mathbf{n}_i, \mathbf{m}_i)\} \,,\ i = 1, \ldots, N$$

- Data amount for CB3M

$$N_{color} + N_{boundary} + N_{structure} + N_{motion}$$
$$= 3N + (8N+3K/8) + 4{*}4N + 4L{*}N_m$$
$$= 27N+3K/8+4LN_m \text{ (bytes)}$$

- N: number of patches
- K: total number of boundary points
- Nm: total number of moving regions
- L: number of motion parameters for each patch

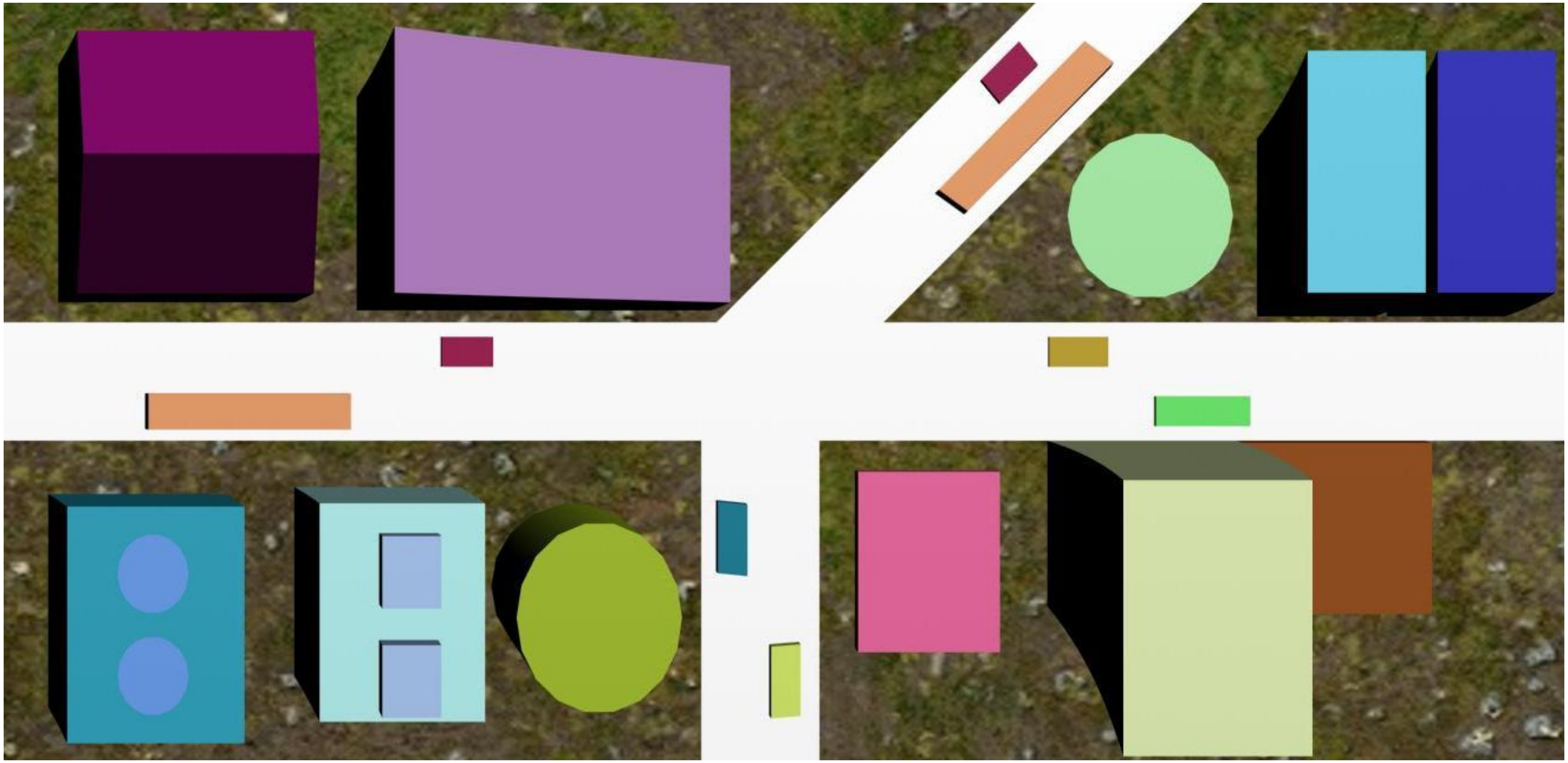
# 3D modeling from video – a mosaic based method using pushbroom geometry

- Dynamic Pushbroom Stereo Mosaic Geometry
- 3D and Motion Content Extraction
- CB3M: Content-Based 3D Mosaics
- Experimental Results and analysis

# Experimental results

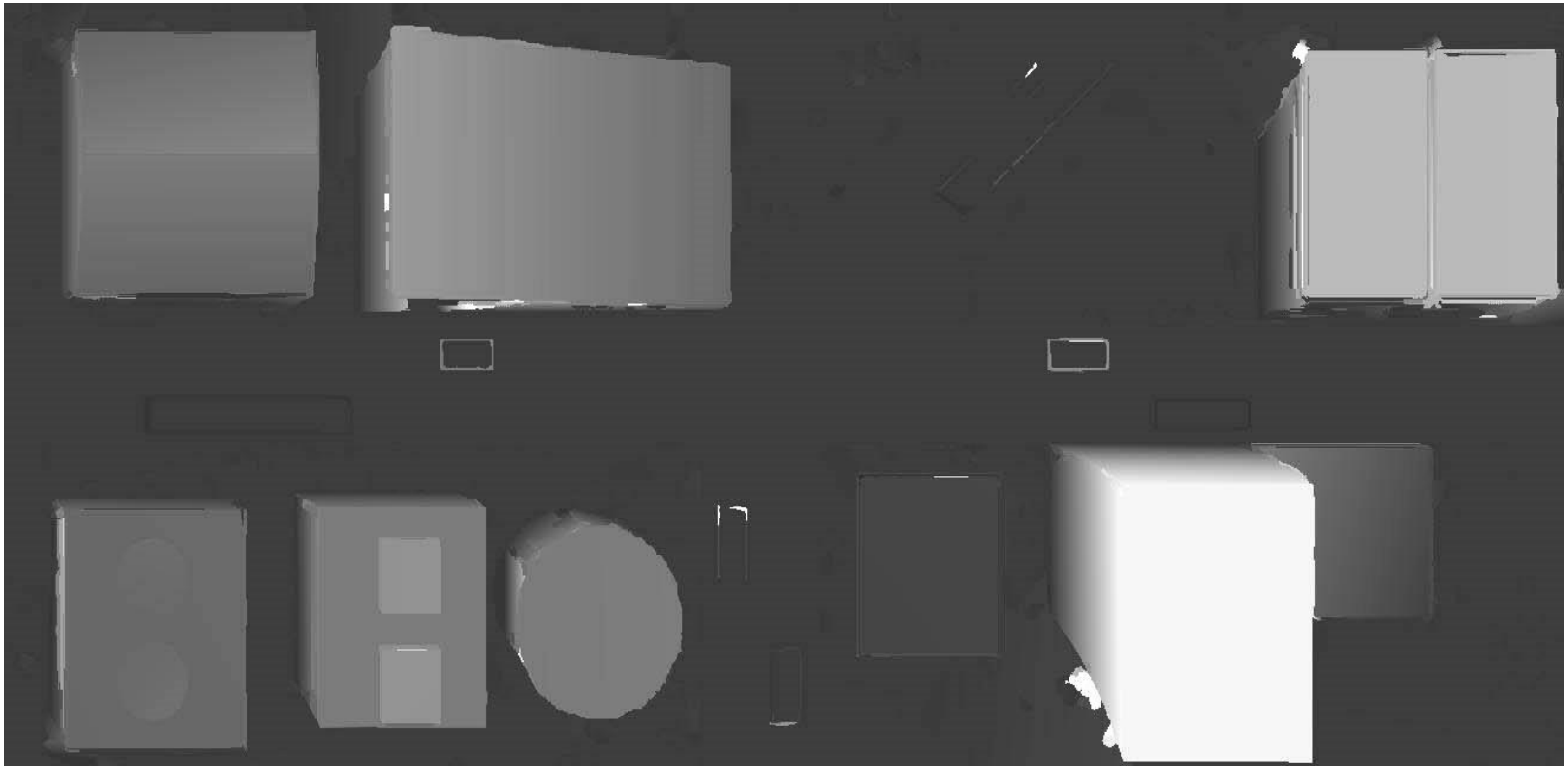
- Simulation data set
- Real data set 1 – UMASS campus
- Real data set 2 – NY City data

# Simulation data



Reference Mosaic **: The Leftmost View**
resolution: 640x1320

# Simulation data

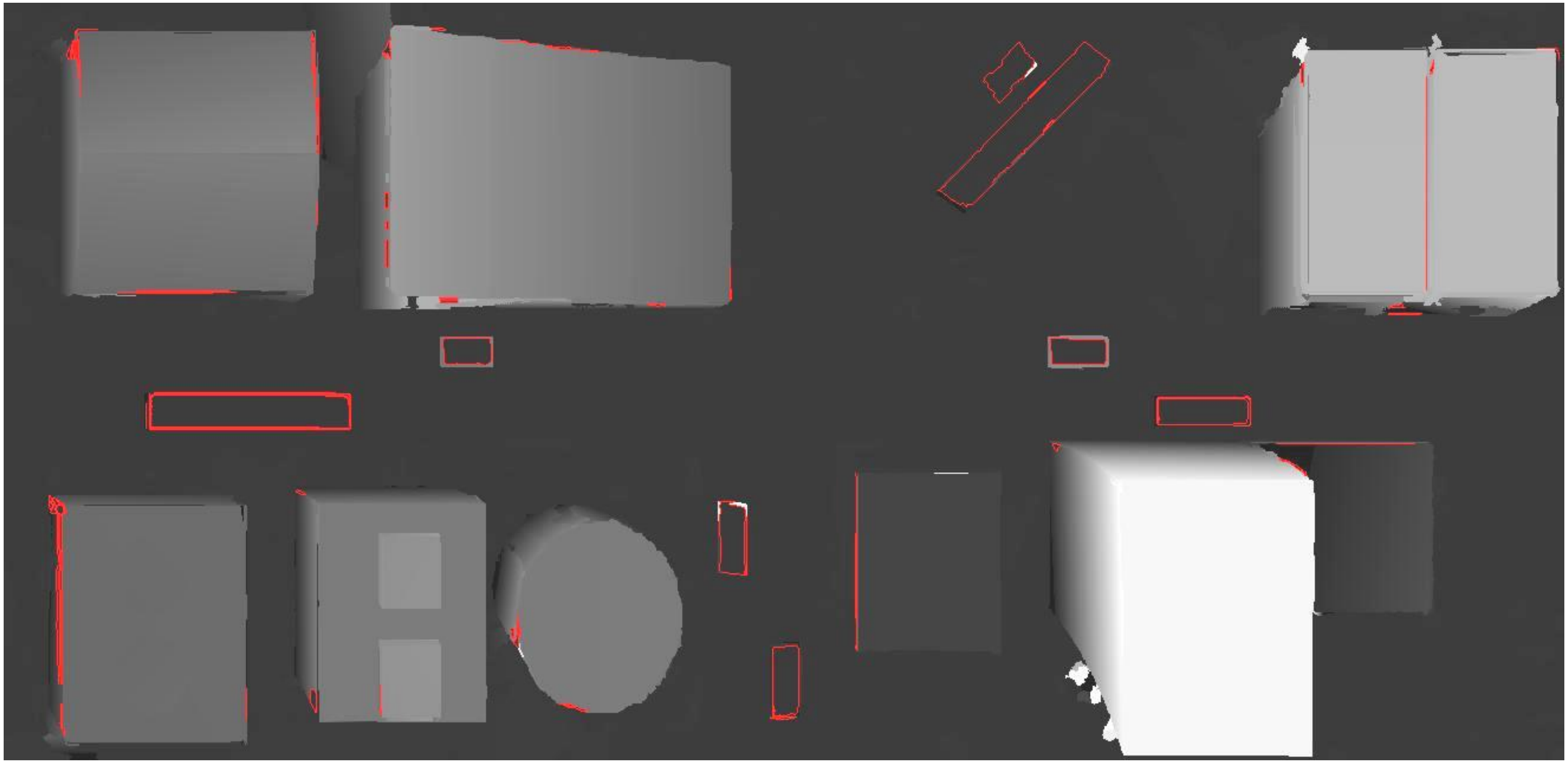


"Height" Map from the First Pair of Mosaics

**Average error ($E_{1st}$) of 85% points is 0.543 meters**

| # pxls | 75% | 85% | 100% |
|---|---|---|---|
| $E_{1st}$ | 0.20m | 0.54m | 5.42 |
| $E_{all}$ | 0.07m | 0.20m | 3.65 |

# Simulation data



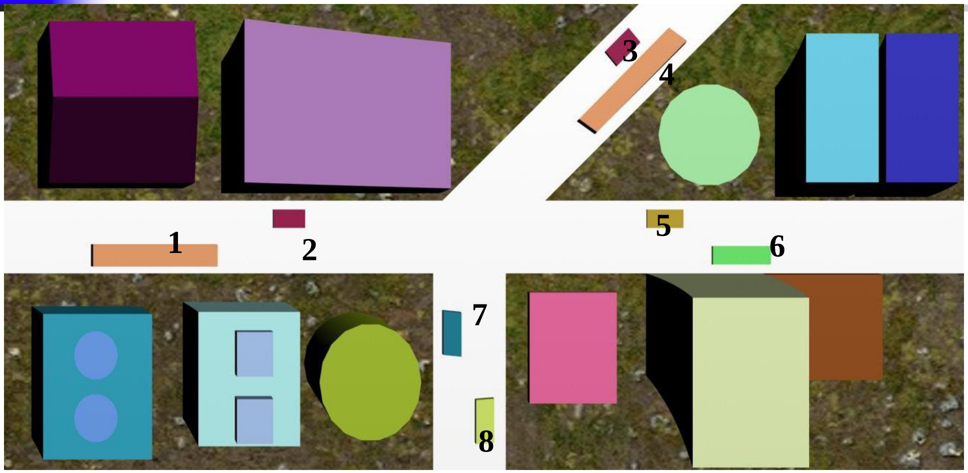
| # pxls | 75% | 85% | 100% |
|---|---|---|---|
| $E_{1st}$ | 0.20m | 0.54m | 5.42 |
| $E_{all}$ | 0.07m | 0.20m | 3.65 |

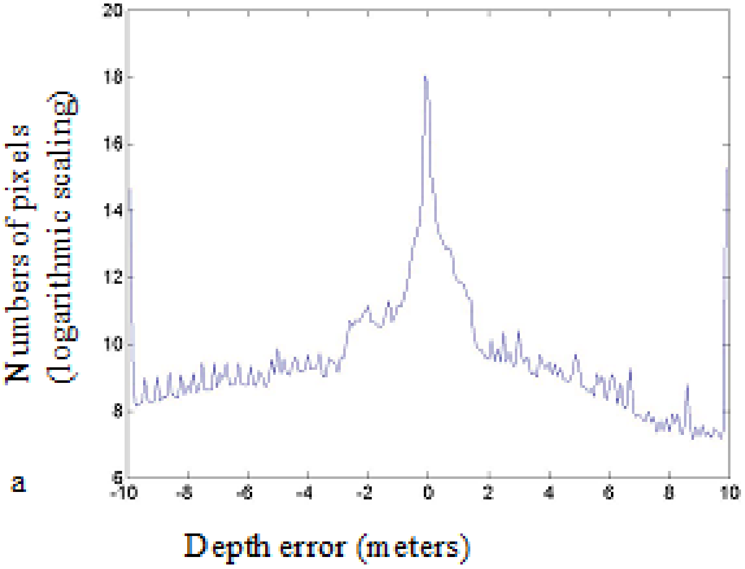
“Height” Map from the All the Nine Mosaics

**Average error ($E_{all}$) of 85% points is 0.195 meters**

Red shows boundaries of 2D search for moving targets

# Simulation data





**Depth error (meters)**

**Video simulation parameters:**

**Camera height H = 300 m, Focal length F = 3000 pixels**
**Image size WxH = 640x480, Number of frames N = 1640**
**Number of mosaics 9; Two farthest mosaicing windows' distance dy = 320 pixels**

$$Z = H\frac{d_y + \Delta y}{d_y}$$

$$\partial Z = \frac{H}{d_y}\partial \Delta y$$

**Scene: road, grass, buildings w/ various roofs, moving targets with ground truth data (Above figure shows one of the 9 mosaics with moving targets # 1- 8 )**

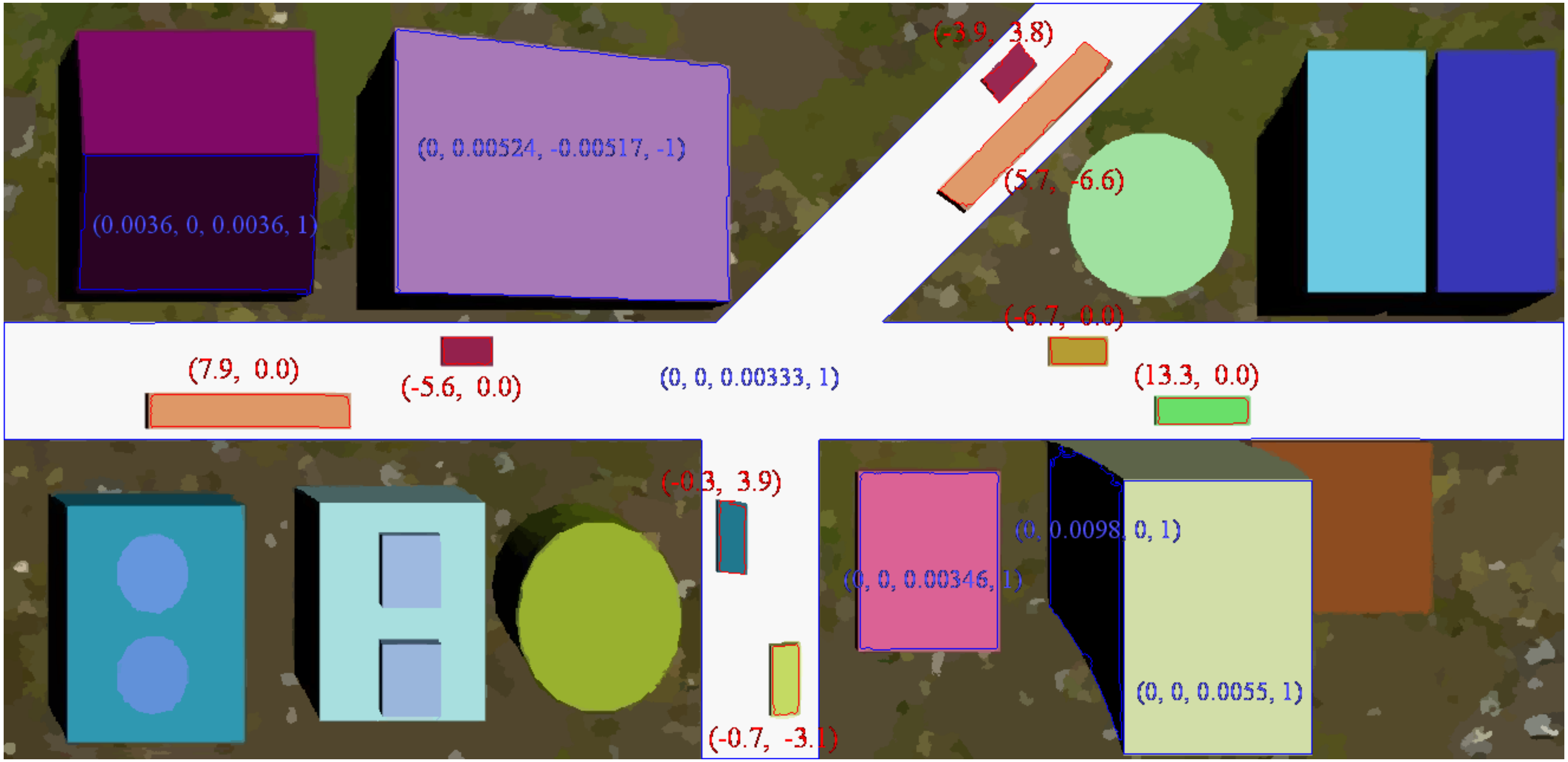
**Results: Using the multi-view approach, the average depth error of 85% points is 0.195 meters**

**Note: 1 pixel error in image matching corresponds to 0.94 meter error (best case) in depth, and we used a sub-pixel match**

## Motion estimation accuracy

| # | Ground Truth (cm/frame) | | Estimated Results (cm/frame) | | Errors (cm/frame) | |
|---|---|---|---|---|---|---|
| | Sx | Sy | Sx* | Sy* | dSx | dSy |
| 1 | 0 | 2.485 | 0 | 1.649 | 0 | 0.836 |
| 2 | 0 | -1.499 | 0 | -1.628 | 0 | 0.129 |
| 3 | 1.064 | -1.262 | 1.053 | -1.08 | 0.011 | -0.181 |
| 4 | -1.414 | 1.414 | -1.444 | 1.247 | 0.031 | 0.166 |
| 5 | 0 | -1.999 | 0 | -2.012 | 0 | 0.013 |
| 6 | 0 | 2.499 | 0 | 2.495 | 0 | 0.003 |
| 7 | 0.999 | 0 | 0.982 | -0.076 | 0.017 | 0.076 |
| 8 | | | | | | |

# Simulation data

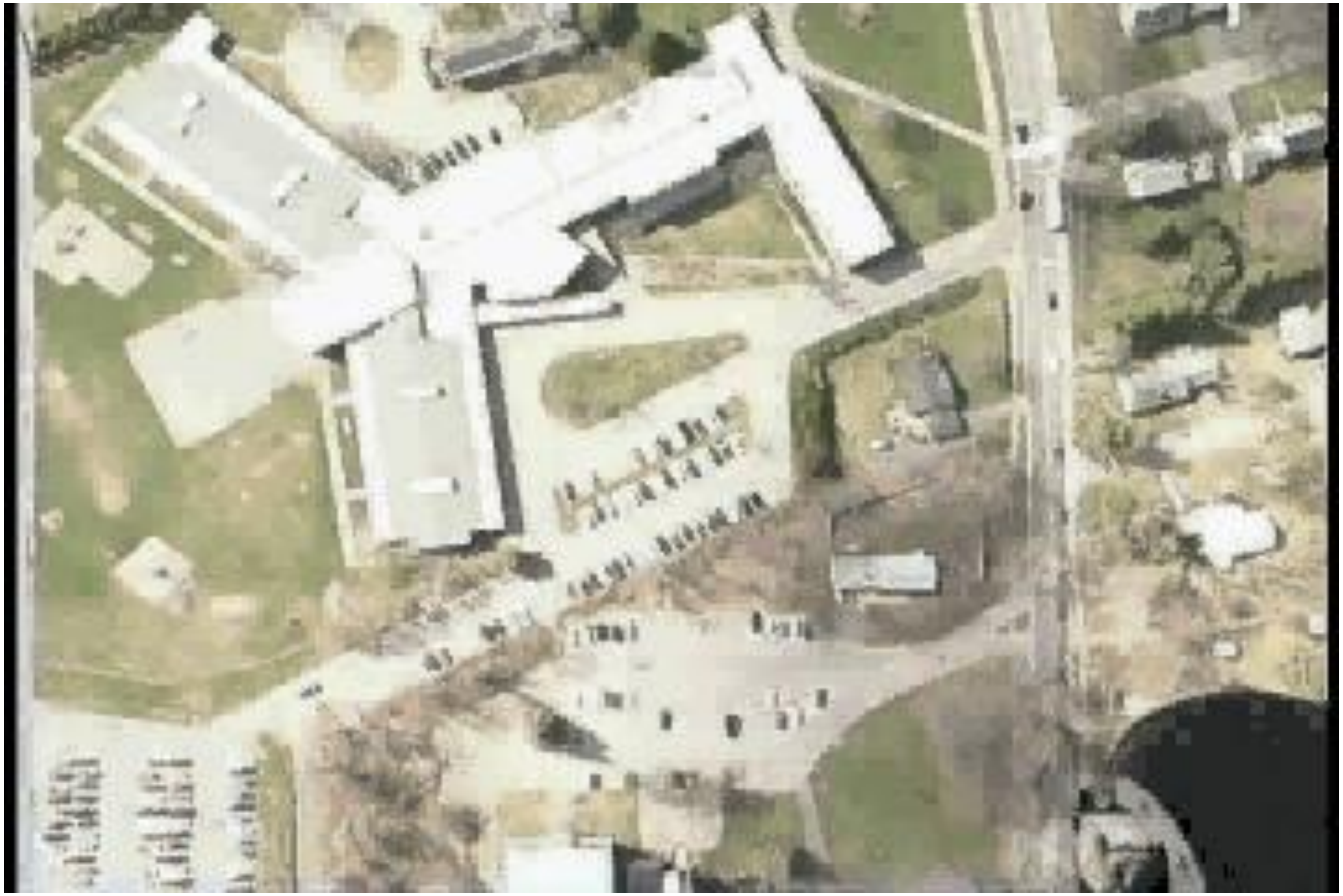


CB3M: The Content-Based 3D Mosaics { (ci, **bi**, **n**i, **mi**)}

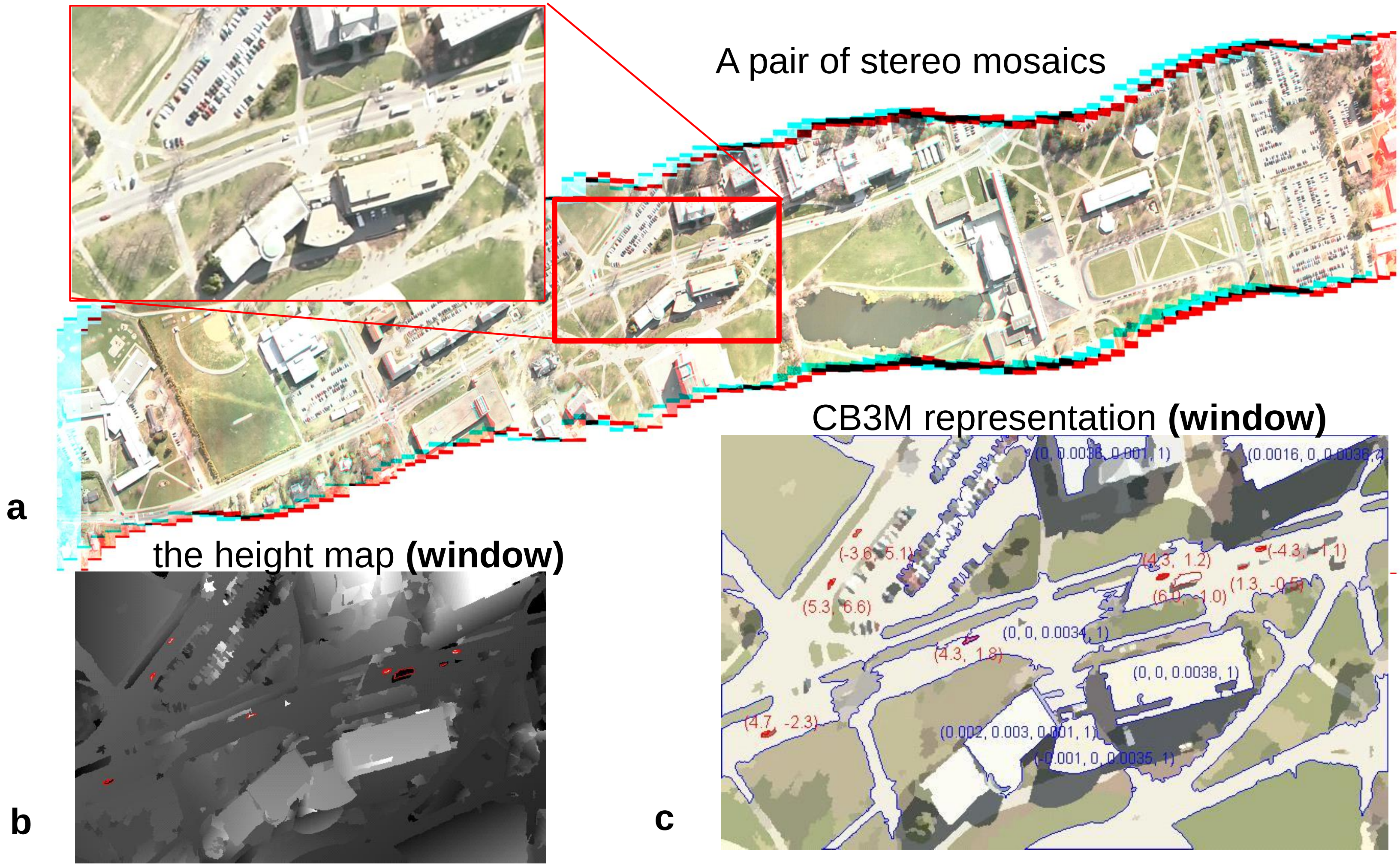
Content Representation: Plane parameters in blue and motion in red

Compression ratio is **76,061**

# UMASS campus



# UMASS campus



Plane parameters in blue and motion in red; **compression ratio is 10,001**

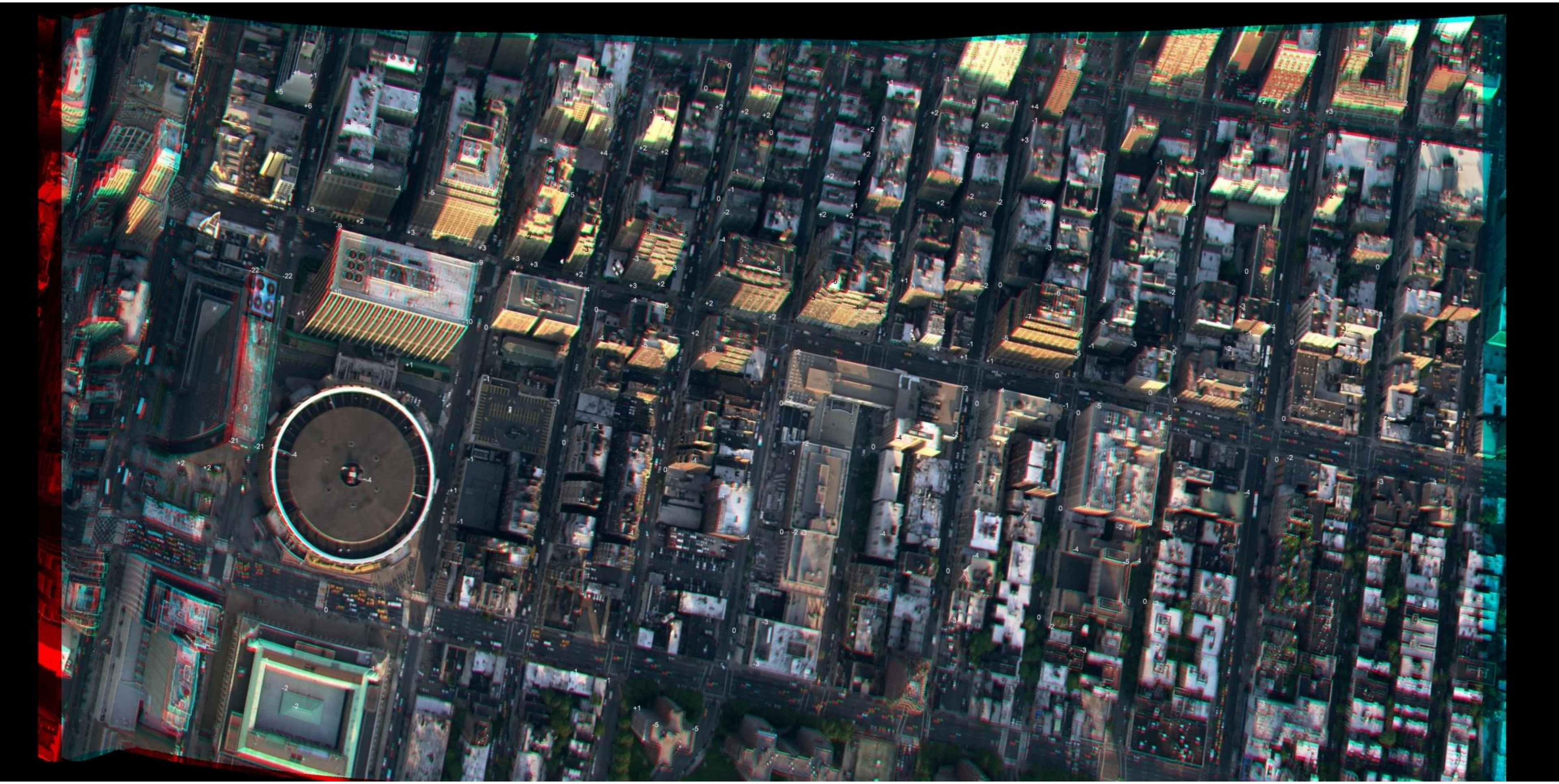
# UMASS campus

- Original Video Sequence: **879 MB**
  - 1000 * 640*480 *3  BMP
- A pair of stereo mosaics: **1.1 MB**
  - 4448*1616*2 = 41 MB in BMP -> 1.1 MB in JEPG
  - Compression ration > 800
- CB3M representation: **90 KB** (real file size)
  - Raw data: 316 KB ( real file size)
  - Lossless winzip file 90 KB
  - Compression ratio **10,001  ($10^4$)** !

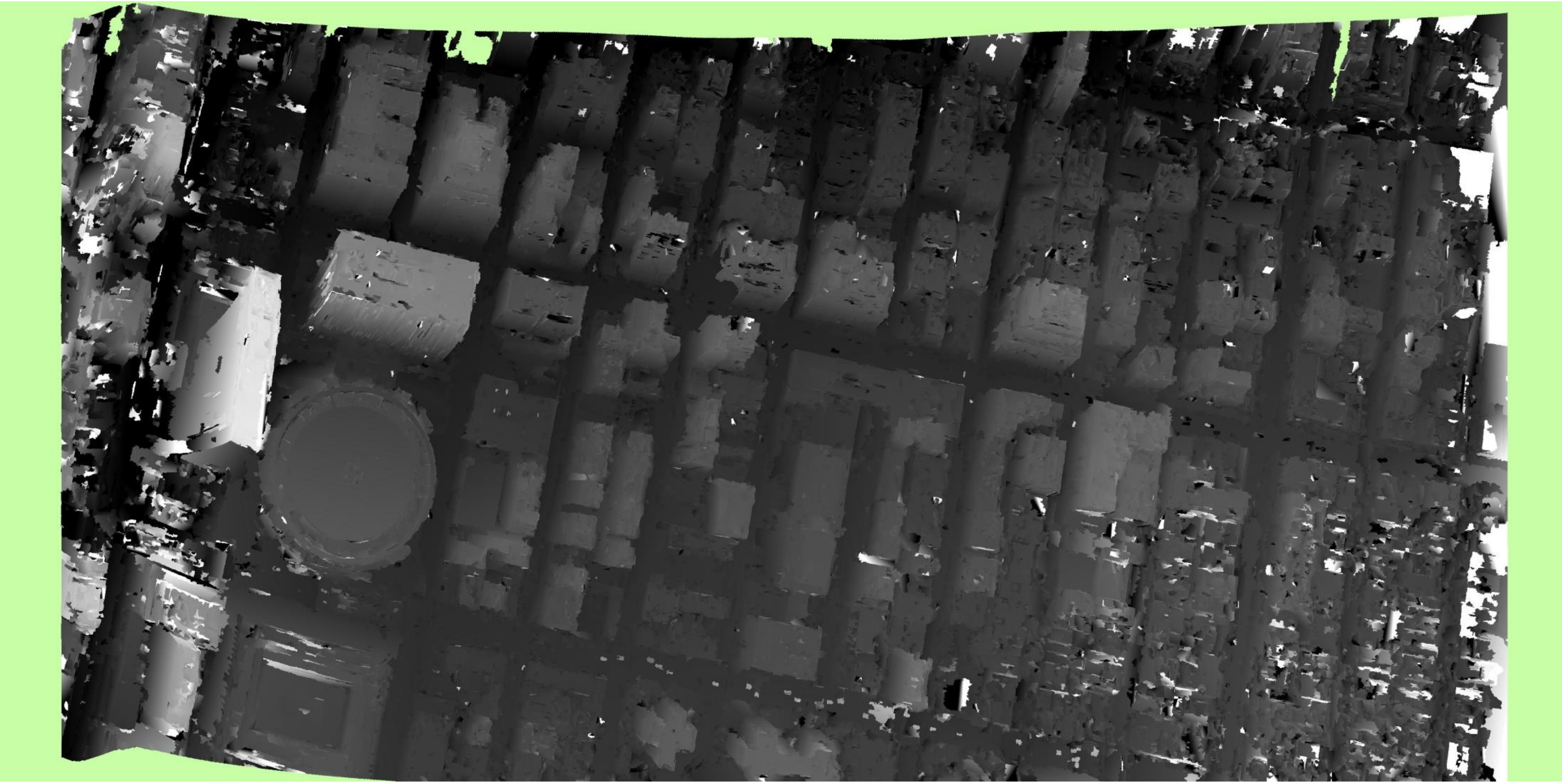
  - **Data with contents (3D and motion parameters)**

# NYC data

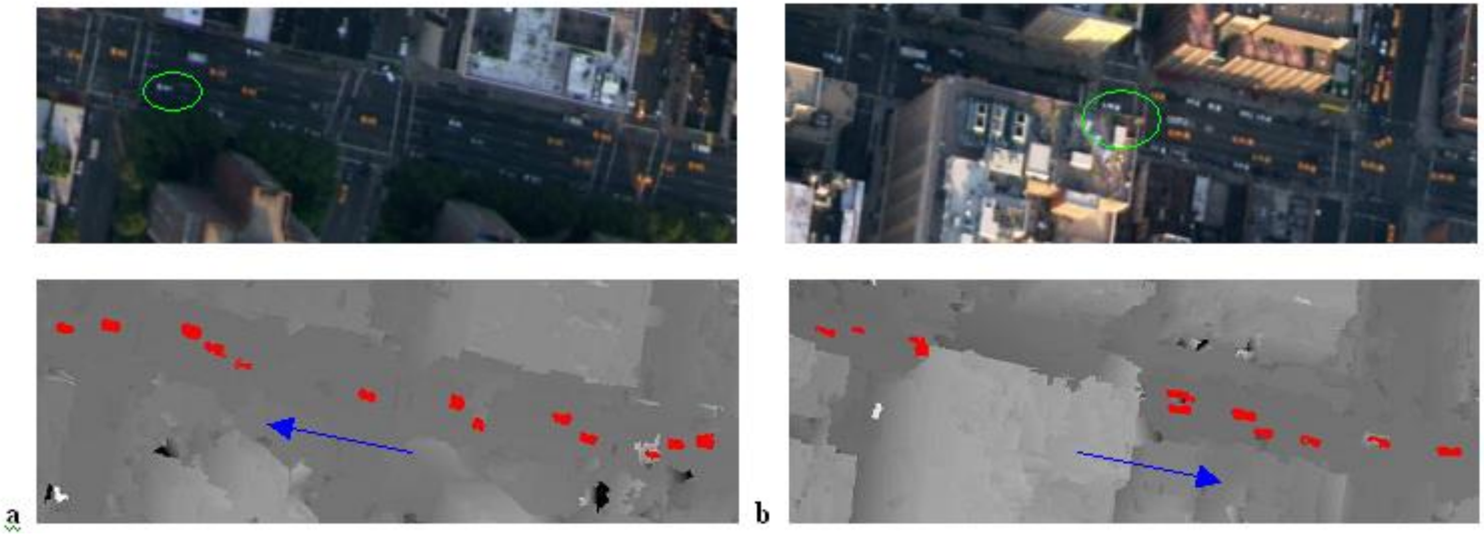
# More Results – NYC data



# Depth map



# Moving targets



# Outline of the Presentation

- **Introduction**
- **Related Work**
- **Research Topics and Methodology**
  - 3D modeling from video – **a mosaic based approach and the core algorithms**
  - **Scene understanding /labeling** from Content Based 3D Mosaics
  - 3D modeling from video – the core algorithms **extended to perspective stereo** images
- **Summary and Future Work**

# Scene labeling from Content Based 3D Mosaics

- A multiple-layer method
  - a surface layer
  - a structural layer
  - a cluster layer
- Using a graph based representation
  - $G = (V, E)$
  - V: patches
  - E: connection between neighboring patches

## Surface layer generation

- Measure the confidence of 3D estimation

$$Conf(i) = C_a(i)\, Conf_{shape}(i)\, Conf_{match}(i)\, Conf_{planesmooth}(i)$$

- Rank confidence
- Start from the patch with maximal confidence
- Merge neighbor regions if they are similar

$$similarity(\mathrm{i,j}) = \mathrm{S_n(i,j)S_d(i,j)S_c(i,j)}$$

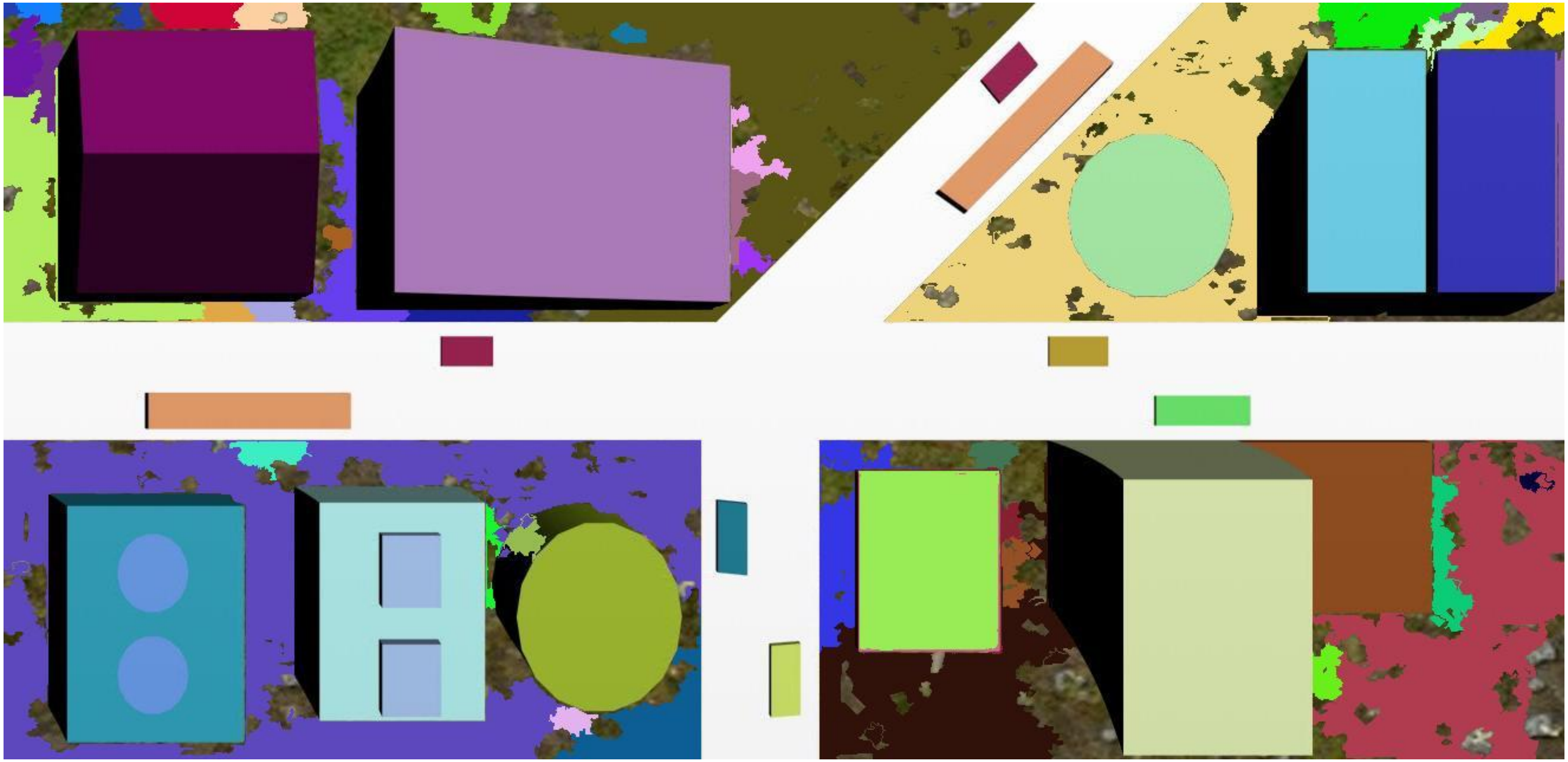
$$S_c(\mathrm{i,j}) = \mathrm{e}^{-\left(\frac{I_i - I_j}{\sigma_I}\right)^2} \quad S_n(\mathrm{i,j}) = \mathrm{e}^{-(\frac{n_i - n_j}{\sigma_n})^2} \quad S_d\,(\mathrm{i,j}) = 1 - \mathrm{e}^{-(\frac{n_i \bar{x}_j + n_j \bar{x}_i}{\sigma_d})^2}$$

# Surface layer generation

Sort patches by confidences in decreasing order => P; (P(i): ith patch; P(k, j): the jth neighbor patch of the kth patch)

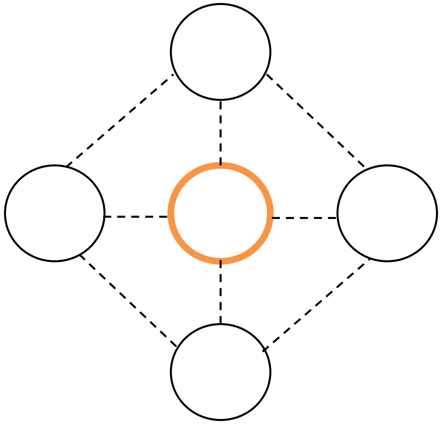
```
Q = NULL; Labels = NULL; i=1;
For i=1…N  //N: the number of patches in image
  If Labels(P(i)) != Null or Conf(P(i)) < $T_{conf}$   //$T_{conf}$: a threshold of confidence
     Continue;
  End-if
  Enqueue(Q, P(i));
  While Q is not empty
       P(k) = Dequeue(Q);
       If Conf(P(k))<=$T_{conf}$
          Continue;
      End-if
      for j=1…M                         //M: the number of the neighbors of P(k)
          if Conf(P(k,j)) >= Tc and Similarity(P(k), P(k,j)) < Ts
                  Enqueue(Q, P(k,j));
                  Labels(P(k,j)) = l;
               End-if
      End-for
  End-while
i++;
End-if
```

# Surface layer generation - result



# Structural layer generation

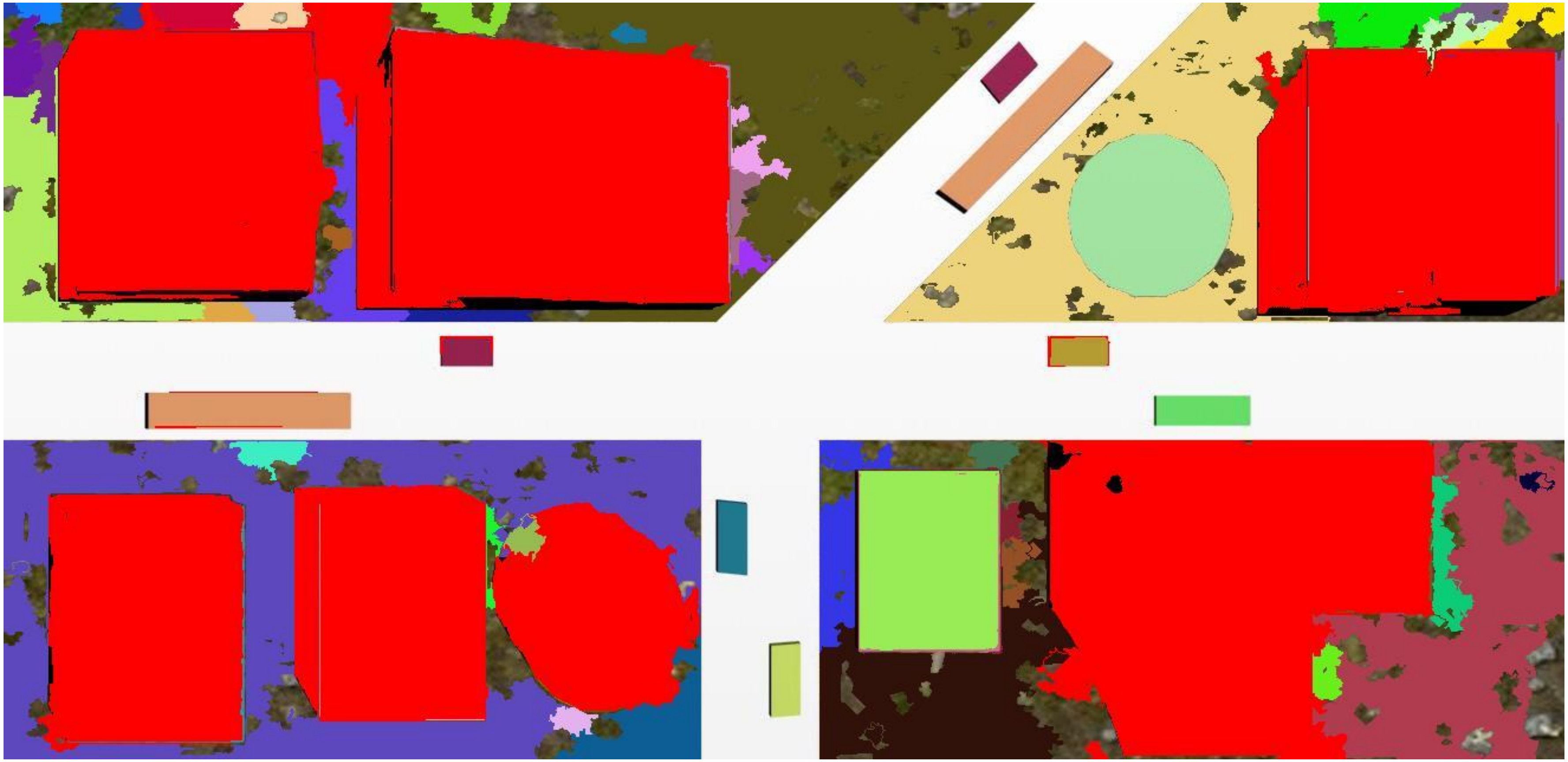
- Group all related patches together



Object graph

- Sub-graph search problem (NP completed)
- Use an exhaustive search

# Structural layer generation - result



# Cluster layer generation

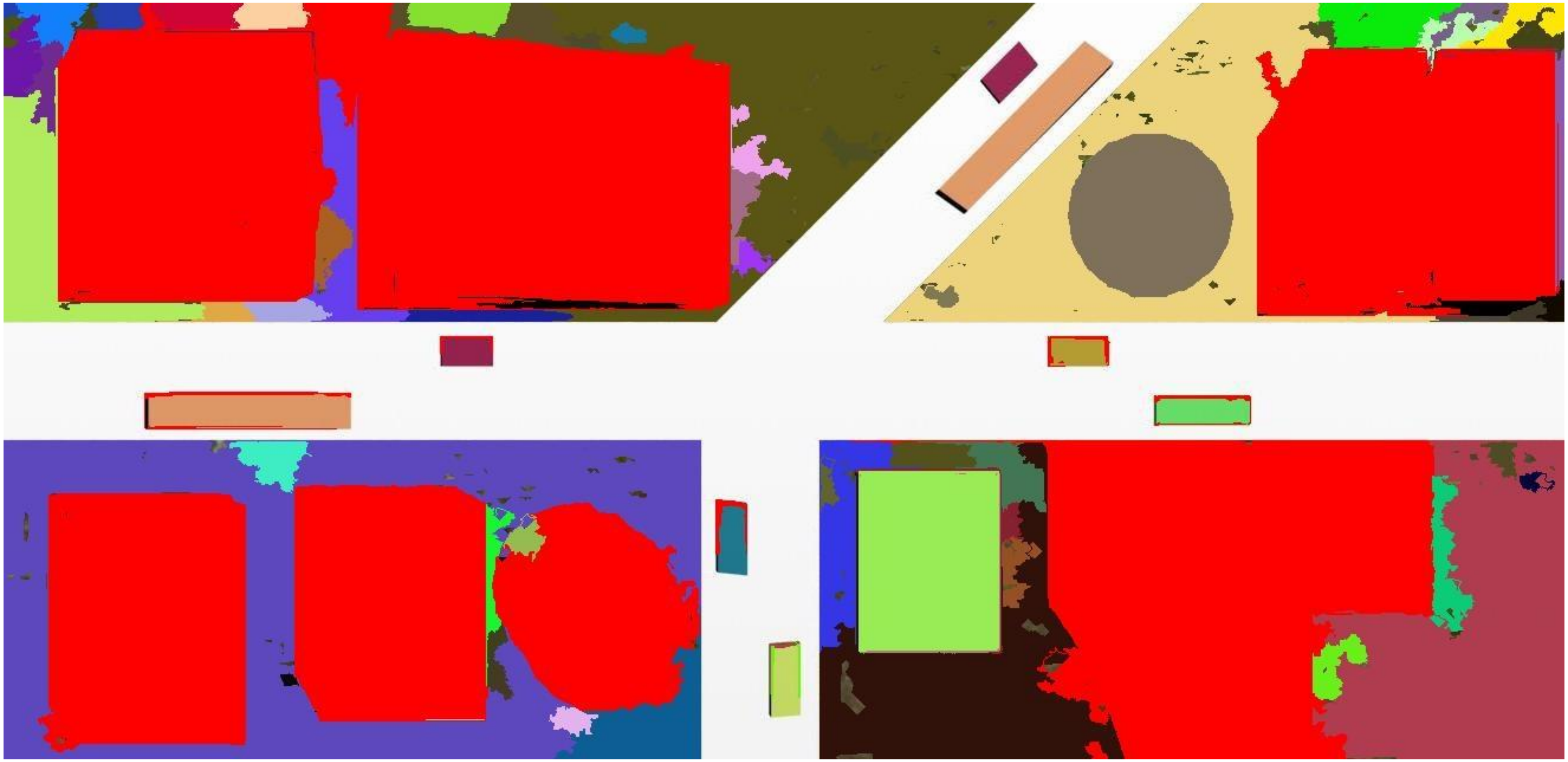
- For small patches, merge them into their neighbor patches

$$similarity_1(\mathrm{i,j}) = \mathrm{S_c(i,j)}\ Conn(i,j)$$

$$\mathrm{Conn(i,j)} = \frac{\mathrm{lb_{i,j}}}{\mathrm{lb_i}}$$

$$S_c(\mathrm{i,j}) = \mathrm{e}^{-\left(\frac{I_i - I_j}{\sigma_I}\right)^2}$$

# Cluster layer generation - result



# Outline of the Talk

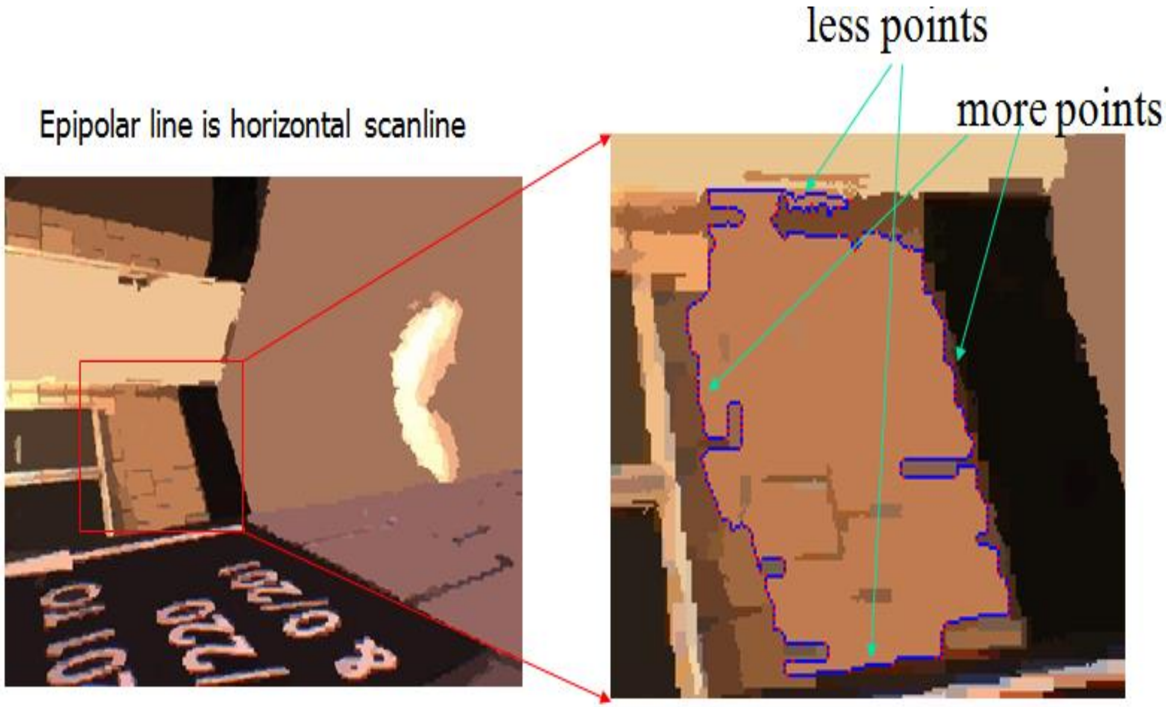
- **Introduction**
- **Related Work**
- **Research Topics and Methodology**
  - 3D modeling from video – **a mosaic based approach and the core algorithms**
  - **Scene understanding /labeling** from Content Based 3D Mosaics
  - 3D modeling from video – the core algorithms **extended to perspective stereo** images
- **Summary and Future Work**

# 3D modeling from video – using perspective geometry

- Using mosaic based method
  - First prepare images with large FOV
  - Then feature extraction become easier and match more efficient
- With perspective view
  - First compute 3D models (done)
  - Then merge multiple models into a large 3D model (future work)
  - Challenges in indoor scene => large textureless regions
- Extension of the core algorithms
  - Patch-based stereo matching

# Extension of Patch-based stereo matching

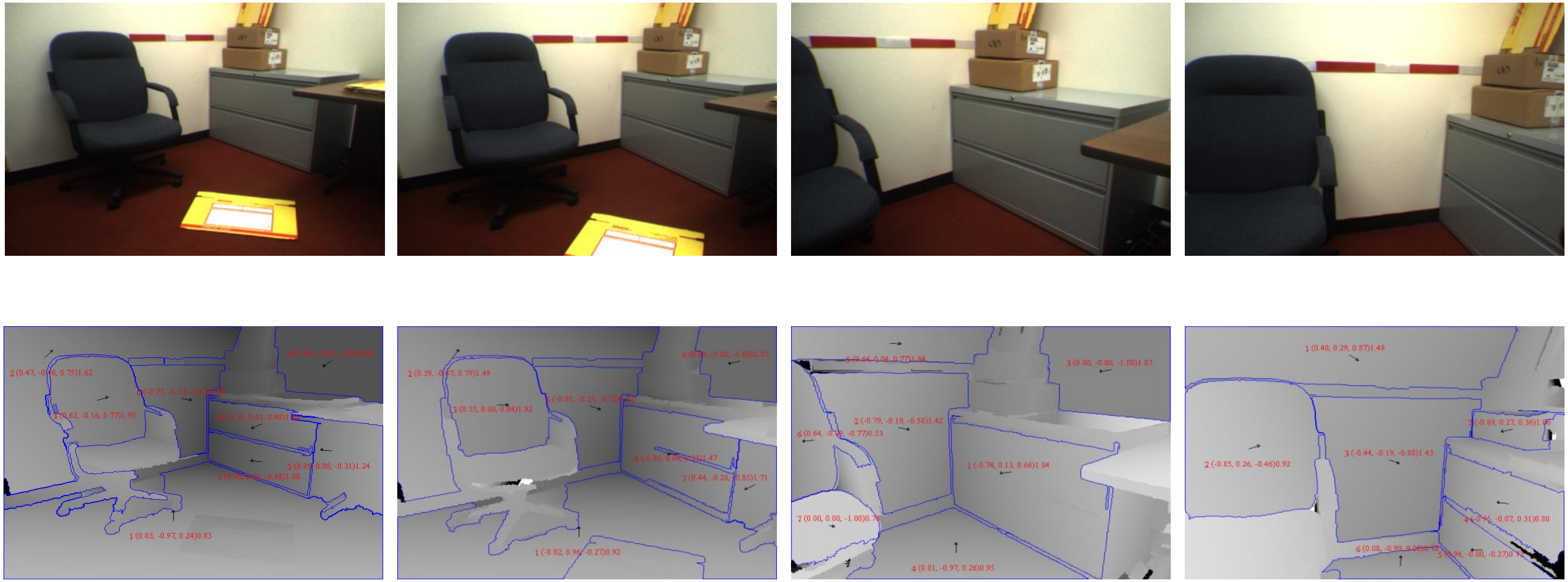
- Different geometry
- Increase number of interest points
- Give less weight to features closed to image border



# Experimental result

- Bumblebee stereovision head
- Baseline = 12 cm
- Focal length = 3.8 mm

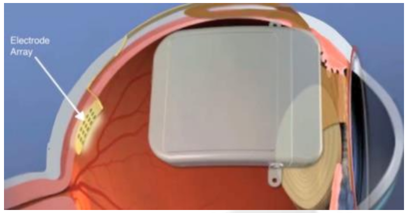
# Experimental result

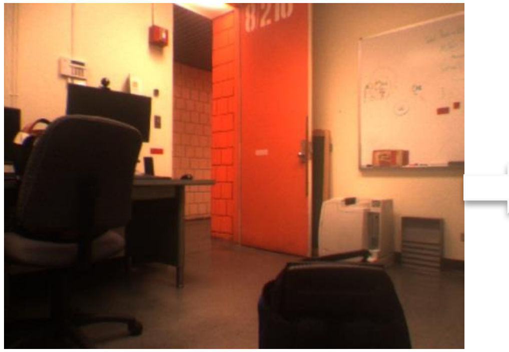


- Speed: 3-5 Sec.
- Can be optimized to near real-time

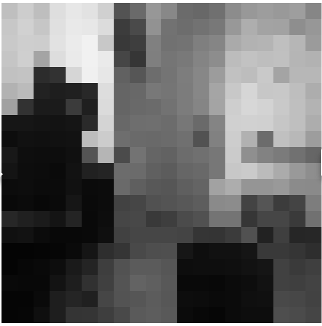
# Apply the 3D modeling onto assistive technology

- Visual prosthetic devices
  - Retinal implant, Brainport
  - Transfer visual information from a digital video camera to your retinal or tongue
  - Low resolution: less than <= 400 pixels
  - Functionalities: recognize high-contrast objects, their location, movement

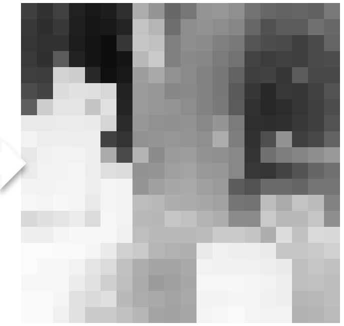




Original image



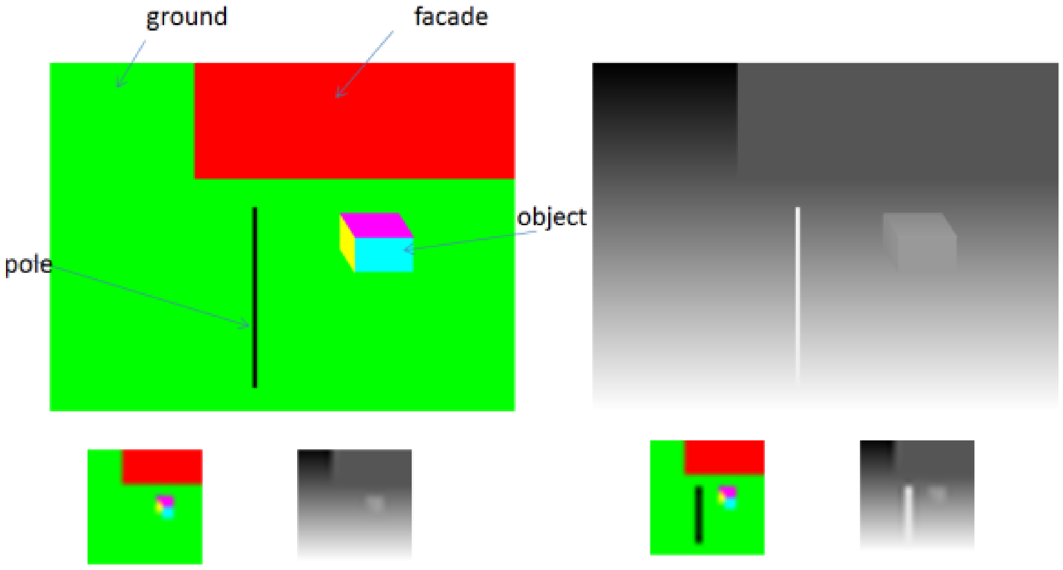
Sampled image



Inversed Sampled image

# Limitation of prosthetic device

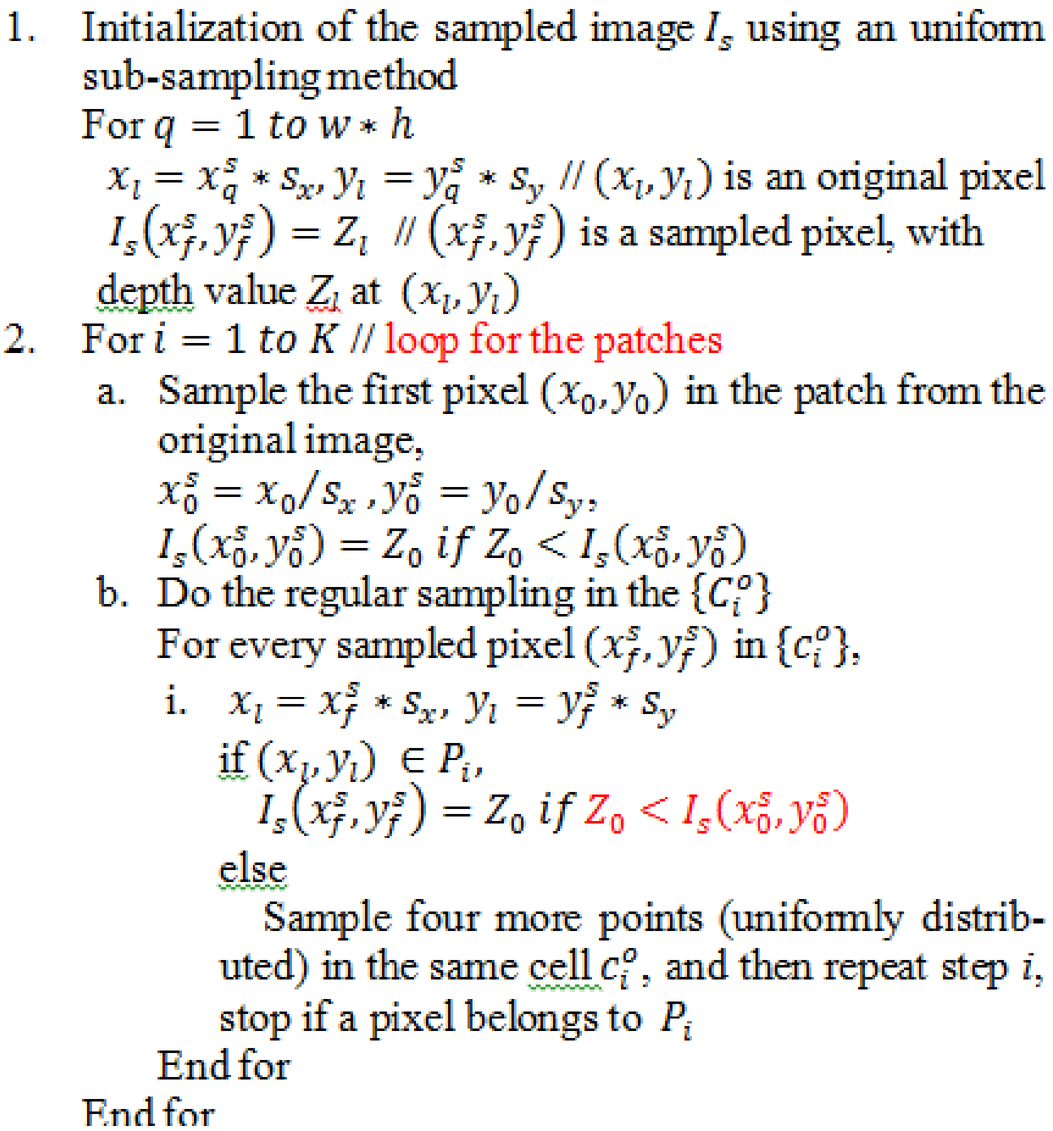
- Use intensity information only
- Low resolution
- No distance information
- Ignore small objects while down-sampling



Sampling

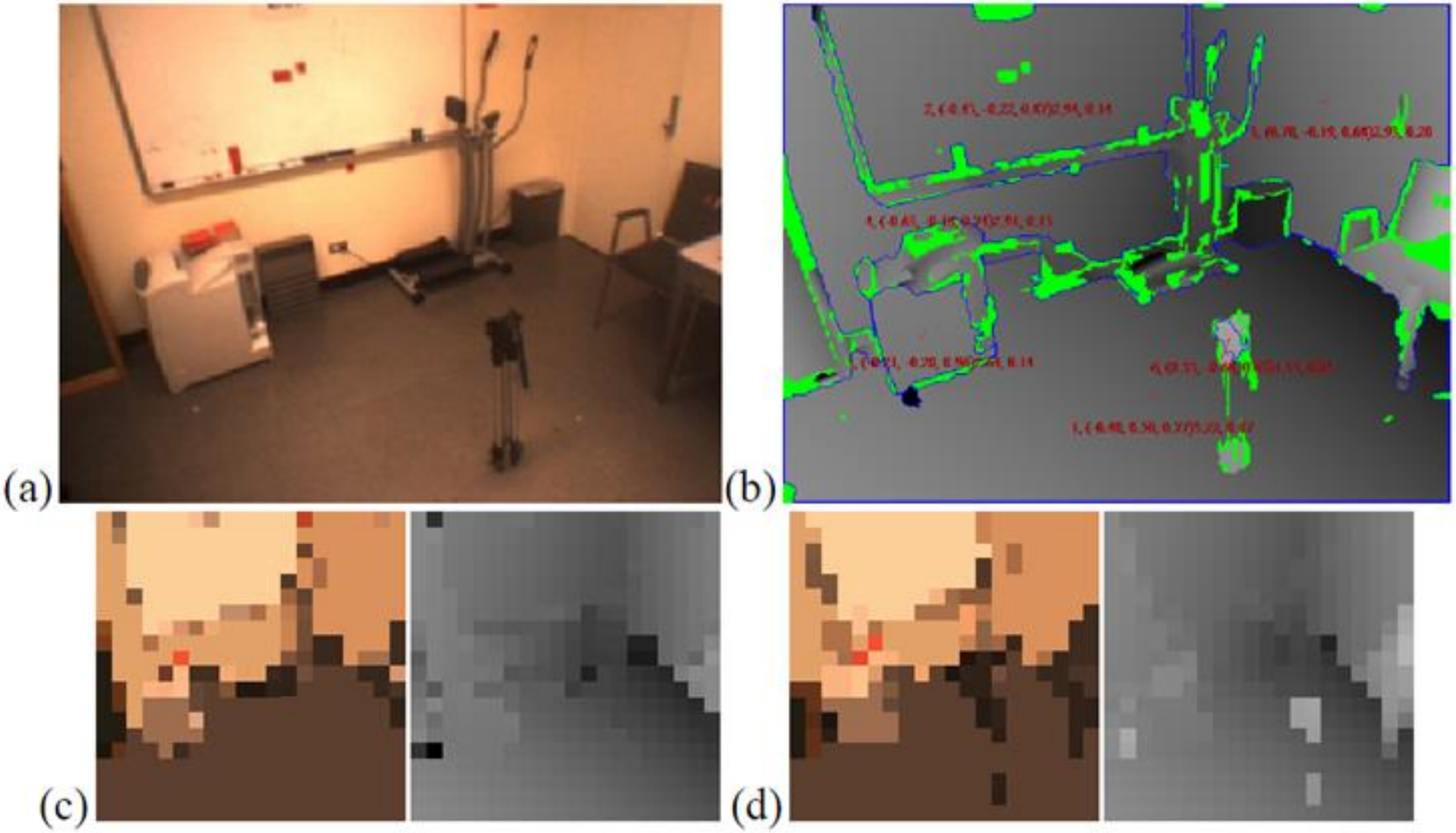
Smart sampling

# Smart SubSampling - algorithm



```
1.  Initialization of the sampled image $I_s$ using an uniform sub-sampling method
    For $q = 1$ to $w * h$
      $x_l = x_q^s * s_x, y_l = y_q^s * s_y$ // $(x_l, y_l)$ is an original pixel
      $I_s(x_f^s, y_f^s) = Z_l$ // $(x_f^s, y_f^s)$ is a sampled pixel, with
      depth value $Z_l$ at $(x_l, y_l)$
2.  For $i = 1$ to $K$ // loop for the patches
      a.  Sample the first pixel $(x_0, y_0)$ in the patch from the original image,
          $x_0^s = x_0 / s_x, y_0^s = y_0 / s_y$,
          $I_s(x_0^s, y_0^s) = Z_0$ if $Z_0 < I_s(x_0^s, y_0^s)$
      b.  Do the regular sampling in the $\{C_i^o\}$
          For every sampled pixel $(x_f^s, y_f^s)$ in $\{c_i^o\}$,
            i.  $x_l = x_f^s * s_x, y_l = y_f^s * s_y$
                if $(x_l, y_l) \in P_i$,
                  $I_s(x_f^s, y_f^s) = Z_0$ if $Z_0 < I_s(x_0^s, y_0^s)$
                else
                  Sample four more points (uniformly distributed) in the same cell $c_i^o$, and then repeat step $i$,
                  stop if a pixel belongs to $P_i$
          End for
    End for
```

# Result of Smart Sampling



# Outline of the Presentation

- **Introduction**
- **Related Work**
- **Research Topics and Methodology**
- **Summary and Future Work**
  - Summary
  - Remaining Work

# Summary

- Model:
  - Extend the previous work on stereo mosaics from static 3D scenes to **dynamic** 3D scenes
- Algorithm:
  - An effective and efficient **patch-based stereo matching** algorithm
- Experimental analysis:
  - Thorough **experimental analysis** of the robustness and accuracy of parallel-perspective stereo mosaics
- 3D understanding:
  - A **graph-based higher-level scene labeling** approach
- patch-based method in different geometry:
  - Combined with **Smart sampling**

# Outline of the Presentation

- **Introduction**
- **Related Work**
- **Research Topics and Methodology**
- **Summary and Future Work**
  - Summary
  - Remaining work

# Remaining Work

- 3D modeling refinement
  - Scene labeling results may help model refinement
- Apply scene labeling method on real data
  - Car classifier & Road network
- Conduct more experiments on Brainport
  - Work with NYISE

# List of Candidate’s Publications

- ***Journal Paper:***
  - **H. Tang** and Z. Zhu, Content-Based 3D Mosaics for Representing Videos of Dynamic Urban Scenes, *IEEE Transactions on Circuits and Systems for Video Technology*, 2012
  - **H. Tang** and Z. Zhu, Modeling and Representing Large Scale 3D Scenes from Video Images, submitted to International Journal of Computer Vision, 2012

- ***Conference /workshop papers:***
  - **H. Tang**, Z. Zhu, Smart Sampling in 3D Scene Transducing for Helping the Blind, submission to IEEE International Conference on Multimedia and Expo (ICME), July 15 to 19, 2013.
  - **H. Tang**, E. Molina, Z. Zhu and P. Chang, Uncertainty preserving patch-based online modeling for 3D model acquisition and integration from passive motion imagery, SPIE Defense, Security, and Sensing 2012, 23 - 27 April 2012, Baltimore, Maryland, USA
  - **H. Tang** and Z. Zhu, A Segmentation-based Stereovision Approach for Assisting Visually Impaired People, 13th International Conference on Computers Helping People with Special Needs, July, 2012, Austria
  - **H. Tang**, Z. Zhu and J. Xiao, Stereovision-Based 3D Planar Surface Estimation for Wall-Climbing Robots, *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*, October 11-15, 2009, St. Louis, USA
  - **H. Tang** and Z. Zhu, Exploiting Local and Global Scene Constraints in Modeling Large-Scale Dynamic 3D Scenes from Aerial Video, *Workshop on Search in 3D (S3D),* June 27, 2008.
  - E. Molina, **H. Tang**, Z. Zhu, O. Mendoza, Mosaic-based Modeling and Rendering of Large-Scale Dynamic Scenes for Internet Applications, *NAECON 2008 - National Aerospace and Electronics Conference*, Dayton, Ohio, United States, Jul 16-18, 2008
  - **H. Tang**, Z. Zhu, G. Wolberg and J. R. Layne, Dynamic 3D Urban Scene Modeling Using Multiple Pushbroom Mosaics, the *Third International Symposium on 3D Data Processing, Visualization and Transmission (3DPVT 2006),* University of North Carolina, Chapel Hill, USA, June 14-16, 2006.

# List of Candidate’s Publications (cont.)

- ***Conference /workshop papers:***
  - W. Li, **H. Tang** and Z. Zhu, Vision-Based Projection-Handwriting Integration in Classroom, *IEEE International Workshop on Projector-Camera Systems (PROCAMS’06)*, New York City, June 17, 2006
  - Z. Zhu, **H. Tang**, Content-Based Dynamic 3D Mosaics, I*EEE Workshop on Three-Dimensional Cinematography (3DCINE'06)*, June 22, New York City
  - W. Li, **H. Tang**, C. McKittrick and Z. Zhu, Classroom Multimedia Integration, *IEEE International Conference on Multimedia & Expo (ICME)*, Toronto, Canada, July 9-12 2006
  - Z. Zhu, **H. Tang**, B. Shen, G. Wolberg, 3D and Moving Target Extraction from Dynamic Pushbroom Stereo Mosaics, *IEEE Workshop on Advanced 3D Imaging for Safety and Security*, June 25, 2005, San Diego, CA, USA

# Sponsored Projects

- Air Force Research Laboratory
  - Award No. FA8650-05-1-1853 (RASER)
- Army Research Laboratory
  - Award No W911NF-05-1-0011 (City Climber)
- National Science Foundation
  - Award No. CNS-0551598 (PRISM)
  - Award No. EFRI-1137172 (M3C)

- Thank you and questions?