[Paper Seminar 7]

CVPR2003, Vol.1, pp.195-202

High-Accuracy Stereo Depth Maps Using Structured Light

Daniel Scharstein
Middlebury College

Richard Szeliski
Microsoft Research

2012. 05. 30.
Yeojin Yoon

Graphics & Virtual Reality Lab

SEJONG UNIVERSITY

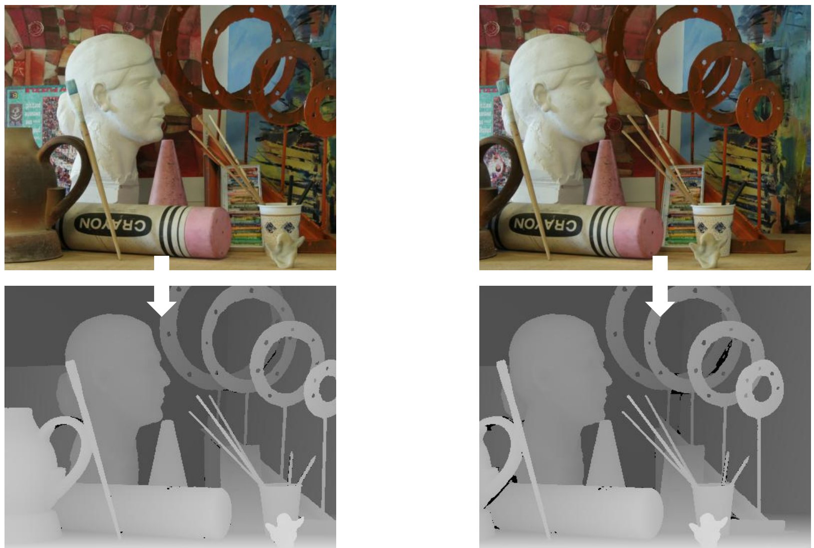
- Introduction
- Overview of our approach
- Structured light
- Disparity computation
- Results
- Conclusion

CONTENTS

Graphics & Virtual Reality Lab

SEJONG UNIVERSITY

1 Introduction



1 Introduction

- **In this paper…**
  - Automatically acquire high-complexity stereo image pairs with pixel-accurate correspondence information

1 Introduction

- **Previous work**
  - Stereo correspondence algorithm
    - hand-labeling a small number of images[Nakamura et al. 96]
    - setting up scenes with a small number of slanted planes[Scharstein et al. 02]
    - Synthetic images[Hoff et al. 1989, Frohlinghaus et al. 96]

1 Introduction

- **Previous work**
  - Structured-light
    - Projecting one or more special light patterns onto a scene
    - Directly acquire a range map of the scene
    - Using single camera and a single projector
    - Random light patterns to provide artificial texture[Kang et al. 95]
    - Register range data with stereo image pairs[Mulligan et al. 01]

2 Overview of our approach

- **Experimental setup**

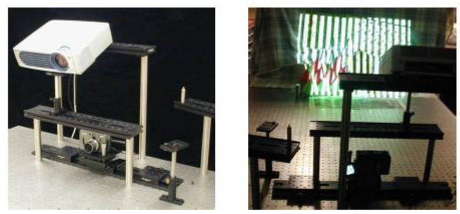


Figure 1. *Experimental setup, showing the digital camera mounted on a translation stage, the video projector, and the complex scene being acquired.*

  - Single digital camera(Canon G1), one or two light projectors from different direction

2 Overview of our approach

- **Terminology**
  - Two primary camera views L(left), R(right)
  - The illumination sources from which light patterns are projected are {0, 1, …}

2 Overview of our approach

- **Processing pipeline**

1. Acquire all desired views under all illuminations
2. Rectify the images to obtain the usual horizontal epipolar geometry, using either a small number of corresponding features or dense 2D correspondences.
3. Decode the light patterns to get (u, v) codes at each pixel in each view.
4. Use the unique codes at each pixel to compute correspondences. The results of this correspondence process are the *view disparities*.

2 Overview of our approach

5. Determine the projection matrices for the illumination sources from the view disparities and the code labels.
6. Reproject the code labels into the two-view geometry. This results in the *illumination disparities*.
7. Combine disparities from all different sources to get a reliable and accurate final disparity map
8. Optionally crop and downsample the disparity maps and the views taken under ambient lighting.

3 Structured light

- We project a series of structured light images onto the scene and decode the set of projected intensities at each pixel to give it a unique label.
- Two kinds of structured light : *binary Gray-code patterns*, and *series of sine waves*.

3 Structured light

- **Gray codes**
  - Only contain black and white(on/off) pixel values
  - For our projector with 1024 x 768 pixels, it is sufficient to illuminate the scene with 10 vertical and 10horizontal patterns encode the (u, v) position at each pixel.
  - Decide whether it is illuminated or not
    - Two reference images(all-white and all-black)
    - Reference images measure the albedo of each scene point
    - This does not work well due to interreflections in the scene and "fogging" inside the projector.

3 Structured light

– Only reliable way of thresholding pixels into on/off is to project both the code pattern and its inverse.

– Using patterns and their inverses may not be enough!

→ Use two different exposure times(0.5 and 0.1 sec.)

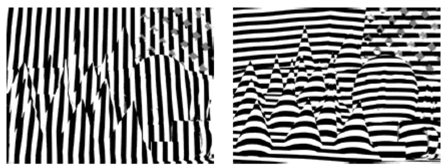


Figure 2. *Examples of thresholded Gray-code images. Uncertain bits are shown in gray. (Full-size versions of all images in this paper are available at http://www.middlebury.edu/stereo/.)*

Graphics & Virtual Reality Lab

SEJONG UNIVERSITY

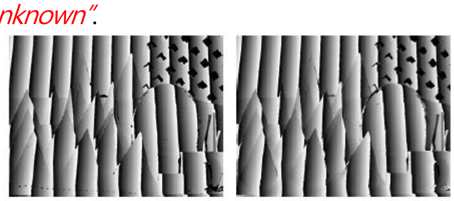
3 Structured light

– If this largest difference is still below a threshold(color bands <32), the pixel is labeled *"unknown"*.

– The projectc ra has 2048 x 1536

→ Each illur vide.

1. Clean up t caused by unknown bit values.



(a): Gray code (b): sine wave

Figure 4. *Computed u coordinates (only low-order bits are shown).*

2. We then ir lution and avoid multiple pixels with the same code.

Graphics & Virtual Reality Lab

SEJONG UNIVERSITY

3 Structured light

- **Sine waves**
  - Projecting a continuous function onto the scene takes advantage of the gray-level resolution available in modern LCD projectors.
  - A sine wave pattern avoids the discontinuities of a sawtooth, but introduces a further two-way ambiguity in phase.
  - Our algorithm
    - The first frequency has a period equal to whole image width or height.
    - The second has 10 periods per screen.

3.2 Sine waves

- **How do we compute the phase and hence(u, v) coordinates at each pixel?**

$$\vec{I}_{kl}(x, y) = \vec{A}(x, y)B_{kl}[\sin(2\pi f_k u + \emptyset_l) + 1] \qquad (1)$$

  - $\vec{A}(x, y)$: the albedo corresponding to scene pixel(x, y)
  - B_{kl} : intensity of the (k, l)th projected pattern
  - f_k : frequency
  - $\emptyset_l$: phase
  - Similar equation can be obtained for horizontal sine wave patterns by replacing u with v.

3.2 Sine waves

- We only have a single frequency $\boldsymbol{f_k}$ and let $\boldsymbol{c_l = cos\emptyset_l, s_l = sin\emptyset_l, c_u = cos(2\pi f_k u), s_u = sin(2\pi f_k u)}$, and $\vec{\boldsymbol{C}} = \vec{\boldsymbol{A}}(\boldsymbol{x, y})\boldsymbol{B}$.

$$\vec{I}_{kl} = \vec{C}[s_u c_l + c_u s_l + 1] \qquad (2)$$

- As estimate of the u signal can then be recovered using

$$u = p_u{}^{-1}(\frac{1}{2\pi} tan^{-1}\frac{s_u}{c_u} + m) \qquad (3)$$

- $p_u = W/f_u$: sine period(in pixels)
- m : integral phase wrap count
- To solve the phase wrapping problem, we first estimate the value of u using wave($f_1 = 1$), and then repeat the estimation with $f_2 = 10$

3.2 Sine waves

- Since we are using least squares, we can compute a certainty for $\boldsymbol{u}$ estimate.
- Computing certainties allows us to merge estimates from different exposures

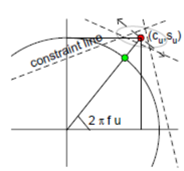
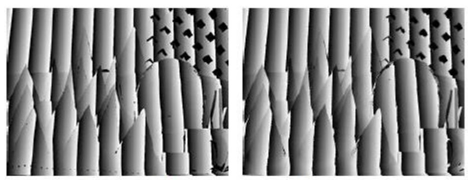


Figure 3. *Phase estimation from (c_u, s_u) least squares fit. The red dot is the least squares solution to the constraint lines, and the ellipse around it indicates the two-dimensional uncertainty.*

3.2 Sine waves

– Results of recovering the u positions using sine patterns

– For these experiments, we use all 12 phase($\emptyset = 0^{\circ}, 30^{\circ}, \ldots, 330^{\circ}$) and two different exposures(0.1 and 0.5sec)



(a): Gray code (b): sine wave

Figure 4. *Computed u coordinates (only low-order bits are shown).*

Graphics & Virtual Reality Lab SEJONG UNIVERSITY

3.3 Comparison

- **The total number of light patterns**
  - 80 for the Gray codes
  - 100 for the sine waves
  - Gray codes yield better(less noisy) results
  - Sine wave patterns
    - potentially yields higher resolution
    - could be done with fewer images
    - more susceptible to non-linearities of camera and projector and to interreflections in the scene

Graphics & Virtual Reality Lab SEJONG UNIVERSITY

4 Disparity computation

- Given N illumination sources, the decoding stage yields a set of labels $\left(u_{ij}(x,y), v_{ij}(x,y)\right)$, for each illumination $i \in \{0, \ldots N-1\}$ and view $j \in \{L, R\}$.
- Uniquely identify each scene point
- Encode the coordinates of the illumination source

Graphics & Virtual Reality Lab

SEJONG UNIVERSITY

4.1 View disparities

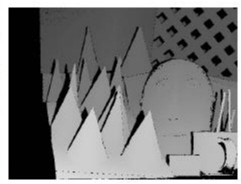
- **Several practical issues**
  - Some pixels may be partially occluded
  - Some pixels may have unknown code value in some illumination due to shadows and reflections
  - A perfect matching code value may not exist due to aliasing or interpolation errors.
  - Several perfect matching code values may exist due to the limited resolution of the illumination source.
  - The correspondences computed from different illuminations may be inconsistent.

Graphics & Virtual Reality Lab

SEJONG UNIVERSITY

4.1 View disparities

- **The number of unknown code values**
  - Reduced by using more than one illumination source
  - Consistency check
    - All points illuminated by at least one source
    - Establish high-confidence correspondences
    - Establish disparities d_{LR} and d_{RL} independently and cross-check for consistency

Graphics & Virtual Reality Lab

SEJONG UNIVERSITY

4.2 Illumination disparities

- Compute another set of disparities:
  →Between the cameras and the illumination sources
- The role of one camera is played by the illumination source(traditional structured lighting system)
- Our final goal is…
  - Express all disparities in the rectified two-view geometry
  - we can treat the view disparities as a 3D reconstruction of the scene
  - Solve for the projection matrix of each illumination source

Graphics & Virtual Reality Lab

SEJONG UNIVERSITY

4.2 Illumination disparities

- **Left view L and illumination source 0**
  - Scene point $S = [x\, y\, d\, 1]^T$ with projective depth $d = d_{LR}(x, y)$.
  - Pixel's code values (u_{0L}, v_{0L}) also represent its x and y coordinates in the illumination pattern
  - We can writhe these coordinates as homogenous 2D point $P = [u_{0L}\ v_{0L}\ 1]^T$

$$P \cong M_{0L} S$$

    - M_{0L} is the unknown 4 x 3 matrix of illumination source 0 respect to the left camera

4.2 Illumination disparities

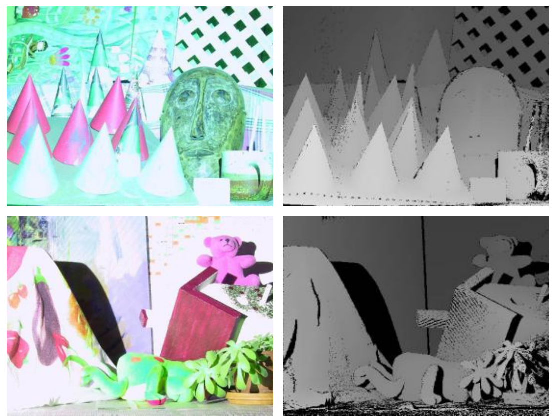
- m_1, m_2, m_3 denote the three rows of M_{0L}, this yields

$$\begin{aligned} u_{0L} m_3 S &= m_1 S, \\ v_{0L} m_3 S &= m_2 S \end{aligned} \qquad (4)$$

  - We set $m_{34} = 1$
- Small number of pixels with large disparity errors can strongly affect the least-square fit. →outlier detection by iterating the above process(4 iterations with successively lower error threshold)
- Given the projection matrix M_{0L}, we can now solve Equation (4) for d at each pixel.

4.2 Illumination disparities

- These disparities are available for all points illuminated by source 0, even those that are not visible from the right camera.



4.3 Combining the disparity estimates

- **Remaining task is to combine 2N + 2 disparity maps**
- **1st step**
  - Create combined maps for each of L and R separately
- **2nd step**
  - The left and right(combined) maps are checked for consistency
  - For unoccluded pixels, this means that

$$d_{LR}(x, y) = -d_{LR}(x + d_{LR}(x, y), y)$$

and vice versa.

4.3 Combining the disparity estimates

4.3 Combining the disparity estimates

- Two final steps
- Cropping
  - Full size images(2048 x 1536), disparity range from 210 to about 450
  - Bring disparities closer to zero by cropping(0-240, image width of 1840)
- Downsampling
  - Most current stereo implementations work with much smaller image sizes and disparity ranges
  - Downsample the images and disparity maps to quarter size(460 x 384).

4.3 Combining the disparity estimates

- **A remaining issue is that of holes(unknown disparity values)**
  - Small holes : filled by interpolation
  - Large holes : remain in areas where no illumination codes
  - Two main sources for this:

    (1) Surfaces that are highly specular or have very low albedo

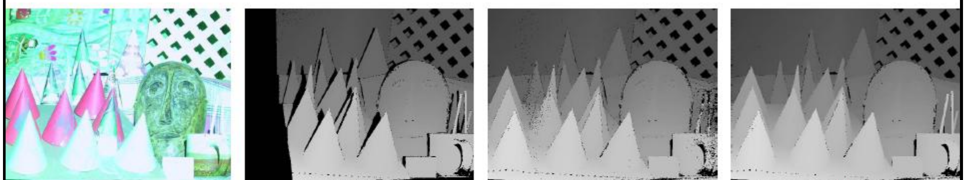
    (2) areas that are shadowed under all illuminations

5 Results

Figure 5. *The two image pairs:* Cones *(left) and* Teddy *(right).*

- Views 3 and 7 out of total of 9 images
- The Cones scene was constructed with a single light source
- For the Teddy we used two illumination directions

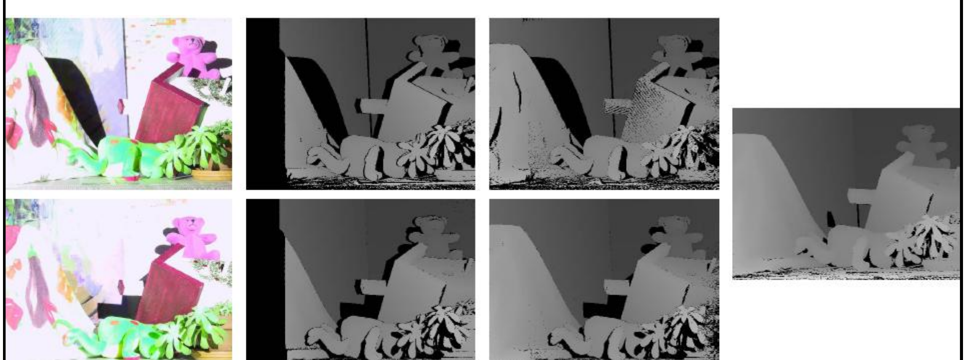
5 Results



(a) (b) (c) (d)

Figure 6. *Left view of* Cones *(one illumination source is used): (a) scene under illumination (note absence of shadows except in upper-right corner); (b) view disparities; (c) illumination disparities; (d) final (combined) disparity map. Unknown disparities are shown in black.*

5 Results



(a) (b) (c) (d)

Figure 7. *Left view of* Teddy *(two illumination sources are used): (a) scene under the two illuminations; (b) view disparities; (c) illumination disparities; (d) final (combined) disparity map. Unknown disparities are shown in black.*

5 Results

- Comparisons with several algorithms

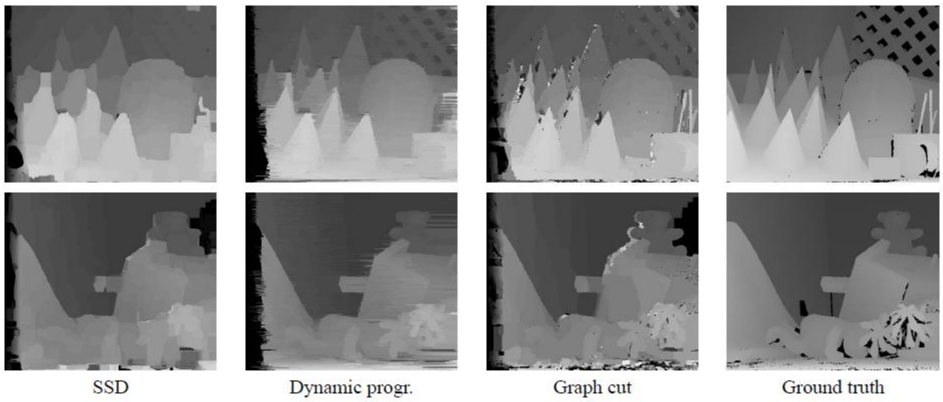


Figure 8. *Stereo results on cropped and downsampled images:* Cones *(top) and* Teddy *(bottom).*

Graphics & Virtual Reality Lab

SEJONG UNIVERSITY

5 Results

- Quantitative performance in non-occluded areas
  - Difficulties to the matching algorithms
    - Large disparity range
    - Complex surface shapes
    - Textureless areas
    - Narrow occluding objects
    - Ordering-constraint violations

Bad %	Cones		Teddy	
	$t=1$	$t=2$	$t=1$	$t=2$
SSD	17.8%	9.3%	26.5%	12.8%
DP	17.1%	9.8%	30.1%	10.5%
GC	12.6%	7.0%	29.3%	11.4%

Table 1. *Performance of SSD, dynamic programming, and graph cut stereo methods on our data sets. The table shows the percentage of pixels whose disparity error is greater than threshold t for* $t=1,2$.

Graphics & Virtual Reality Lab

SEJONG UNIVERSITY

6 Conclusion

- New methodology to acquire highly precise and reliable ground truth disparity with stereo image pairs
- These high-quality data is essential to evaluate the performance of stereo correspondence algorithms.
- The structured lighting enables us to uniquely code each scene pixel, which makes inter-camera correspondence much easier and more reliable
- We have two different kinds of structured light: binary Gray codes, and continuous sine waves
- Gray codes give us more reliable estimates of projector coordinates.

6 Conclusion

- **Future work**
  - We plan to develop the sine wave approach
  - Reduce number of acquired images
  - Fill in the remaining pixels marked as unknown in disparity maps

Yeojin Yoon
yjyoon@gce.sejong.ac.kr

THANK YOU.

Graphics & Virtual Reality Lab

SEJONG UNIVERSITY