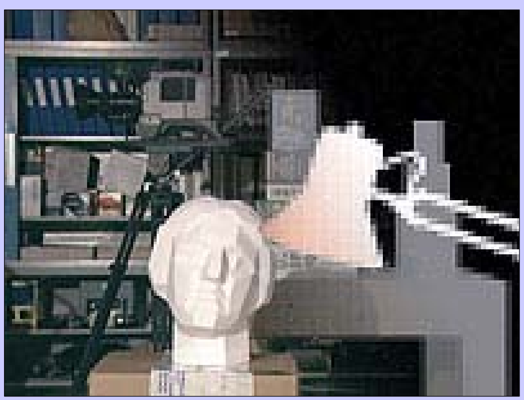
Berlin University of Technology

A Benchmarking Dataset for Surface Reconstruction Algorithms

Anke Bellmann, Olaf Hellwich, Volker Rodehorst and Ulaş Yılmaz

Computer Vision & Remote Sensing
Berlin University of Technology



www.cv.tu-berlin.de

Introduction

- **ISPRS Photogrammetric Computer Vision & Image Analysis WG III/2**
  - Describe and improve **automatic surface reconstruction methods**
    - Shape-from-**stereo**
    - Shape-from-**motion**
    - Shape-from-**silhouette**
    - Shape-from-**texture**
    - Shape-from-**shading**
    - Shape-from-**(de-)focus**
  - No taxonomy exist to **compare the performance** of these methods
- **Contribution**
  - A **benchmark dataset** for the evaluation of shape-from-*X* algorithms
  - A methodology for comparing their results

Problem

- **Controversial Requirements**
  - Some methods prefer **textured** and/or **piecewise planar** objects
  - Other approaches deal with **curved Lambertian** surfaces
- **Different Results**
  - Dense **disparity** or **depth maps**
  - Unstructured **point clouds** or **polygon meshes**
  - Surface **orientations**
- **Usefulness**
  - A comparison is very **general**
  - The algorithms have
    - different properties
    - overall characteristics
    - rely on different assumptions
  - The lowest common denominator is small

Outline

- Introduction
- Benchmarking Data
  - Object selection
  - Calibration (camera, texture/light projector)
  - Image acquisition
  - Ground truth generation
- Surface Reconstruction from Images
  - Shape-from-stereo
  - Shape-from-motion
  - Shape-from-silhouette
  - Shape-from-texture
  - Shape-from-shading
  - Shape-from-(de-)focus
- Evaluation Procedure
- Conclusions and Future Work

Object Selection

- **Challenge**
  - Find an object that satisfies the requirements of all reconstruction algorithms
- **Surface Geometry**
  - **Problem:** Smooth curved parts vs. piecewise planar patches
  - **Solution:**
    - Combine various objects with different surface geometries
    - Moderate occlusions
- **Surface Radiometry**
  - **Problem:** Lambertian surface vs. enough textural information
  - **Solution:**
    - White plaster approximate matte reflectance properties
    - Texture is projected
  - **Remarks:**
    - The effect of **different texture patterns** on final result can be analyzed
    - Artificial **shadows** penalize texture-based approaches

Benchmarking Scene

Camera Calibration

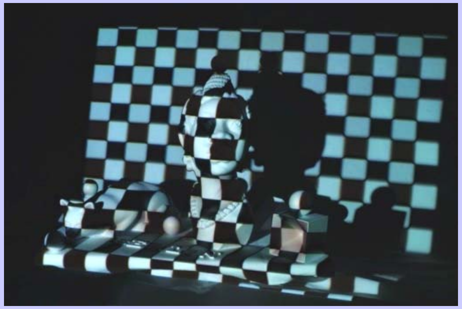
- **Digital SLR Camera Canon EOS-1D Mark II**
  - 28.7×19.1 mm **CMOS** sensor, **50 mm** lens
  - Image size is 3504×2336 pixels (8MP)
  - Scaled and cropped to **1360×904 pixels**
- **Interior Orientation**
  - **Bundle adjustment** software Australis 6
  - Image axes are orthogonal, aspect ratio is 1.0
  - **Radial distortion** < 1.8 pixels is eliminated by bicubic resampling
- **Algebraic Representation**
  - Line-preserving 3×4 **projection matrices** $\mathbf{P}$
  - **Constant calibration matrix** $\mathbf{K}$ with a principle distance in pixels:

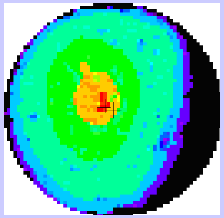
Control point field for camera calibration

$$w \cdot \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \mathbf{P} \cdot \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \qquad \mathbf{P} = \mathbf{KR}\left[\mathbf{I} \middle| \text{-}\mathbf{C}\right] \qquad \mathbf{K} = \begin{bmatrix} 2909.1 & 0 & 748.4 \\ 0 & 2909.1 & 408.7 \\ 0 & 0 & 1 \end{bmatrix}$$

Projector Calibration

- **Texture Projector**
  - LCD-projector with 1024×768 pixel
  - Interior orientation from technical **reference sheet**
  - Exterior orientation is computed by projecting a synthetic **chessboard structure** with known 3D coordinates
  - **Problem:** Raster artifacts with blank images
- **Light Projector**
  - Slide-projector model a **point light source**
  - Positions are measured manually
  - Illumination **directions** are verified by analyzing Styrofoam sphere highlights
  - Verification accuracy is < 2.5 degrees

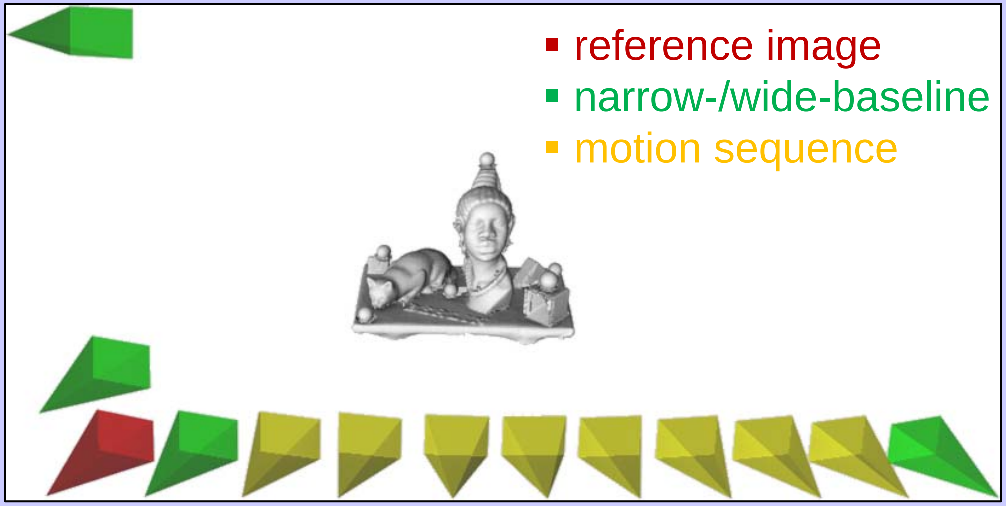
Controlled Camera Motion

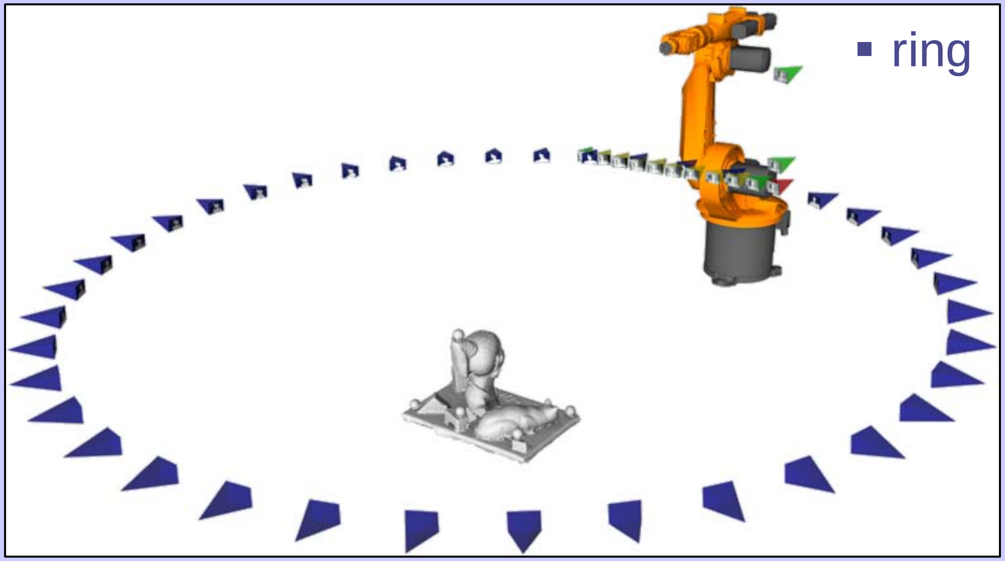
- **Industrial Robot**
  - The camera was firmly mounted on a KUKA KR C1 robot arm
  - Stable and repeatable camera motion (relative precision of 0.1 mm)
  - Remote image capturing through firewire connection
  - Images are acquired at **13 different positions**

- **Turntable**
  - **Scene rotation** simulate a camera motion around the scene.
  - Computer-controlled MICOS Pollux high resolution positioning device
  - Acquisition of **40 images** in 9-degrees steps

  - **Problem:** Rotate the projectors synchronous with the turntable
  - **Solution:** Ambient light in the room and no texture projection

Imaging Geometry

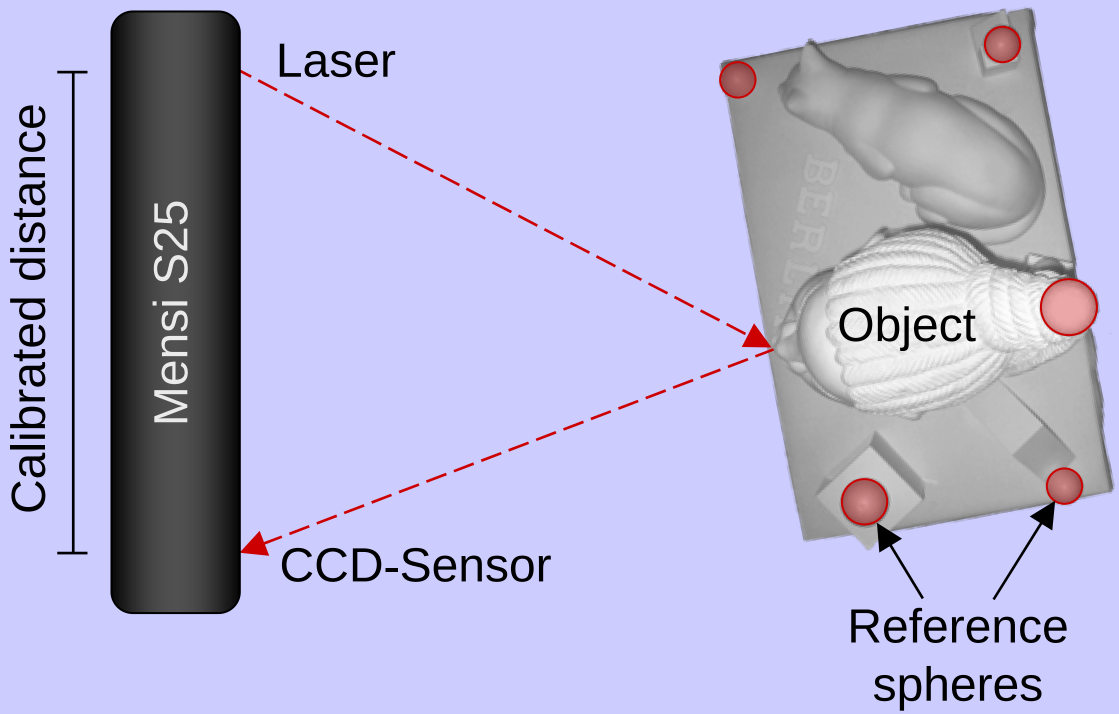




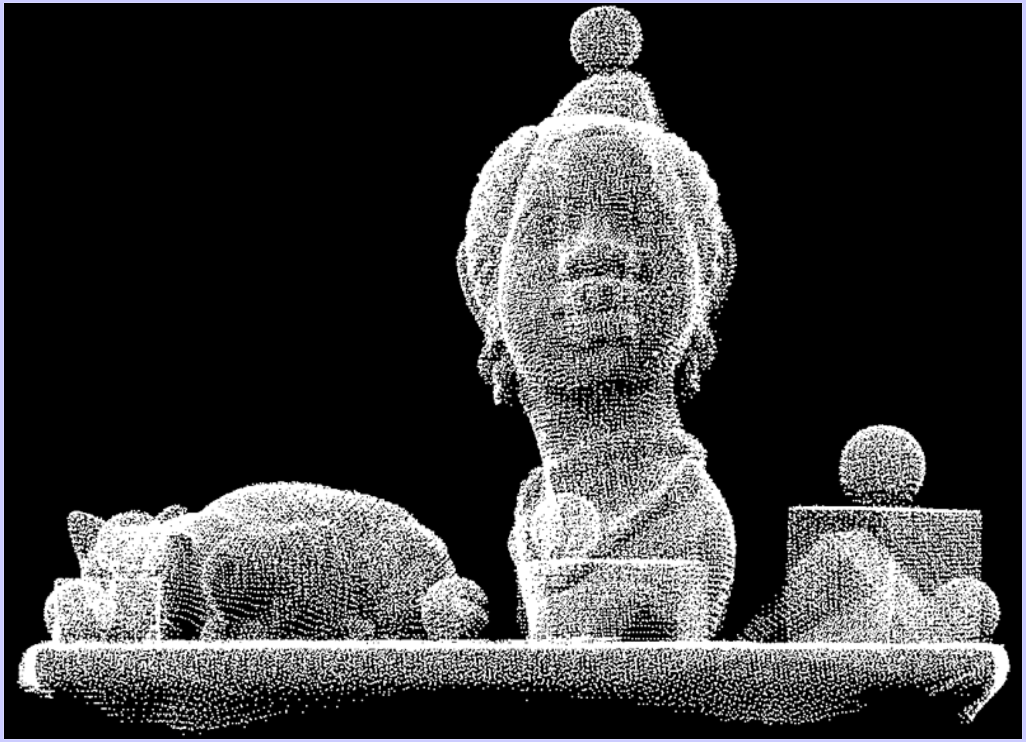
KUKA robot arm with mounted digital camera

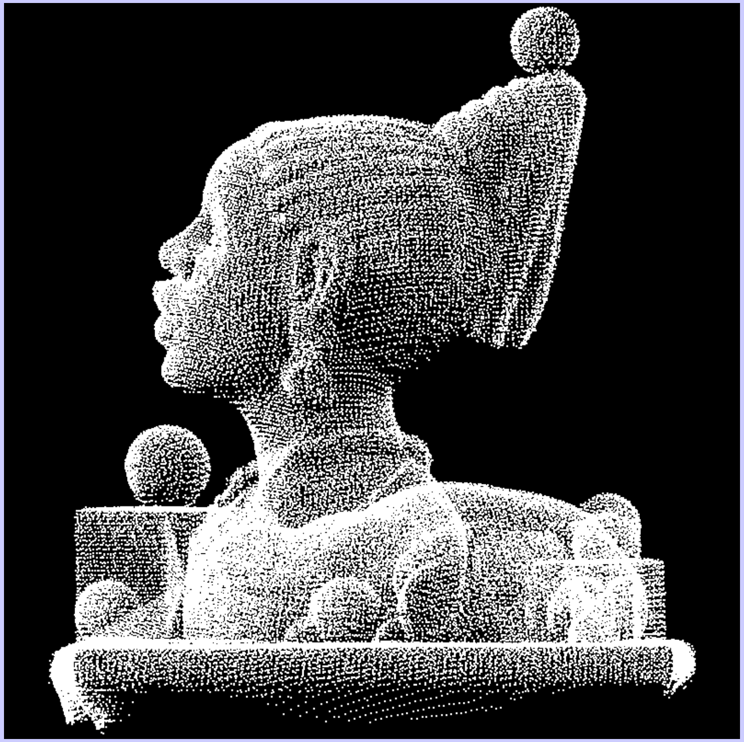
Ground Truth Generation

- Generation of ground truth information using a **laser scanner**
- Triangulation of 3D object points from **7 viewing directions**
- Point cloud registration using **five reference spheres**
- Density of the registered point cloud is 2 mm



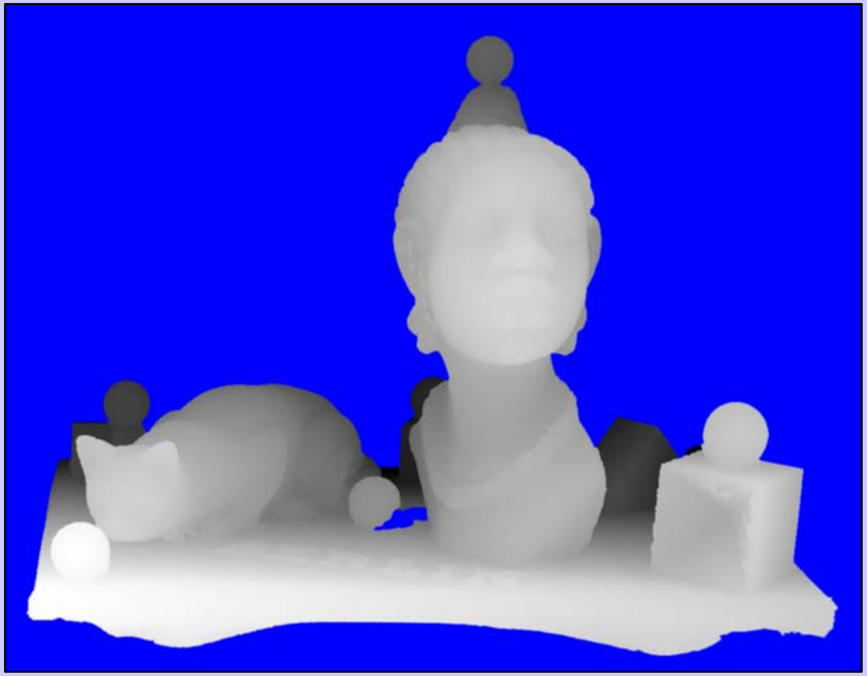
Laser Scanner Accuracy



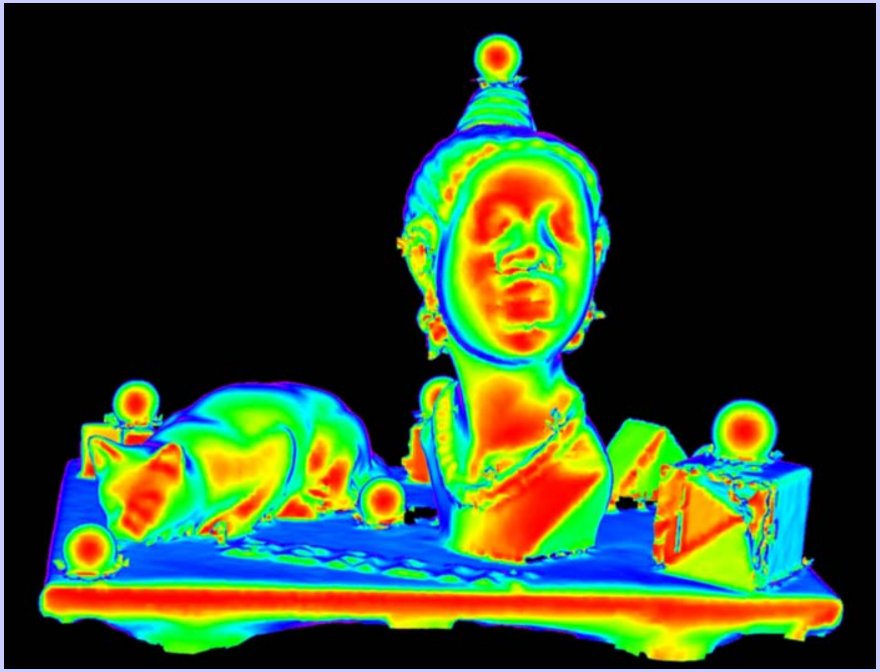


- The **depth accuracy** is 0.2 mm for the object distance of 4 m
- After registering the individual views the **overall accuracy** is 0.8 mm
- Expected depth resolution from images is 1.5 mm
  (e.g. pixel-accurate wide-baseline matching)

Ground Truth Data Maps



Normalized depth map



Color-coded orientation map

- **Depth map:** Each intensity value is related to a depth value
- **Orientation map:** Each pixel is related to a surface normal vector

Outline

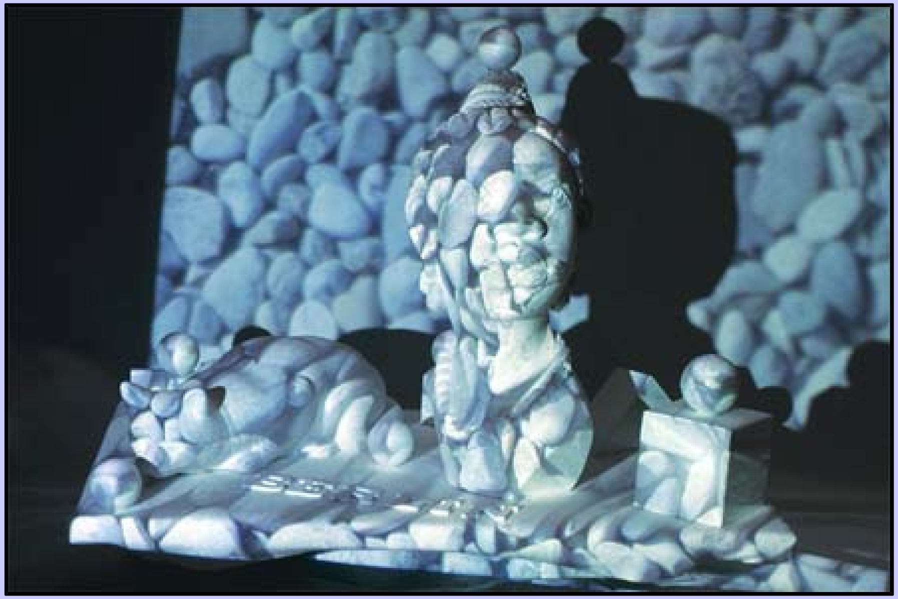
- **Introduction**
- **Benchmarking Data**
  - Object selection
  - Calibration (camera, texture/light projector)
  - Image acquisition
  - Ground truth generation
- **Surface Reconstruction from Images**
  - Shape-from-stereo
  - Shape-from-motion
  - Shape-from-silhouette
  - Shape-from-texture
  - Shape-from-shading
  - Shape-from-(de-)focus
- **Evaluation Procedure**
- **Conclusions and Future Work**

1. Shape-from-Stereo

Stereo is well-known and works analogue to **human vision**

- **Basic Idea**
  - The scene must be observed from **different viewpoints**
  - Search for **corresponding image points** is most challenging
- **Related Work** [Scharstein & Szeliski, 2002]
  - Dynamic programming [Selzer & Yang, 2006]
  - Graph cuts [Kolmogorov & Zabih, 2001]
  - Belief propagation [Klaus, Sormann & Karner, 2006]
  - Semi-global matching [Hirschmüller, 2006]
- **Narrow-baseline Stereo**
  - Small camera distance leads to similar images and **simplifies matching**
  - Inaccurate triangulation for a small angle (**glancing intersection**)
- **Wide-baseline Stereo**
  - Large camera distance makes **spatial intersection more accurate**
  - **Difficult matching** due to perspective distortions and stronger occlusions

Narrow- / Wide-baseline Stereo



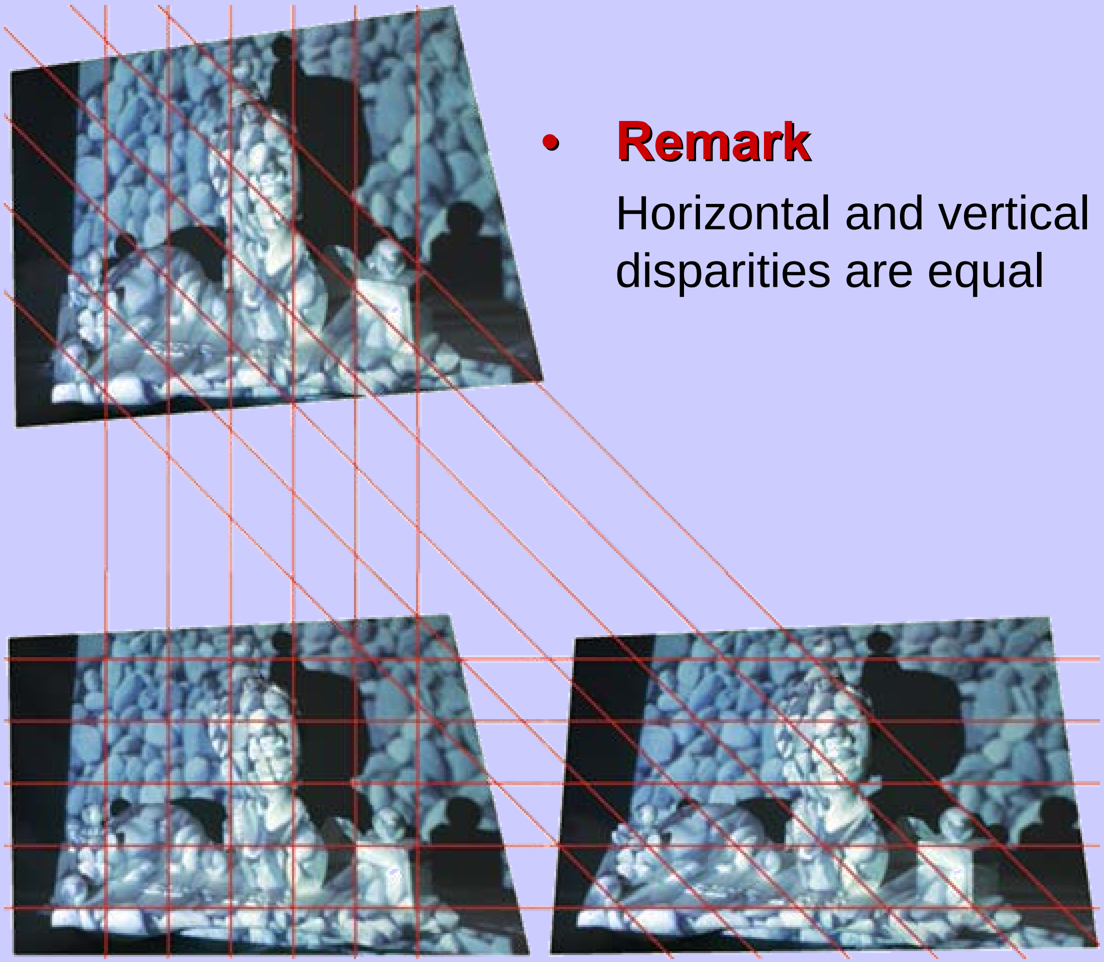
Narrow-baseline (10 cm + 10 cm)

Wide-baseline (1 m + 0.5 m)

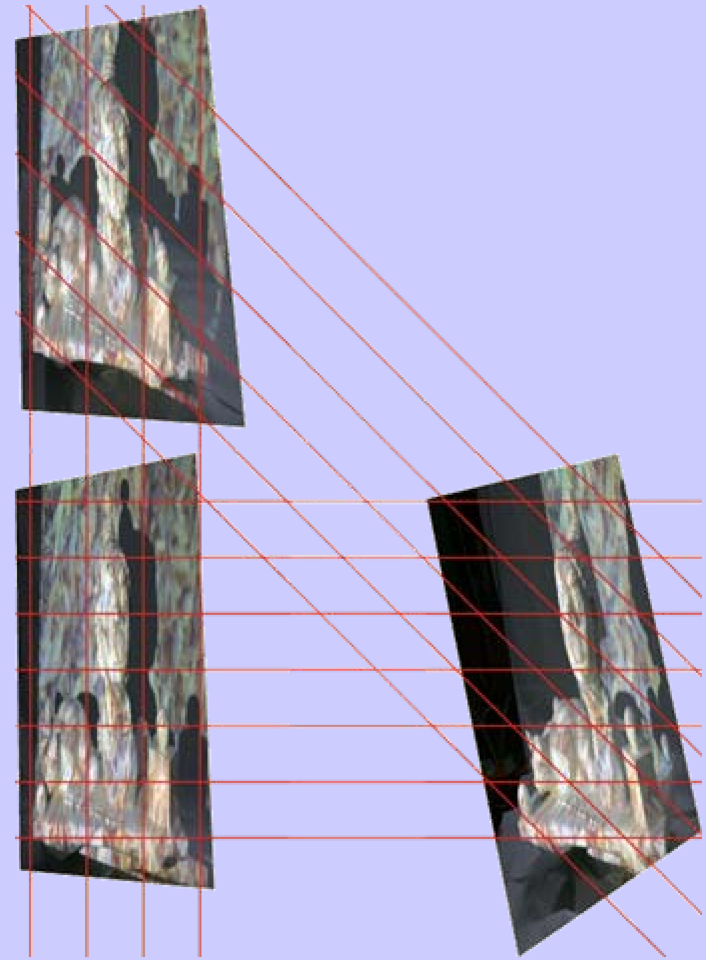
- **Trinocular Stereo**
  - A stereo matching candidate can be **verified** using a third image
  - Convergent stereo triplets are **rectified** to the **stereo normal case**
  - Disparity **search range** is approximately 150 pixels

Rectified Stereo Image Triplets

- **Remark**
  Horizontal and vertical disparities are equal



Narrow-baseline (23 textured image triplets)



Wide-baseline (23 images)

2. Shape-from-Motion

Linear camera motion (253 textured images)

- **Basic Idea**

  Multi-view stereo may combine advantages of narrow- and wide-baseline stereo
  - **Simplified image matching** is using neighboring images
  - **Baseline extension** for spatial intersection by tracking features

3. Shape-from-Silhouette

- **Basic Idea**
  - 3D shape estimation using multiple **contour images** of the object
- **Related Work**
  - Voxel space carving [Martin & Aggarwal, 1983]
  - Marching intersections [Tarini et al., 2002]
  - Generalized cone intersections [Matusik et al., 2000]

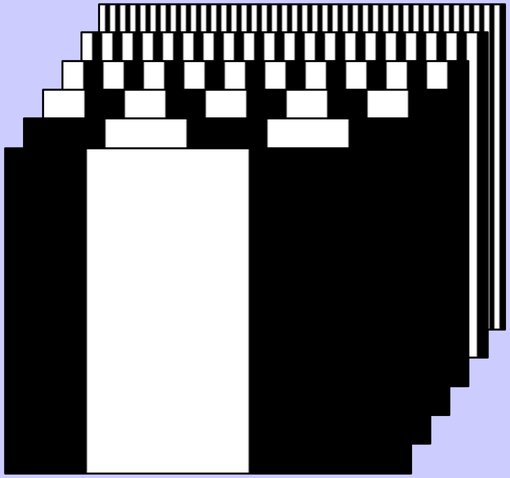
  The resulting **visual hull** still suffers from **concavities** and **insufficient views**
- **Voxel Coloring / Photoconsistency**
  - Visual hull refinement by analyzing **consistent colors** [Seitz & Dyer, 1999]
- **Shadow Carving**
  - Self-shadows may indicate concavities [Savarese et al., 2001]
  - The **illumination direction** must be known
  - Difficult **shadow detection** and the categorization
- **Remark:** No **texture** of **light direction** available for the turntable sequence

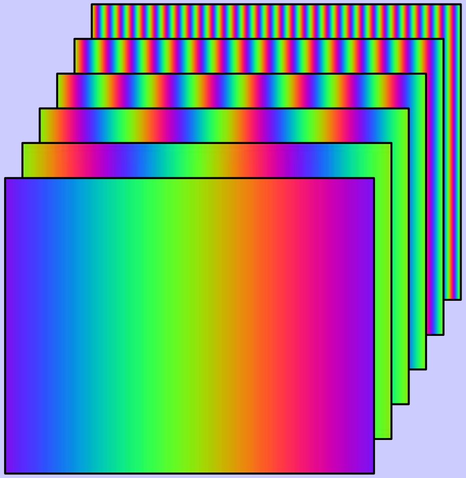
Turntable Sequence

Segmented silhouettes (40 images)

4. Shape-from-Texture

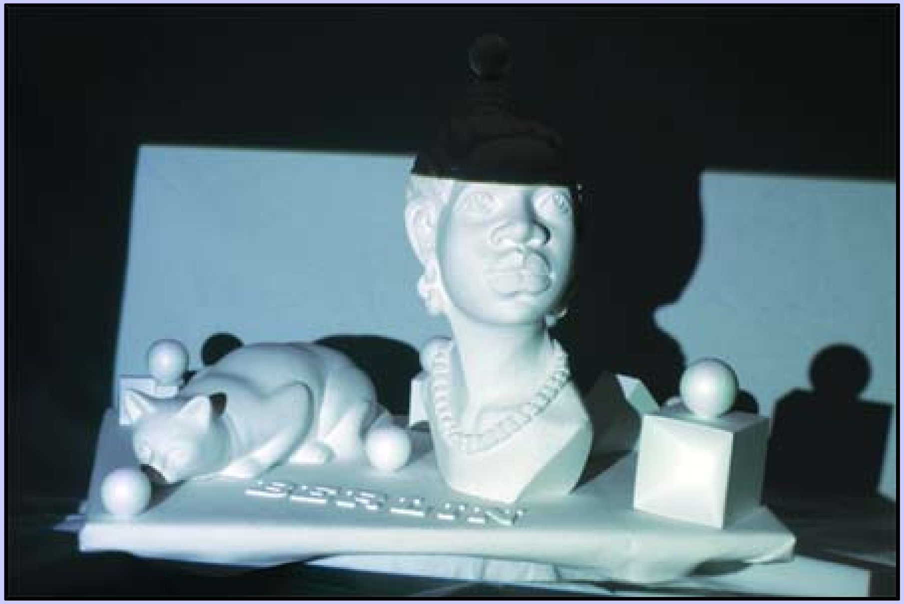
- **Basic Idea**
  - Analyze perspective distortions of individual texture elements (texel)
  - Their variation across the image allows the estimation of surface shape
- **Structured Light** [Scharstein & Szeliski, 2003]
  - **Active triangulation** by replacing the second stereo camera with a projector
  - A calibrated LCD-projector throw **multiple stripe pattern** onto the scene
  - Stripes are **coded** (no correspondence problem)



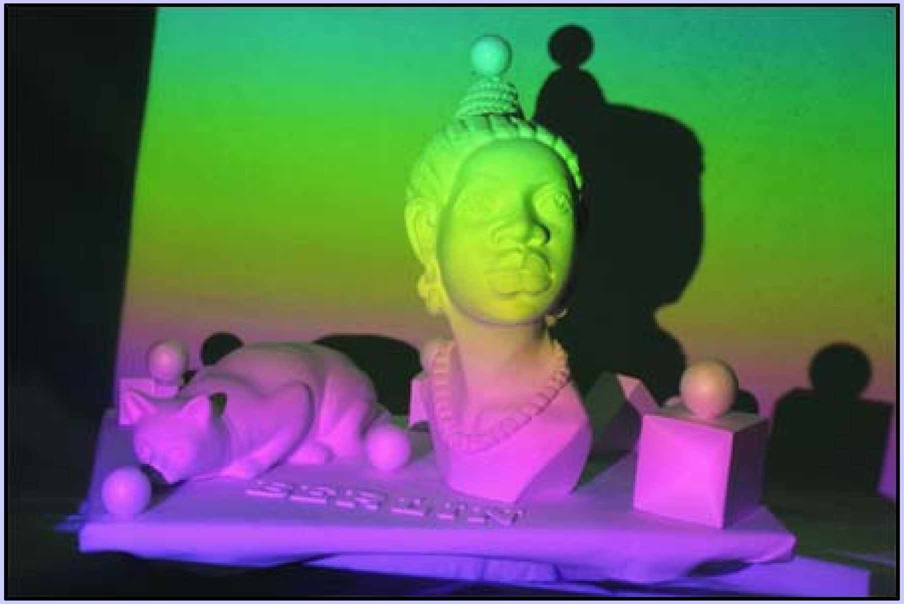


  - Depth can be computed out of the distortion along the detected profiles

Structured Light



Binary coded textures (10+9 images)



Color spectrum textures (10+9 images)

$$S_R = \sin\left(\frac{i}{n}\cdot\pi\right)\cdot\left(\frac{G_{MAX}}{2}-1\right)+\frac{G_{MAX}}{2}$$

$$S_G = \sin\left(\left(\frac{2}{3}+\frac{i}{n}\right)\cdot\pi\right)\cdot\left(\frac{G_{MAX}}{2}-1\right)+\frac{G_{MAX}}{2}, \quad \text{for} \quad i=\{0,\ldots,n\}$$

$$S_B = \sin\left(\left(\frac{4}{3}+\frac{i}{n}\right)\cdot\pi\right)\cdot\left(\frac{G_{MAX}}{2}-1\right)+\frac{G_{MAX}}{2}$$

5. Shape-from-Shading

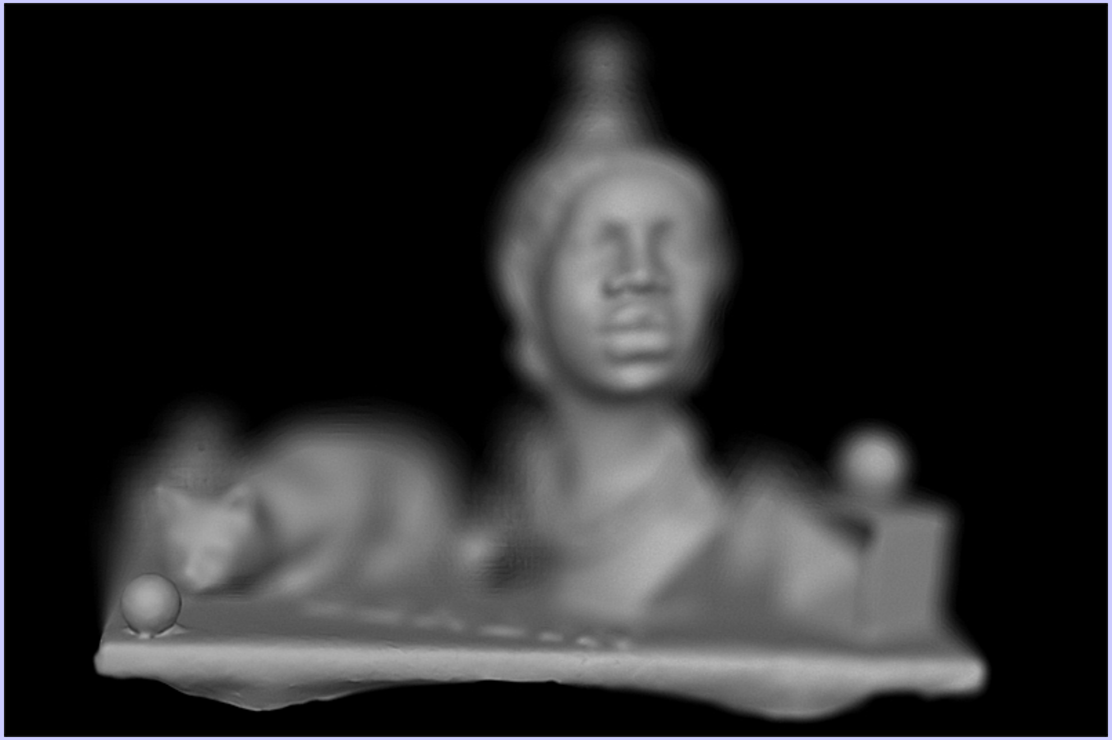
- **Basic Idea**
  - **Single image** technique to estimate a 2.5D surface [Horn & Brooks, 1989]
  - Pixel **intensity** is related to surface patch **orientation**
  - **Lambertian reflectance** is assumed
  - **Known illumination** direction or point light source position
- **Photometric stereo** [Klette, Kozera & Schlüns, 1999]
  - Unique result using 3 monoscopic images with different illumination directions
- **Related Work**
  - Helmholtz reciprocity [Zickler, Belhumeur & Kriegman, 2002]
  - Light transport constancy [Davis, Yang & Wang, 2005]
  - Appearance clustering [Koppal & Narasimhan, 2006]
- **Remark**
  - Reflectance based methods determine surface **orientations** instead of depth

Photometric Stereo

Reference image under different illumination configurations (3 images)

6. Shape-from-(De)Focus

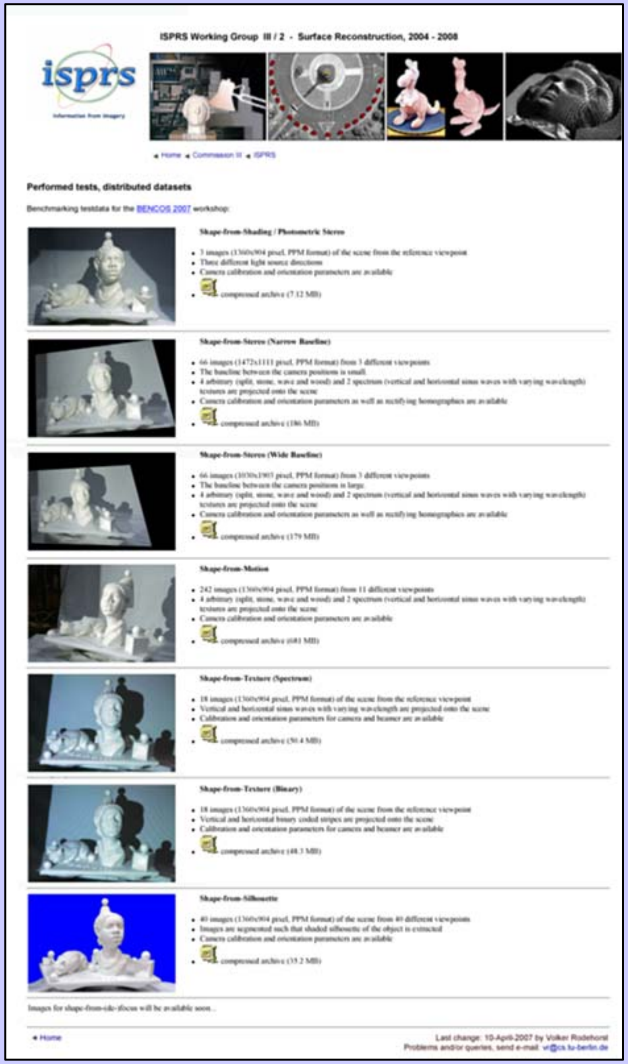
- **Basic Idea**
  - Monoscopic images with **varying focal length** [Favaro & Soatto, 2005]
  - Sharp image regions provide information about the focused depth
  - Requires a sensor with extreme small **depth of field** (e.g. microscope)



Synthetically rendered image sequence (20 images)

Benchmarking Dataset

- **Contents** (~2GB data)
  - 360 **color images** of a real scene
  - 20 **synthetic** rendered scene images
  - 4 real and 38 synthetic texture patterns
  - **Orientation** and **calibration** data
    - 52 camera positions
    - Texture-mapping LCD-projector
    - 3 light source positions
  - **Ground truth** (not public)
    - 3D point cloud
    - Depth map
    - Surface orientation map



Outline

- Introduction
- Benchmarking Data
  - Object selection
  - Calibration (camera, texture/light projector)
  - Image acquisition
  - Ground truth generation
- Surface Reconstruction from Images
  - Shape-from-stereo
  - Shape-from-motion
  - Shape-from-silhouette
  - Shape-from-texture
  - Shape-from-shading
  - Shape-from-(de-)focus
- **Evaluation Procedure**
- **Conclusions and Future Work**

Evaluation Object

- **Problem**
  - Convert different reconstruction results to a **comparable form**

Method	Output
Stereo	Disparity map, depth map
Motion	Point cloud, polyhedron
Silhouette	Volumetric data, polyhedron
Texture	Disparity map, depth map
Shading	Surface orientation, needle map
Focus	Depth map

  - Most formats (i.e. polyhedrons) can easily converted to a **depth map**
  - **Reflectance-based** methods prefer surface **orientation maps**
- **Solution**
  - Compare **2.5D reconstructions** for one **reference view**

Evaluation Criteria

- **Accuracy**
  - Difference between the reconstruction R and ground truth G for n given pixels

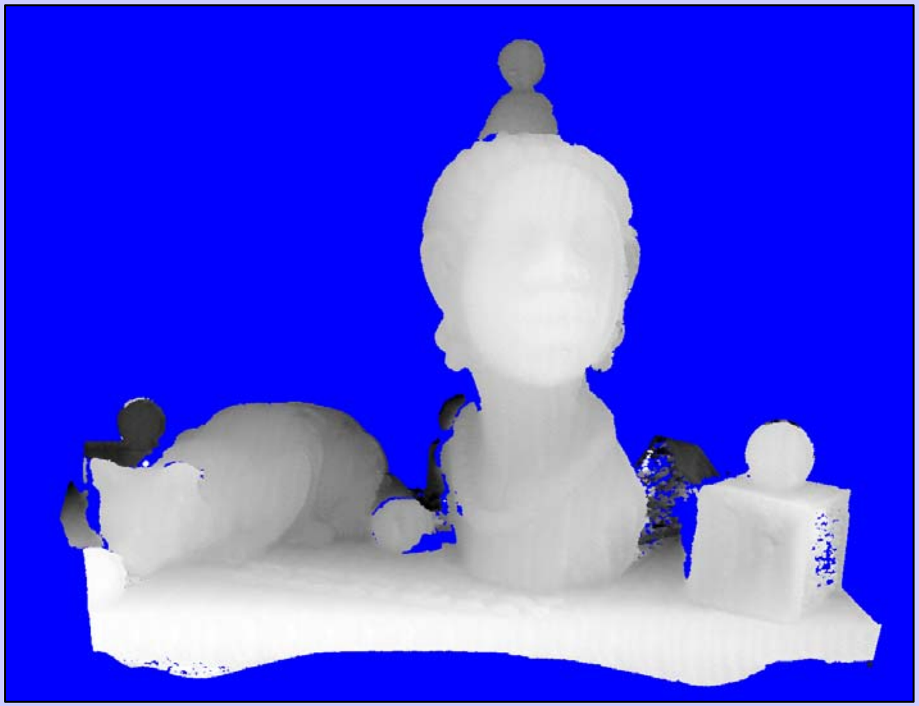
$$a = 1 - \sqrt{\frac{1}{n} \sum_{i,j} \left| \frac{R(i,j) - G(i,j)}{g} \right|^2}$$

- **Completeness**
  - Ratio of well-reconstructed pixels to all possible m pixels

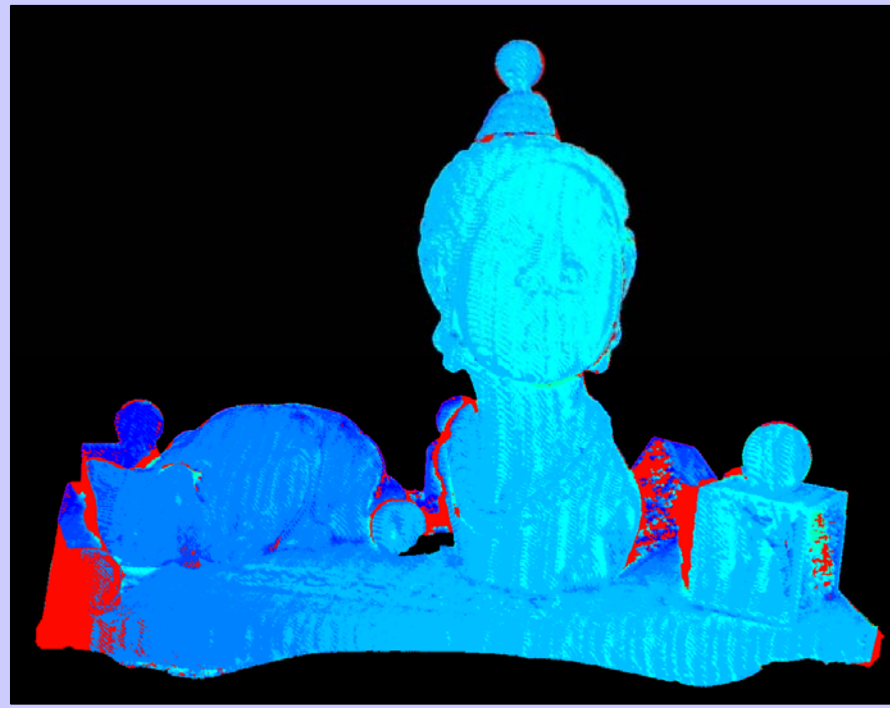
$$c = \frac{1}{m} \sum_{i,j} \left(\frac{R(i,j) - G(i,j)}{g} \leq \delta \right)$$

- **Ranking**
  - Quantitative evaluation allows an **objective comparison** of different methods
  - Provides an informative basis for the **combination** of reconstruction methods

Example: Narrow-baseline Stereo



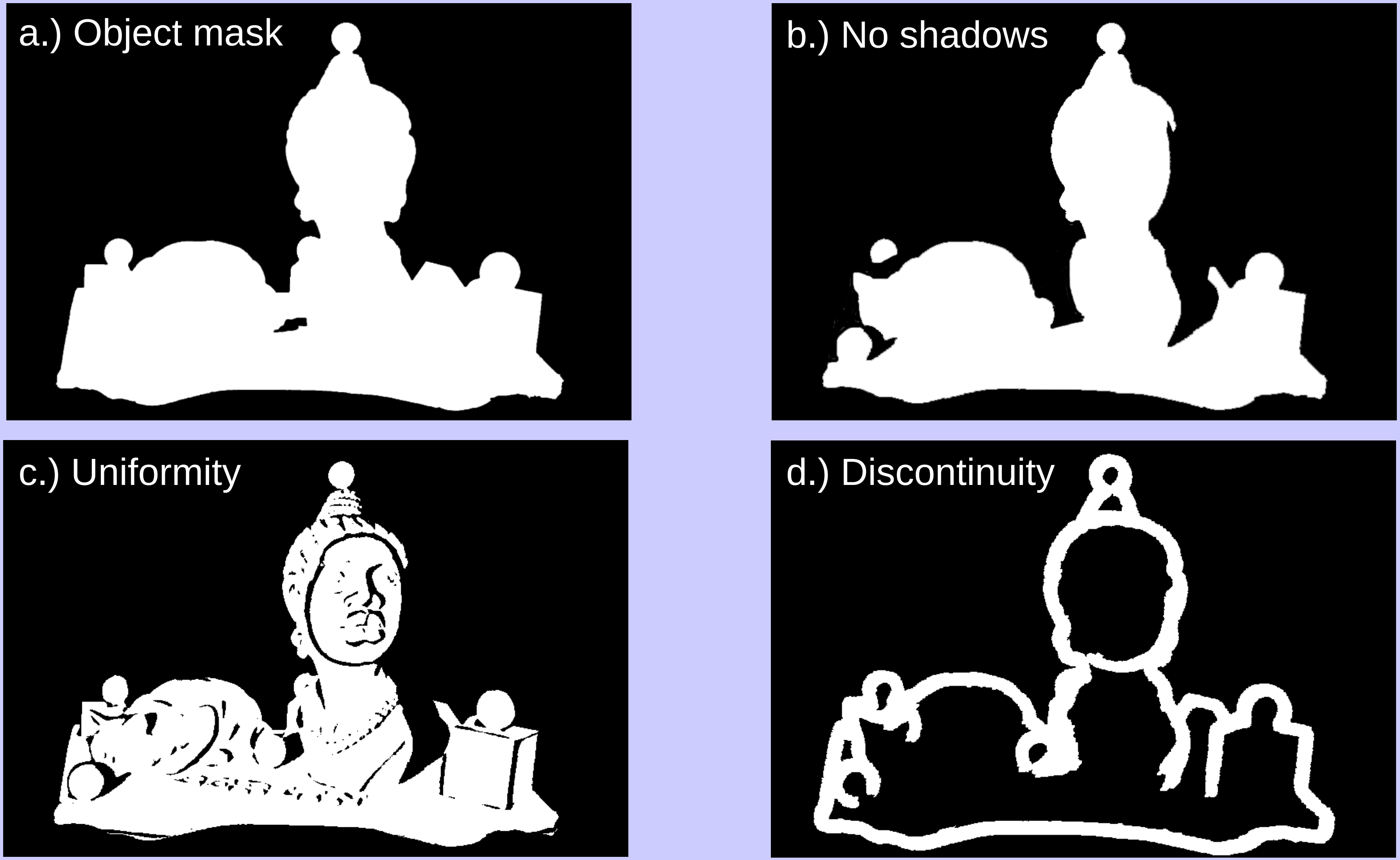
Stereo depth map using modified SGM



Color-coded differences

- **Accuracy:** a = 0.895
- **Completeness:** c = 0.934

Masks for Detailed Investigations



Conclusions

- **Contribution**
  - True **benchmark dataset** for the evaluation of shape-from-*X* algorithms
  - **Test procedure** for an objective comparison of the reconstruction results
- **Invitation**
  - Help to improve **automatic surface reconstruction** techniques
  - **Download** the benchmarking datasets from

    www.cv.tu-berlin.de/isprs/wgiii2
  - Return your **comments**, **experiences** and/or **reconstruction results**
- **Future Work**
  - Benchmarking **contest** (new improved dataset, full resolution)
  - **Online submission** and instant evaluation
  - Textured images and light control for **shape-from-silhouette**
  - More realistic images for **shape-from-(de-)focus**