Challenges on 3D shape acquisition for VR/AR/MR systems

2008 September 8th
Oxford Brookes University, Wheatley

Hiroshi Kawasaki (Saitama University)
joint research with
Katsushi Ikeuchi (University of Tokyo) and
Ryo Furukawa (Hiroshima city University)

My research areas

Reconstruct real-world in a computer
- Acquisition
- How human recognize real-world visually?

- Acquisition
  - Geometry – Active scanning method
  - Photometry –BTF (BRDF) sampling machines
- Modeling
  - Geometry – Registration, integration and hole-filling
  - Photometry – BTF (BRDF) analysis and compression
- Photo-realistic rendering
  - Image based rendering and model based rendering

- Application
  - VR, ITS, CAD, digital archiving and Web

Current research projects

- City modeling project (ITS) granted by motor company, map company and government
- Active 3D scanning granted by MOF
- Texture acquisition and analysis granted by SCOPE
- Digital archiving project (Ikeuchi Lab.)

Research and project matrix

	Acquisition		Modeling		Rendering
	Geometry	Photometry	Geometry	Photometry	
City modeling (ITS)	△	◎		△	◎
Active 3D scanner	◎		◎		
Texture analysis		◎		◎	◎
Digital archiving	◎	△	◎	△	

Research and project matrix

	Acquisition		Modeling		Rendering
	Geometry	Photometry	Geometry	Photometry	
City modeling (ITS)	△	◎		△	◎
Active 3D scanner	◎		◎		
Texture analysis		◎		◎	◎
Digital archiving	◎	△	◎	△	

Research and project matrix

	Acquisition		Modeling		Rendering
	Geometry	Photometry	Geometry	Photometry	
City modeling (ITS)	△	◎		△	◎
Active 3D scanner	◎		◎		
Texture analysis		◎		◎	◎
Digital archiving	◎	△	◎	△	

City modeling for ITS

- Reconstruct a large scale scene in a computer -

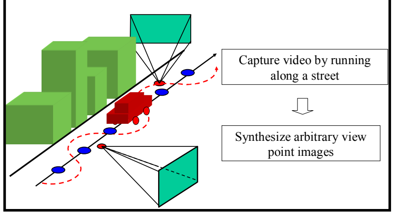
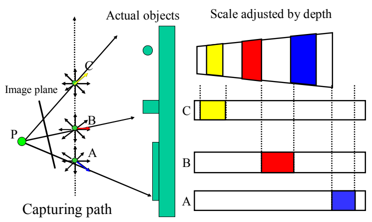
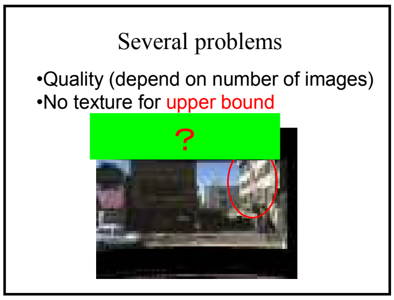


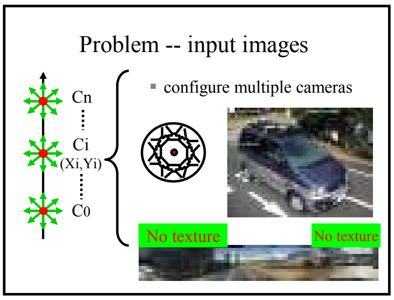
Image based rendering technique

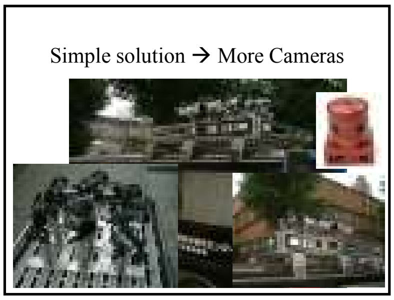
[Kawasaki and Ikeuchi CVPR00]

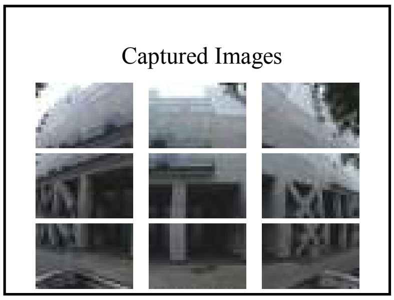


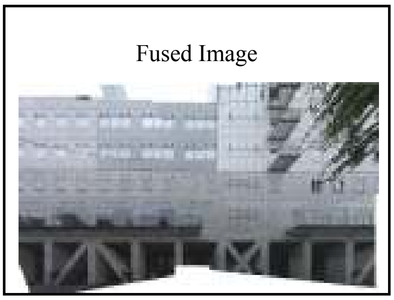
Demo – virtual driving


Several problems
•Quality (depend on number of images)
•No texture for upper bound
?

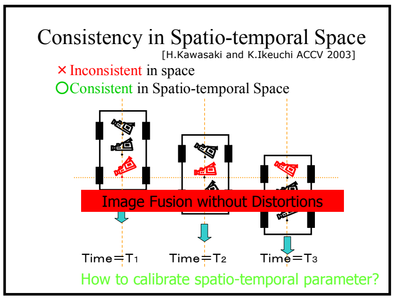

Problem -- input images
Cn
Ci
(Xi,Yi)
C0
configure multiple cameras
No texture
No texture

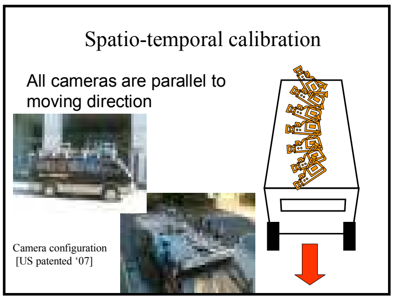

Simple solution → More Cameras

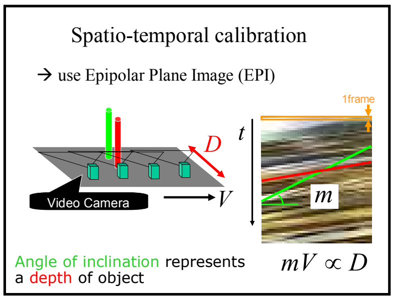

Captured Images

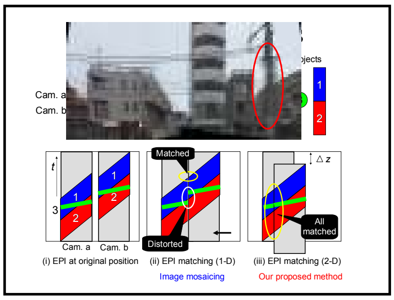

Fused Image

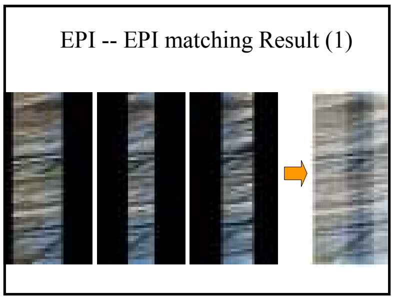

Still problem
Inconsistency of Optical Center

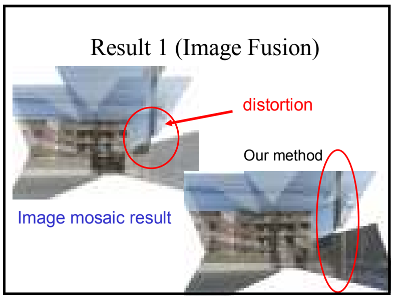

Consistency in Spatio-temporal Space
[H.Kawasaki and K.Ikeuchi ACCV 2003]
× Inconsistent in space
○Consistent in Spatio-temporal Space
Image Fusion without Distortions
Time=T1
Time=T2
Time=T3
How to calibrate spatio-temporal parameter?

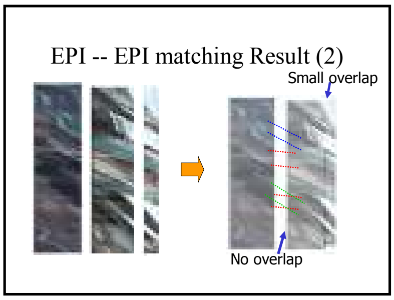

Spatio-temporal calibration
All cameras are parallel to moving direction
Camera configuration
[US patented '07]

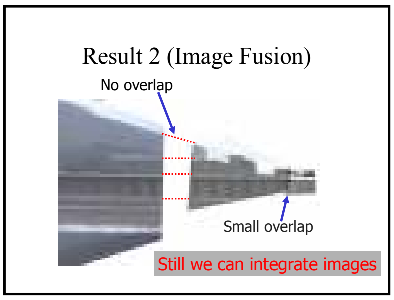

Spatio-temporal calibration
→ use Epipolar Plane Image (EPI)
1frame
D
t
V
Video Camera
m
Angle of inclination represents a depth of object
$mV \propto D$

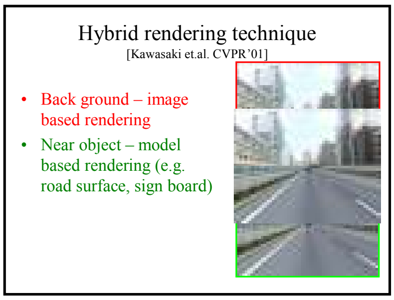

Cam. a
Cam. b
Matched
Distorted
All matched
Cam. a
Cam. b
(i) EPI at original position
(ii) EPI matching (1-D)
(iii) EPI matching (2-D)
Image mosaicing
Our proposed method

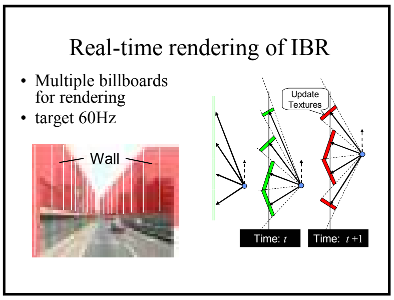

EPI -- EPI matching Result (1)


Result 1 (Image Fusion)
distortion
Our method
Image mosaic result


EPI -- EPI matching Result (2)
Small overlap
No overlap


Result 2 (Image Fusion)
No overlap
Small overlap
Still we can integrate images

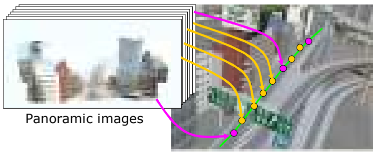

Hybrid rendering technique
[Kawasaki et.al. CVPR'01]
• Back ground – image based rendering
• Near object – model based rendering (e.g. road surface, sign board)


Real-time rendering of IBR
• Multiple billboards for rendering
• target 60Hz
Wall
Update Textures
Time: t
Time: t +1


Near object by MBR
→ Need registration

Registration between map and image

- Using GPS and manual modification



Result – driving simulator

[Kawasaki et.al ACM ISMAR'05]

Current research topics

- Accurate registration of camera position using images (SfM for omini-cameras)
- Real-time rendering using GPU
- Noise removal (reflection on windows, pedestrian and cars)

Research and project matrix

	Acquisition		Modeling		Rendering
	Geometry	Photometry	Geometry	Photometry	
City modeling (ITS)	△	◎		△	◎
Active 3D scanner	◎		◎		
Texture analysis		◎		◎	◎
Digital archiving	◎	△	◎	△	

Why active 3D scanner?

- Passive system
  - ☺Input→only images
  - ☹Unstable
  - ☹Sparse 3D points
- Active system
  - ☺Stable and high accuracy
  - ☺Dense 3D points
  - ☹System→ expensive and heavy

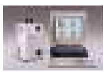
Because…

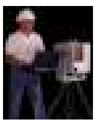
- Passive system
  - ☺Input→only images
  - ☹Unstable
  - ☹Sparse 3D points
- Active system
  - ☺Stable and high accuracy
  - ☺Dense 3D points
  - ☹System→ expensive and heavy

Required for VR and other actual systems!

Active 3D scanner

Minolta (light sectioning)





Cyrax (time of flight)

Z+F (phase-based)

Active Stereo 3D scanner

☺ Stable
☺ Dense 3D points
☹ Complicated and expensive system (lasers, mechanical actuator and laser sensor)

Active Stereo 3D scanner

☺ Stable
☺ Dense 3D points
☹ ~~Complicated and expensive system (lasers, mechanical actuator and laser sensor)~~

Eliminate them by computer vision techniques!

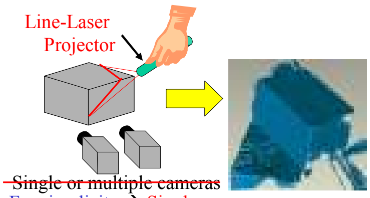
Active Stereo 3D scanner

- ☺ Stable
- ☺ Dense 3D points
- ☹ ~~Complicated and expensive system (lasers, mechanical actuator and laser sensor)~~

Eliminate them by computer vision techniques!

Hand-held 3D scanner

What is a hand-held 3D scanner?

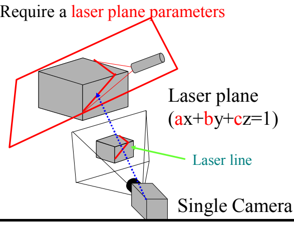


~~Single or multiple cameras~~

For simplicity → Single camera

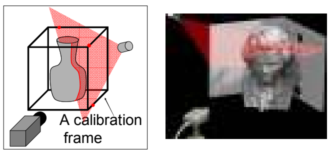
Hand-held scan by single camera

- Light sectioning method (triangulation)
- Require a laser plane parameters



Previous methods

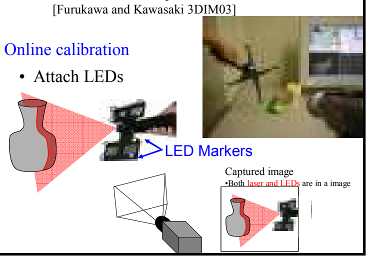
- Online calibration
  - Frame [Chu et.al. 3DIM01]
  - Planes [David 2006 http://www.rob.cs.tu-bs.de/news/david]



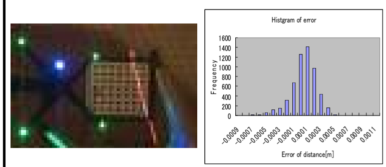
Our first system

[Furukawa and Kawasaki 3DIM03]

- Online calibration
  - Attach LEDs

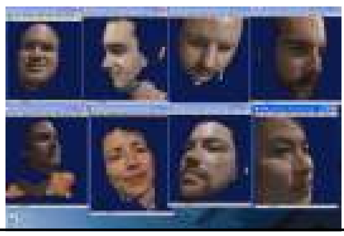


Capturing sequence and precision

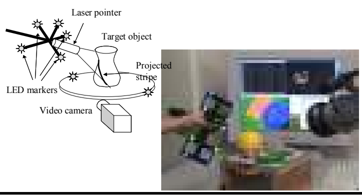


σ=0.00017(m)

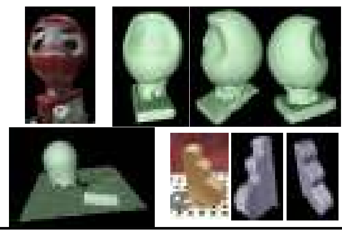
Experiments of shape acquisition Human face



Entire shape acquisition – rotation table



Rotation table results



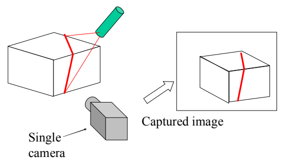
Hand-held scanner with single cam.

- ☺ Simple configuration
- ☺ No precision devices
- ☹ Online-calibration
  - △ Frames or planes are required [david'06]
  - △ LED markers required [kawasaki'03]

Can we eliminate all additional devices ?

Hand-held 3D scanner

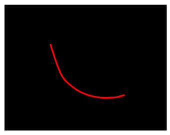
without any additional devices



System configuration

Question 1

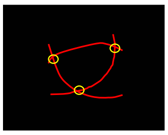
- Can we reconstruct shape from the following image?



NO!

Question 2

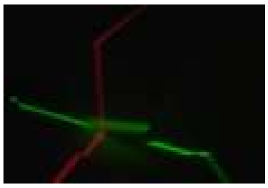
- How about this?



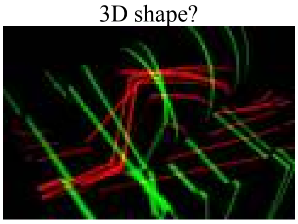
Maybe?

Example of line laser projector (cross pattern by two lasers)

Can you imagine the shape?

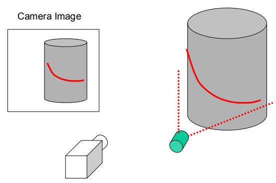


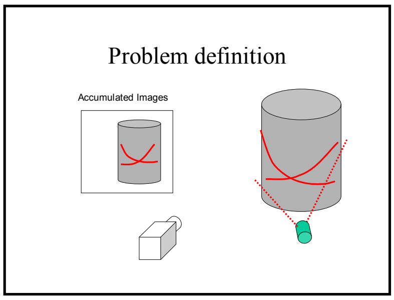
Is it possible to reconstruct the 3D shape?

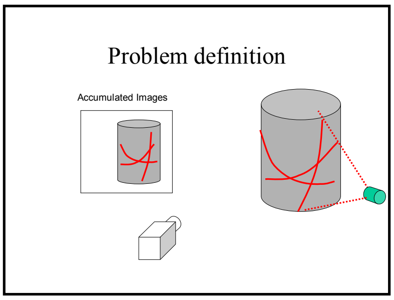


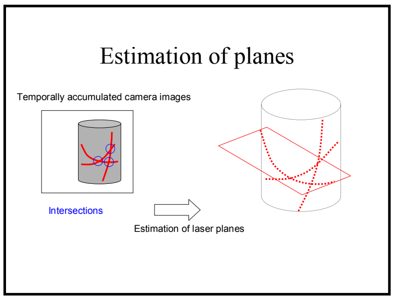
True shape

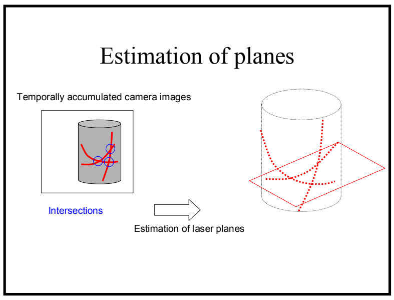
Problem definition

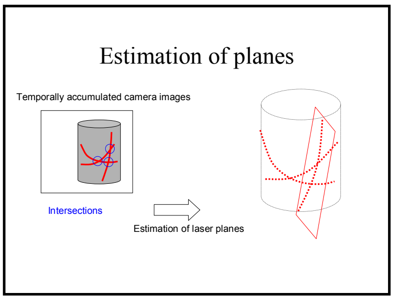


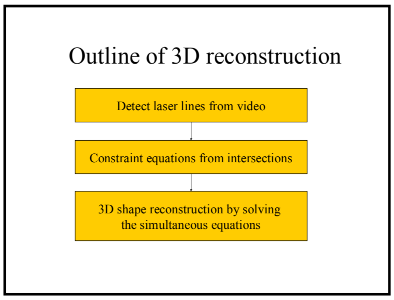

Problem definition
Accumulated Images

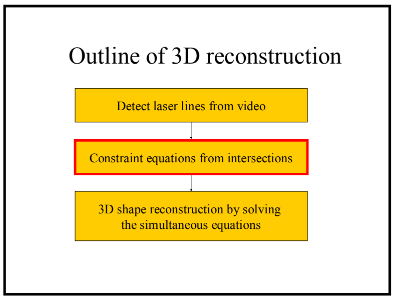

Problem definition
Accumulated Images

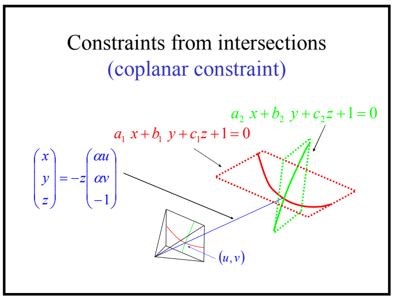

Estimation of planes
Temporally accumulated camera images
Intersections
Estimation of laser planes

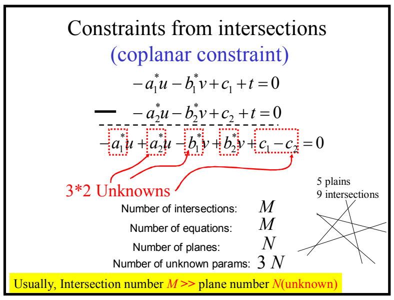

Estimation of planes
Temporally accumulated camera images
Intersections
Estimation of laser planes


Estimation of planes
Temporally accumulated camera images
Intersections
Estimation of laser planes


Outline of 3D reconstruction
Detect laser lines from video
Constraint equations from intersections
3D shape reconstruction by solving the simultaneous equations


Outline of 3D reconstruction
Detect laser lines from video
Constraint equations from intersections
3D shape reconstruction by solving the simultaneous equations


Constraints from intersections
(coplanar constraint)
$a_2 x + b_2 y + c_2 z + 1 = 0$
$a_1 x + b_1 y + c_1 z + 1 = 0$
$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = -z \begin{pmatrix} \alpha u \\ \alpha v \\ -1 \end{pmatrix}$
(u, v)


Constraints from intersections
(coplanar constraint)
$-a_1^* u - b_1^* v + c_1 + t = 0$
$-a_2^* u - b_2^* v + c_2 + t = 0$
$-a_1^* u + a_2^* u - b_1^* v + b_2^* v + c_1 - c_2 = 0$
3*2 Unknowns
5 plains
9 intersections
Number of intersections: M
Number of equations: M
Number of planes: N
Number of unknown params: $3N$
Usually, Intersection number M >> plane number N(unknown)

Matrix form

$$\begin{pmatrix} u_1 & v_1 & 1 & -u_2 & v_2 & -1 & 0 & 0 & 0 & \cdots \\ u_2 & v_2 & 1 & 0 & 0 & 0 & u_2 & v_2 & -1 & \cdots \\ \cdots & & & & & & & & & \end{pmatrix} \begin{pmatrix} a_1 \\ b_1 \\ c_1 \\ a_2 \\ b_2 \\ c_2 \\ a_3 \\ b_3 \\ c_3 \\ \vdots \end{pmatrix} = \mathbf{0}$$

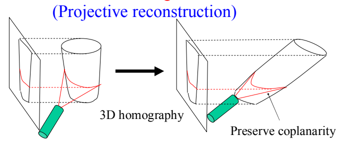
$$\mathbf{Lx} = \mathbf{0}$$

$\mathbf{L}$: $3N*M$ matrix(Intersection num M, Plane num N)

Reconstruction from coplanarity

$$\mathbf{Lx} = \mathbf{0}$$

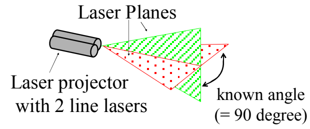
Solution $\mathbf{x}$ has 4 degrees of freedom
(Projective reconstruction)



The 4 DOFs → Found in other research areas.
e.x. Polyhedra analyses in single view reconstruction
Generalized Bas-Relief Ambiguity in photometric stereo

Upgrade to Euclidean solution

- Metric constraints from laser planes

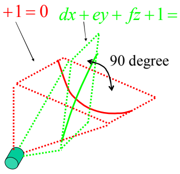


Formulation of metric constraints

Constraints from orthogonality

$ax + by + cz + 1 = 0$ $\quad dx + ey + fz + 1 = 0$

$$ad + be + cf = 0$$



Outline of 3D reconstruction

Detect laser lines from video

↓

Projective reconstruction from intersections

↓ Eliminate 4dofs

Euclid upgrade using metric constraints

Sparse data sets

Outline of dense 3D reconstruction

Detect laser lines from video

↓

Projective reconstruction from intersections

↓

Euclid upgrade using metric constraints

↓

Dense shape reconstruction

Experiments

- Simulation data
- Real data

Real data 2

- Cross line laser

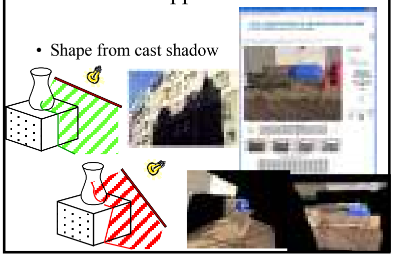
Real data 2 – result

Proposed method

- Only require a line laser and a single camera
  - General solution for “Shape from Coplanarity”
  - Any other applications?

Other applications

- Shape from cast shadow



Shape from cast shadow

Other applications

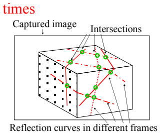
- Single view reconstruction

Other applications

- Single view reconstruction

Summary of the proposed method

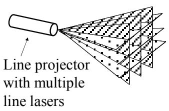
- Temporal accumulation
  - ○ Self-calibration for hand held 3D scanner
  - △ take times



→ Can we make enough intersections at one time?

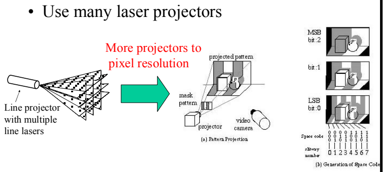
Self-calibration for 3D scanner

- Use many laser projectors



Self-calibration for Coded Structured light

- Use many laser projectors

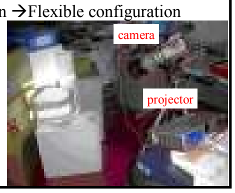


Coded structured light

Self-calibration for Coded Structured light

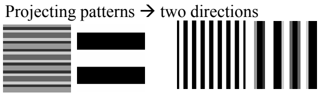
[H.Kawasaki and R.Furukawa, 3DIM2005]

- No mechanical devices
- No pre-calibration →Flexible configuration
- High accuracy
- Dense sampling

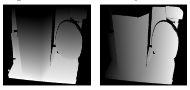


Structured light example

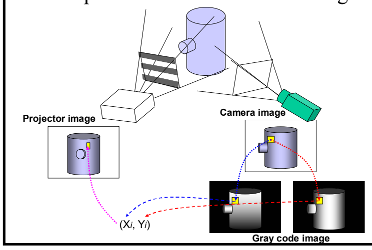
- Projecting patterns → two directions



- Acquired coded images

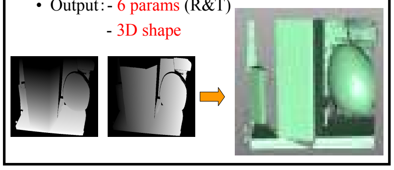


Correspondences from decoded images



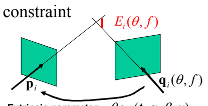
Self-calibration for Coded Structured light
- problem definition -

- Input : - camera params (focal length, etc.)
  - two index images
- Output: - 6 params (R&T)
  - 3D shape



Non-linear optimization

- Epipolar constraint



Extrinsic parameter $\theta := (\mathbf{t}, \alpha, \beta, \gamma)$

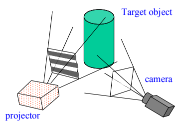
$\mathbf{t}$:Translation α, β, γ :Rotation(Euler angles)

$$E_i(\theta, f) := \left|\mathbf{t} \cdot N(\mathbf{p}_i \times \mathbf{q}_i(\alpha, \beta, \gamma, f))\right|$$

Using $F(\theta, f) := \sum_i \{E_i(\theta, f)\}^2$ as a minimizing function.

Feature of the method

- Single camera and a projector
- Freely movable (both camera and projector)
- High speed scan
- High accuracy



Extended techniques

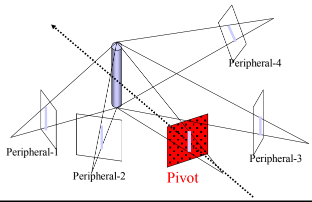
- Wide range reconstruction by pivot scanning
- Simultaneous reconstruction method

Extended techniques

- Wide range reconstruction by pivot scanning
- Simultaneous reconstruction method

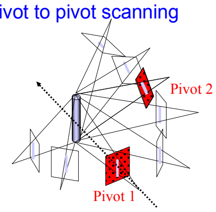
Pivot scanning

Fix pivot device and move peripheral device arbitrarily



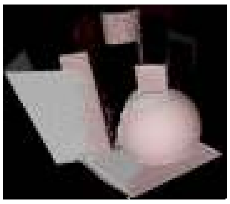
More wide view scanning

Pivot to pivot scanning



Result – pivot to pivot

- No alignment algorithm applied



1+2 result

Demo movie

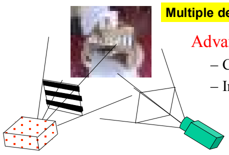
- Pivot scan

Extended techniques

- Wide range reconstruction by pivot scanning
- Simultaneous reconstruction method

Simultaneous reconstruction

- Capture multiple scenes
- 3D reconstruction simultaneously



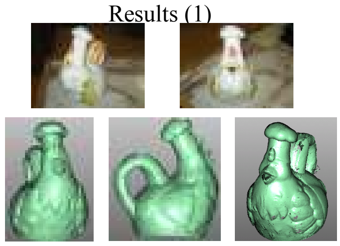
Multiple depths for single pixel

Advantage
  - Consistent scaling
  - Improving result
    - Redundant input

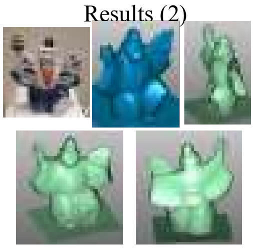
Demo movie

- Simultaneous scan

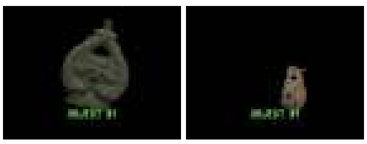
Results (1)



Results (2)

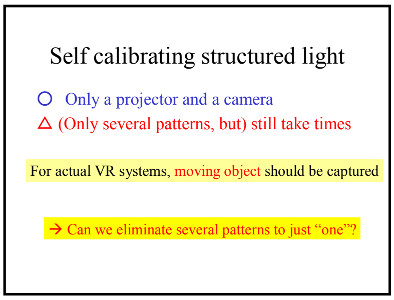


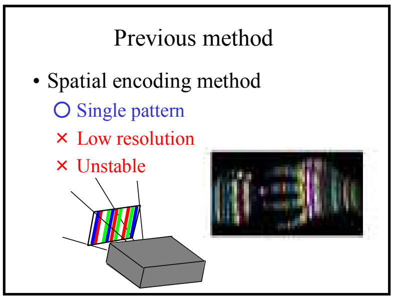
Final results

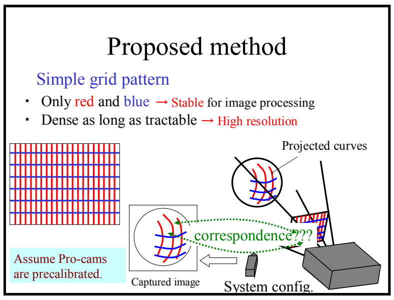


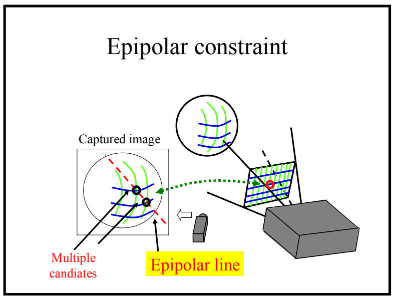
With •Fast mesh integration [Furukawa and Kawasaki 3DIM '05]
•seamless texture [Inose, kawasaki *et.al.* '06 '07]

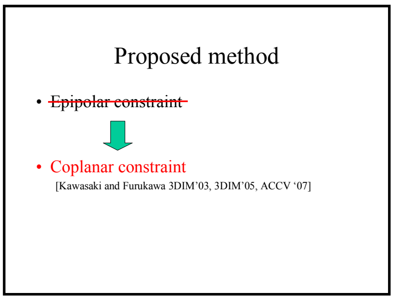
Live Demo

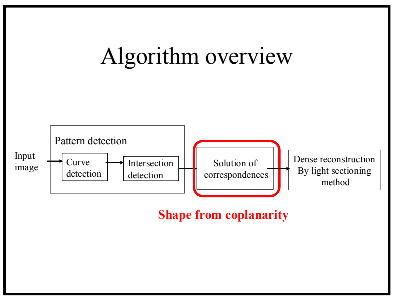

Self calibrating structured light
○ Only a projector and a camera
△ (Only several patterns, but) still take times
For actual VR systems, moving object should be captured
→ Can we eliminate several patterns to just "one"?

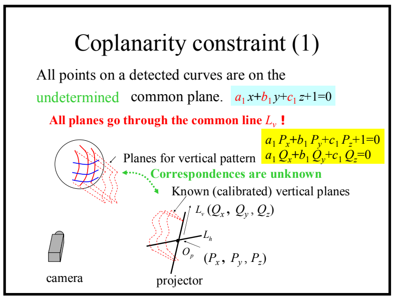

Previous method
• Spatial encoding method
○ Single pattern
× Low resolution
× Unstable

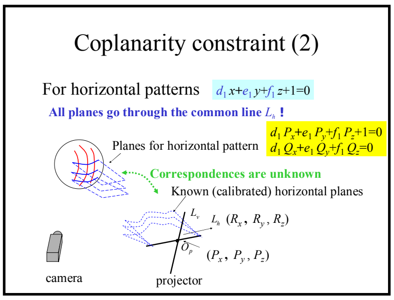

Proposed method
Simple grid pattern
• Only red and blue → Stable for image processing
• Dense as long as tractable → High resolution
Projected curves
correspondence???
Assume Pro-cams are precalibrated.
Captured image
System config.

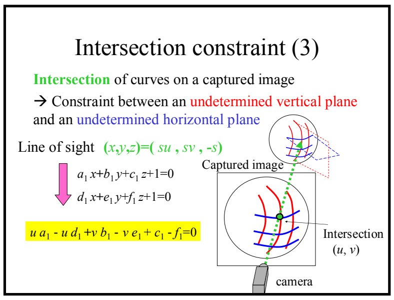

Epipolar constraint
Captured image
Multiple candiates
Epipolar line

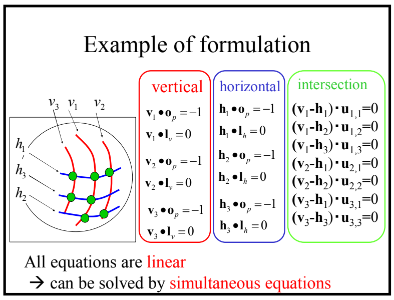

Proposed method
• Epipolar constraint
• Coplanar constraint
[Kawasaki and Furukawa 3DIM'03, 3DIM'05, ACCV '07]

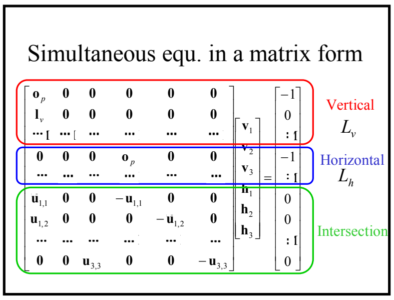

Algorithm overview
Input image
Pattern detection
Curve detection
Intersection detection
Solution of correspondences
Dense reconstruction By light sectioning method
Shape from coplanarity

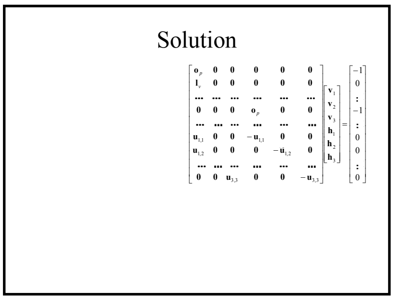

Coplanarity constraint (1)
All points on a detected curves are on the undetermined common plane. $a_1 x+b_1 y+c_1 z+1=0$
All planes go through the common line L_v !
$a_1 P_x+b_1 P_y+c_1 P_z+1=0$
$a_1 Q_x+b_1 Q_y+c_1 Q_z=0$
Planes for vertical pattern
Correspondences are unknown
Known (calibrated) vertical planes
L_v (Q_x, Q_y, Q_z)
L_h
O_p
(P_x, P_y, P_z)
camera
projector

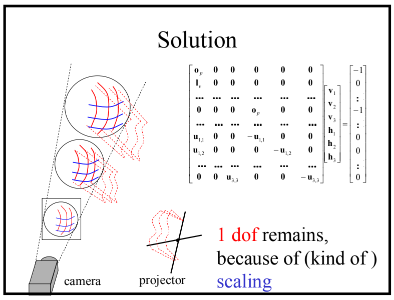

Coplanarity constraint (2)
For horizontal patterns $d_1 x+e_1 y+f_1 z+1=0$
All planes go through the common line L_h !
$d_1 P_x+e_1 P_y+f_1 P_z+1=0$
$d_1 Q_x+e_1 Q_y+f_1 Q_z=0$
Planes for horizontal pattern
Correspondences are unknown
Known (calibrated) horizontal planes
L_v
L_h (R_x, R_y, R_z)
O_p
(P_x, P_y, P_z)
camera
projector

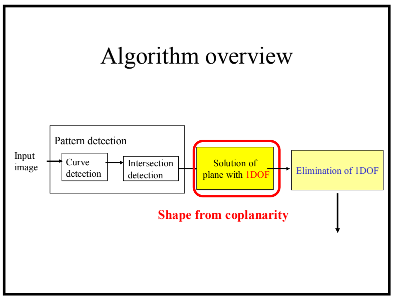

Intersection constraint (3)
Intersection of curves on a captured image
→ Constraint between an undetermined vertical plane and an undetermined horizontal plane
Line of sight $(x,y,z)=(su, sv, -s)$
$a_1 x+b_1 y+c_1 z+1=0$
$d_1 x+e_1 y+f_1 z+1=0$
$u a_1 - u d_1 + v b_1 - v e_1 + c_1 - f_1=0$
Captured image
Intersection (u, v)
camera

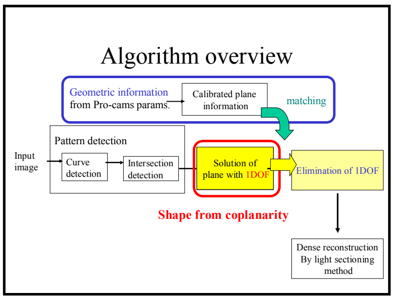

Example of formulation
vertical
horizontal
intersection
All equations are linear
→ can be solved by simultaneous equations

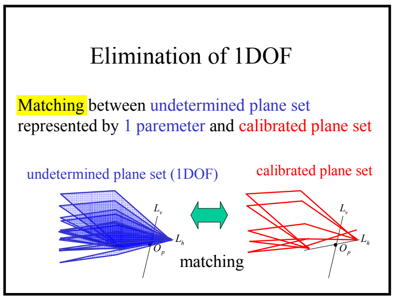

Simultaneous equ. in a matrix form
Vertical
Horizontal
Intersection

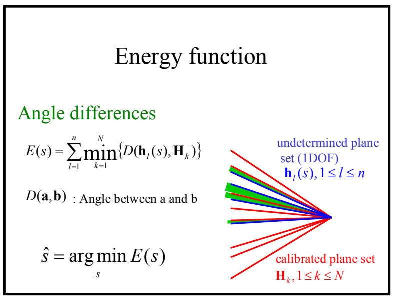

Solution

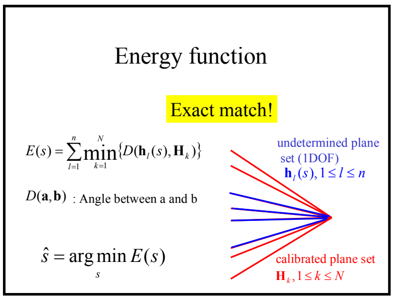

Solution
camera
projector
1 dof remains,
because of (kind of)
scaling


Algorithm overview
Input image
Pattern detection
Curve detection
Intersection detection
Solution of plane with 1DOF
Elimination of 1DOF
Shape from coplanarity


Algorithm overview
Geometric information from Pro-cams params.
Calibrated plane information
matching
Input image
Pattern detection
Curve detection
Intersection detection
Solution of plane with 1DOF
Elimination of 1DOF
Shape from coplanarity
Dense reconstruction By light sectioning method

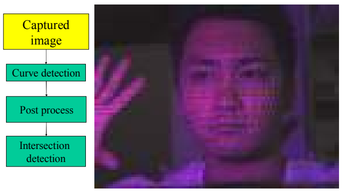

Elimination of 1DOF
Matching between undetermined plane set
represented by 1 paremeter and calibrated plane set
undetermined plane set (1DOF)
calibrated plane set
matching


Energy function
Angle differences
D(a,b) : Angle between a and b
undetermined plane set (1DOF)
calibrated plane set


Energy function
Exact match!
D(a,b) : Angle between a and b
undetermined plane set (1DOF)
calibrated plane set

Experiments

Reconstruction process



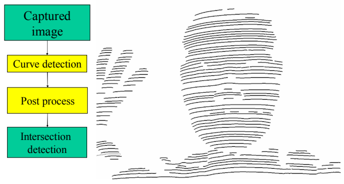
Reconstruction process

Reconstruction process

Vertical direction

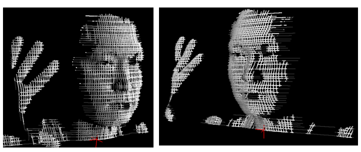
Reconstruction process

Horizontal direction

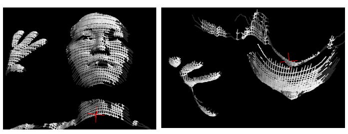


Reconstruction process

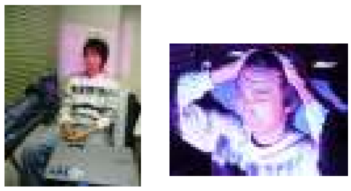
Reconstruction results



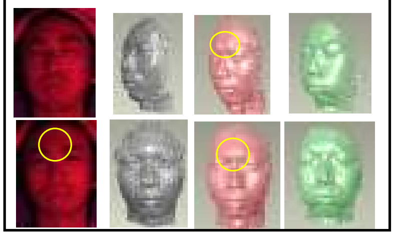
Reconstruction results



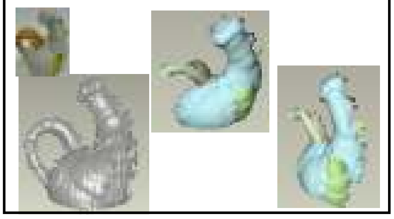
Other example



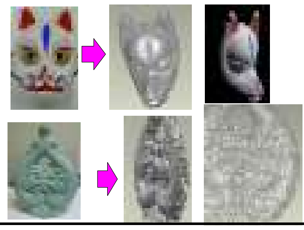
Results



Textured object 1

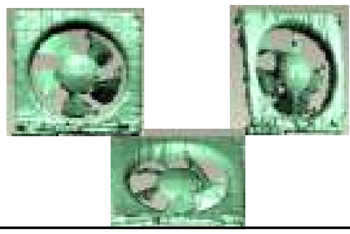


Textured object 2



High speed scan

High speed scan –result –



Live demo

What's next?

- Infrared projector for VR/MR systems
- Recognition human behavior by 3D video
- Ultra high-speed 3D scan

Thanks.

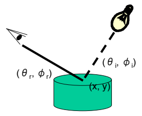
- Any questions or comments?
- Web sites:
  – http://www.cgv.ics.saitama-u.ac.jp
- e-mail:
  – kawasaki@cgv.ics.saitama-u.ac.jp

Research and project matrix

	Acquisition		Modeling		Rendering
	Geometry	Photometry	Geometry	Photometry	
City modeling (ITS)	△	◎		△	◎
Active 3D scanner	◎		◎		
Texture analysis		◎		◎	◎
Digital archiving	◎	△	◎	△	

Texture acquisition

- Surface reflectance of the object
  - Light dependency L(θ , ϕ)
    - Diffuse
    - Specularity
  - View dependency V(θ , ϕ)
    - Velvet
    - Fur



4D function (L (θ , ϕ),V (θ , ϕ))

→ Bidirectional Reflectance Distribution Function (BRDF)

Texture acquisition device

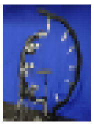
- Single camera (1D)
- Light stage v.1-v.3 (2D)
  - [Devebec SIGGRAPH01-
- Polynomial texture map (4D)
  - [Malzender SIGGRAPH01]
- University of Bonn[EG04] (4D)
- CUReT [Columbia Univ.] (4D)

Assume plane (2D) geometry

Surface Reflectance Sampler

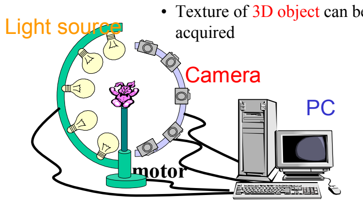
[Aritaki, Kawasaki et.al. ICIP01]

- Full automatic 4D data capturing machine
  - Japan Patent Office #3601031



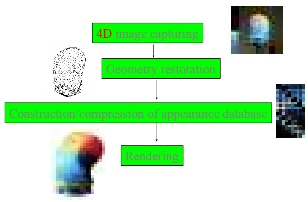
Surface Reflectance Sampler

- Texture of 3D object can be acquired

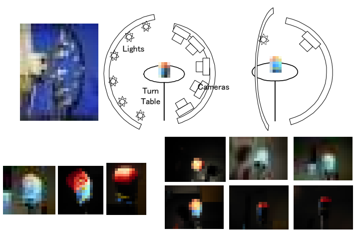


Rendering system

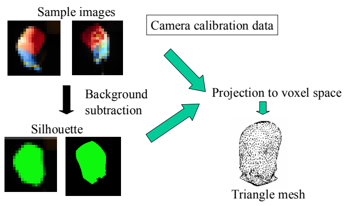
[Furukawa, Kawasaki et.al. EGSR03]



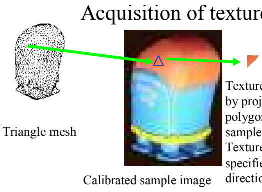
Capturing of real image samples



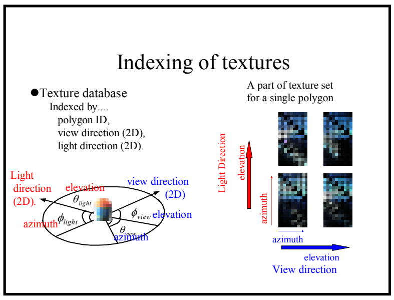
Restoration of geometry(Voxel Curving)

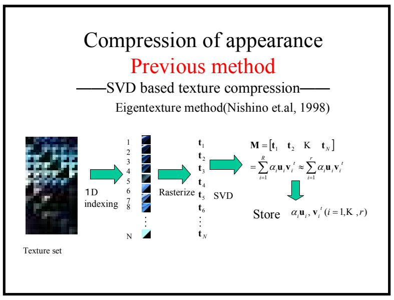


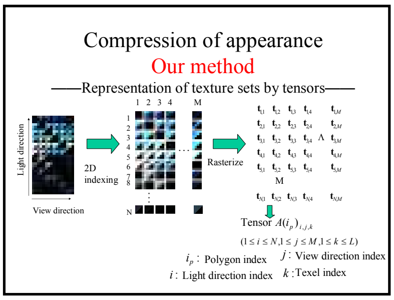
Acquisition of textures

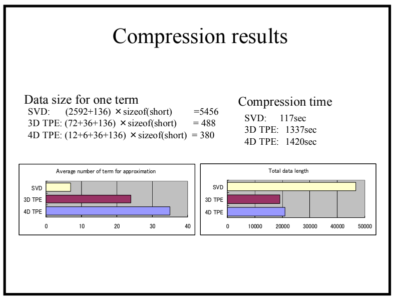


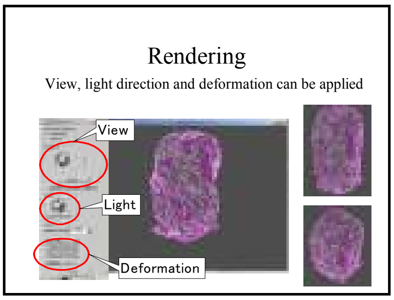
Textures are obtained by projecting each polygon into calibrated sample images. Texture can be specified by view/light directions.

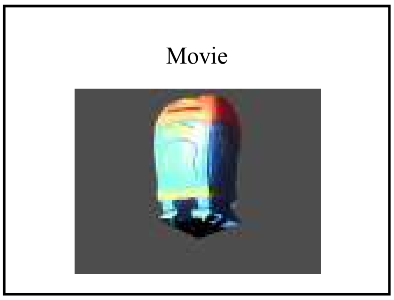

Indexing of textures
●Texture database
Indexed by....
polygon ID,
view direction (2D),
light direction (2D).
A part of texture set
for a single polygon
Light
direction
(2D)
elevation
azimuth
view direction
(2D)
elevation
azimuth
Light Direction
View direction

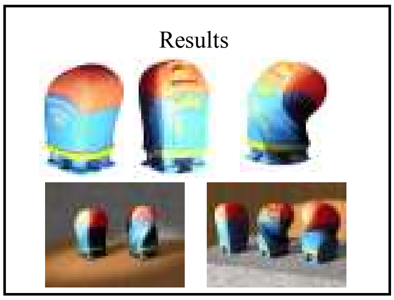

Compression of appearance
Previous method
SVD based texture compression
Eigentexture method(Nishino et.al, 1998)
1D
indexing
Rasterize
SVD
Store
Texture set

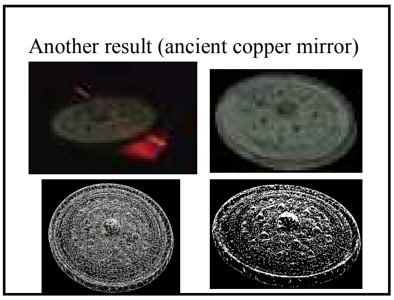

Compression of appearance
Our method
Representation of texture sets by tensors
Light direction
View direction
2D
indexing
Rasterize
Tensor
i_p : Polygon index
j : View direction index
i : Light direction index
k : Texel index


Compression results
Data size for one term
SVD: (2592+136) ×sizeof(short) =5456
3D TPE: (72+36+136) ×sizeof(short) = 488
4D TPE: (12+6+36+136) ×sizeof(short) = 380
Compression time
SVD: 117sec
3D TPE: 1337sec
4D TPE: 1420sec
SVD
3D TPE
4D TPE


Rendering
View, light direction and deformation can be applied
View
Light
Deformation


Movie


Results


Another result (ancient copper mirror)

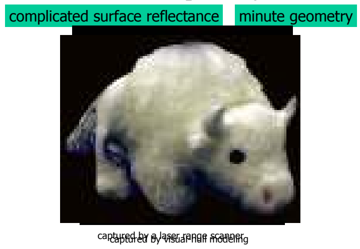
Real-time rendering results

- Using pixel and vertex shader

Discussion

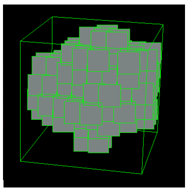
- ○ BTF – efficiently capture, compress and render 3D objects
- × **Intricate shaped** object is still difficult to render

Intricate Shaped Objects



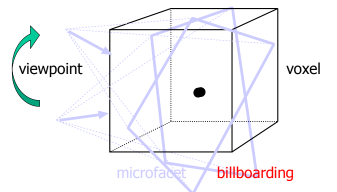
Solution → MBR + IBR

1. Rough surface model
2. Resampling
3. Multiresolution
4. A set of microfacets

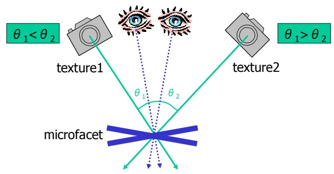


Put view dependent texture on microfacets

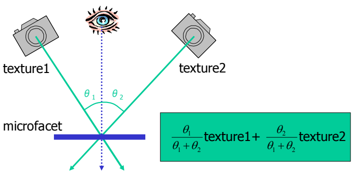
Billboarding Microfacets



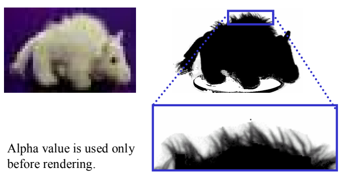
Selection of Texture Images



Blending of Texture Images



Alpha Estimation



- Alpha value is used only before rendering.

Final Synthesis

Previous method　　Preposed method

Movie

Thanks.

- Any questions or comments?
- Web sites:
  - http://www.cgv.ics.saitama-u.ac.jp
- e-mail:
  - kawasaki@cgv.ics.saitama-u.ac.jp