Computer Vision

No. 1
What is the Computer Vision?

Instructor

- Katsushi Ikeuchi
- Takeshi Oishi
- Shintaro Ono
- Bo Zheng
- Masataka Kagesawa
- Yasuhide Okamoto

Pointers

- 03-5452-6242
- cvl-sec@cvl.iis.u-tokyo.ac.jp
- 東京都目黒区駒場４－６－１
  東大生研　Ee405

Pre-requisite

- This course mainly reads recent research papers on computer vision
- If you have not taken one of the followings:
  - Grad: 学際情報学基礎IV　（学際情報学府）
  - UnGrad: コンピュータビジョン（理学部情報科学）

  I strongly discourage you to take this course

Evaluation

- attendance 50%
- Report 50%

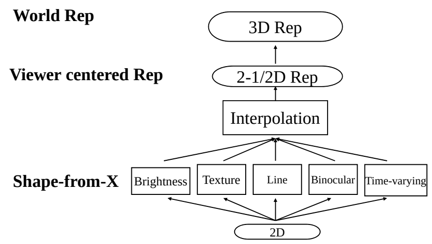
Class document

- Report submission:
  cvl-class-2014w@cvl.iis.u-tokyo.ac.jp
- Hand-out
  http://www.cvl.iis.u-tokyo.ac.jp/class/grad

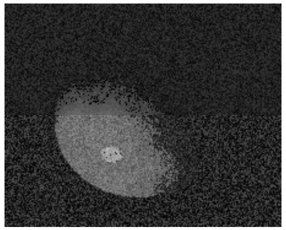
Computer Vision (CV)

- To make a computer to recognize the 3D world as we do
- To generate 3D representations from 2D images

Marr's Paradigm

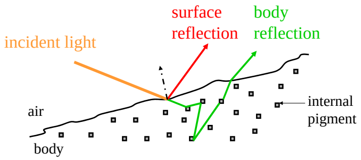


Shape-from-Shading



Surface and body reflection

- Surface reflection and body reflection



surface reflection=gloss,highlights very directional(specular)
body reflection =object color all direction(diffuse)
plastic, paint have both
metal has only surface reflection

Model for body reflection

Diffuse---scatters in all directions
common approximation:
- equal in all directions
- "lambertian"Lambertian's cosine law
- "perfectly diffuse reflector"

reflectance=constant * geometric factor
- f(i,e,g) = Kb * cos i

why cos i ?

angle of incidence affects "density" of illumination.(irradiance)

irradiance=light/area
light=1
area=1/cos i
irradiance = cos i

Calculating a reflection map (Lambertian)

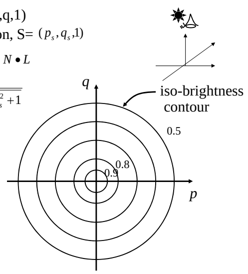
- for each(p,q), N=(p,q,1)
- light source direction, S= $(p_s, q_s, 1)$

$$R(p,q,p_s,q_s) = \cos i = N \bullet L$$

$$= \frac{p \cdot p_s + q \cdot q_s + 1}{\sqrt{p^2+q^2+1}\sqrt{p_s^2+q_s^2+1}}$$

$$(p_s, q_s) = (0,0)$$

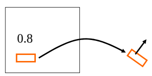
$$R(p,q) = \frac{1}{\sqrt{p^2+q^2+1}}$$



Definition of SFS

Relationship between surface orientation and brightness
⇔ obtain surface orientation from brightness

Brightness
（1 dim）



Orientation
（2 dim）

=> 2 unknown under one equation!

Need extra constraint

Smoothness constraints:
Neighboring points have similar characteristics

$$\left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2 + \left(\frac{\partial g}{\partial x}\right)^2 + \left(\frac{\partial g}{\partial y}\right)^2 = 0$$

f and *g* -- surface orientations

Variational Approach

- Set up a minimization problem.

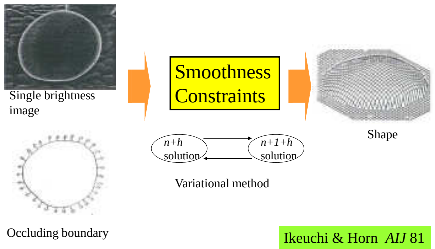
$$E = \iint (E(x,y) - R(f(x,y), g(x,y)))^2 + \lambda\left\{\left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2 + \left(\frac{\partial g}{\partial x}\right)^2 + \left(\frac{\partial g}{\partial y}\right)^2\right\} dxdy \rightarrow \min$$

- Get iterative formula using the calculus of variation

$$f^{n+1}(x,y) = \frac{1}{4}\left\{f^n(x-1,y)...\right\} + \lambda\{...\}$$

$$g^{n+1}(x,y) = \frac{1}{4}\left\{g^n(x-1,y)...\right\} + \lambda\{...\}$$

Shape-from-shading with smoothness constraints

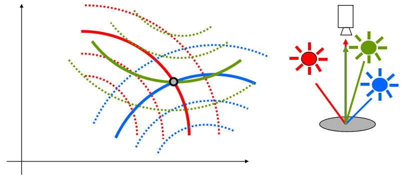


Shape-from-shading

For robot vision

Need high-speed algorith
Photometric stereo

Photometric stereo



Analytic solutions

$$n=(n_x,n_y,n_z)=\left(\frac{p}{\sqrt{p^2+q^2+1}},\frac{q}{\sqrt{p^2+q^2+1}},\frac{1}{\sqrt{p^2+q^2+1}}\right)$$

$$S_1=(S_{1x},S_{1y},S_{1z})=\left(\frac{p_{1s}}{\sqrt{p_{1s}^2+q_{1s}^2+1}},\frac{q_{1s}}{\sqrt{p_{1s}^2+q_{1s}^2+1}},\frac{1}{\sqrt{p_{1s}^2+q_{1s}^2+1}}\right)$$

$$S_2=(S_{2x},S_{2y},S_{2z})$$

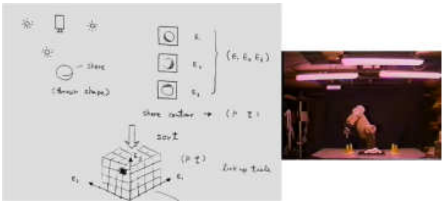
$$S_3=(S_{3x},S_{3y},S_{3z})$$

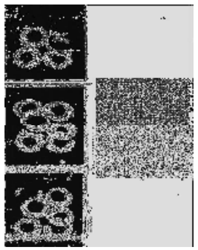
$$\left\{\begin{array}{l}E_1=\vec{S}_1\bullet\vec{n}\\ E_2=\vec{S}_2\bullet\vec{n}\Rightarrow\vec{E}=A\bullet\vec{n}\\ E_3=\vec{S}_3\bullet\vec{n}\end{array}\right.\qquad A=\begin{pmatrix}S_{1x}&S_{1y}&S_{1z}\\S_{2x}&S_{2y}&S_{2z}\\S_{3x}&S_{3y}&S_{3z}\end{pmatrix}$$

$$A^{-1}\bullet\vec{E}=A^{-1}\bullet A\bullet\vec{n}$$

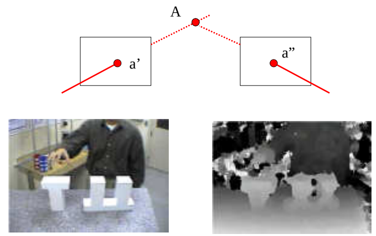
$$\vec{n}=A^{-1}\bullet\vec{E}$$

Lookup table method

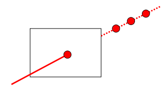




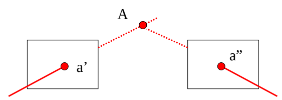
Binocular Stereo



Binocular stereo

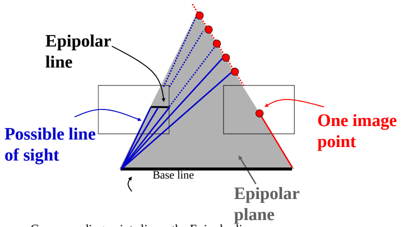


single image is ambiguous



another image taken from a different direction
gives the unique 3D point

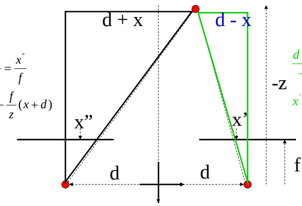
Epipolar line constraints



Corresponding points lie on the Epipolar lines

→ *Epipolar line constraints*

Basic binocular stereo equation



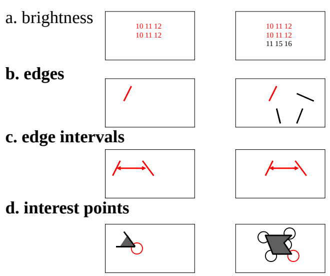
$$\frac{d+x}{-z}=\frac{x''}{f}\qquad x''=-\frac{f}{z}(x+d)$$

$$\frac{d-x}{-z}=\frac{-x'}{f}\qquad x'=-\frac{f}{z}(x-d)$$

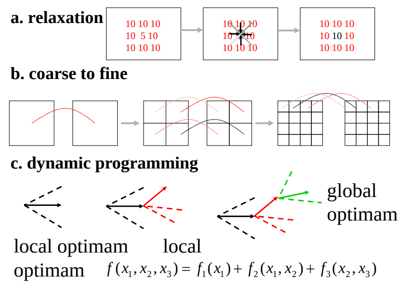
$$x''-x'=-\frac{f}{z}(x+d-x+d)=-\frac{f}{z}2d$$

z=-2df/(x"-x')
x"-x': disparity
2d : base line length

Features for matching

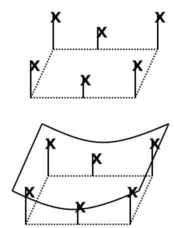


Strategies for matching



Interpolation

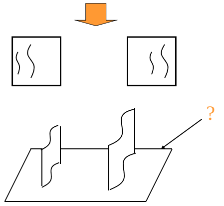
Surface Interpolation



Purpose of surface interpolation 1

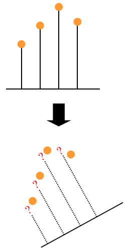
(Computer Vision)

Binocular Stereo



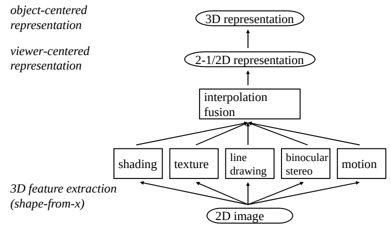
Purpose of surface interpolation 2

(Graphics)



Purpose of surface interpolation 3

(Psychology and computer vision)

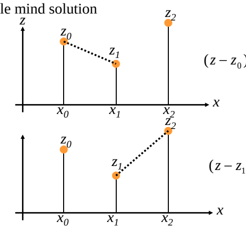


Purpose of surface interpolation 4

(Psychology)

Method 1

simple mind solution



$$(z - z_0) = \frac{z_1 - z_0}{x_1 - x_0}(x - x_0)$$

$$(z - z_1) = \frac{z_2 - z_1}{x_2 - x_1}(x - x_1)$$

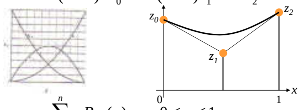
Problem:

orientation discontinuity at x_1

Method 2

Bezier curve

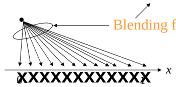
3 points $z = (1-x)^2 z_0 + 2x(1-x) z_1 + x^2 z_2$



$$z = \sum_{i=0}^{n} z_i B_{i,n}(x) \qquad 0 \le x \le 1$$

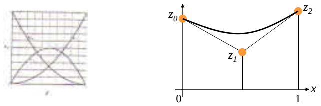
$$B_{i,n}(x) = {}_nC_i u^i (1-u)^{n-i}$$

Blending function

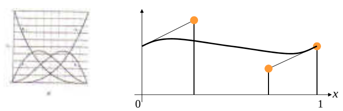


Example of Bezier curve

3 points $z = (1-x)^2 z_0 + 2x(1-x) z_1 + x^2 z_2$

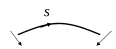


4 points $z = (1-x)^3 z_0 + 3x(1-x)^2 z_1 + 3x^2(1-x) z_2 + x^3 z_3$



Least energy curve (Snake)

- To find a curve which minimizes
  1) internal energy --- bending

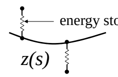


$$E_{internal} = \int_0^S bending-energy(s)ds$$

2) external energy --- intensity

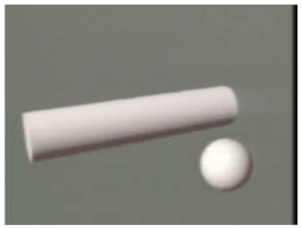
$$E_{external} = \int_0^S I(s)ds$$

3) constraint energy --- boundary

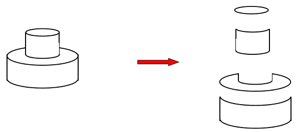


$$E_{constraint} = k(z(s) - z_0)^2$$

Falling Rug



Object Representation

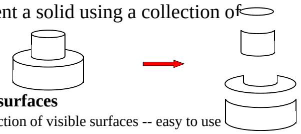


Classification of representation

- **Surface based representation**
  - represented as a collection of surfaces
    - » curvature primal sketch
    - » extended gaussian image
    - » aspect graph
    - » b-rep (winged-edge)
    - » well-tessellated surface
- **Function based representation**
  - represented as a function and its parameters
    - » generalized cylinder
    - » superquadric
    - » symmetry seeking
    - » spherical attribute image
- **Volumetric based representation**
  - represented as a collection of primitive solids
    - » constructive solid geometry
    - » occupancy graph
    - » oct-tree

Surface based representation

- Represent a solid using a collection of surfaces

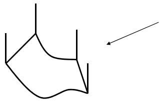


- **which surfaces**
  - * collection of visible surfaces -- easy to use
  - * collection of all surfaces -- easy to maintain
- **what is a reasonable definition**
  - * orientation discontinuity -- if range data is available, one of the most common definitions
  - * Gaussian curvature and mean curvature represent a solid using a collection of surfaces
  - * color (brightness) -- view dependent only

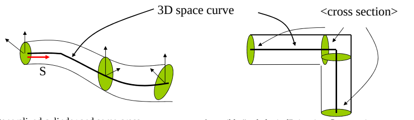
Function based representation

1. Classification of representation
2. Superquadric
3. Generalized cylinder
4. Symmetry seeking
5. Spherical attribute image

Generalized Cylinder



"the sweep volume" of a two-dimensional set moved along a three-space curve <axis>

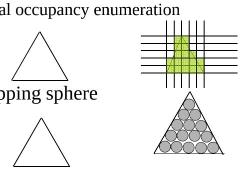


A generalized cylinder and some cross-sectional coordinate systems

A possibly "pathological" situation. Cross sections may be simply described as circles centered on the axis, but then their intersection makes volume calculations (for instance) less straightforward.

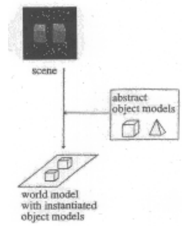
Volume based representation

- Represent a solid as a collection of primitive solids
  - spatial occupancy enumeration
- overlapping sphere



- requires many spheres or voxels to represent to represent a relatively simple smooth solid

Object Recognition

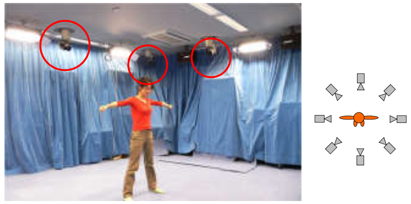


VISION FOR NAVIGATION

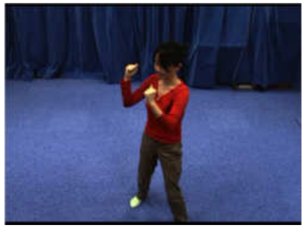
Autonomous Driving

VISION FOR MOTION MODELING

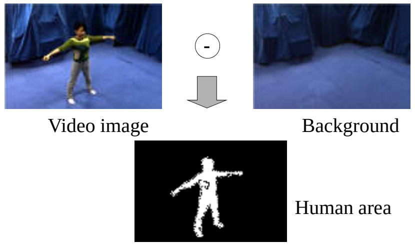
One way of Observation



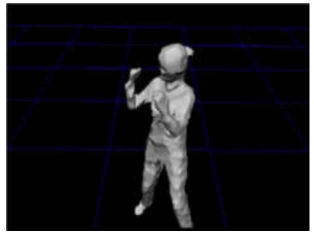
One of eight sequences



Background subtraction



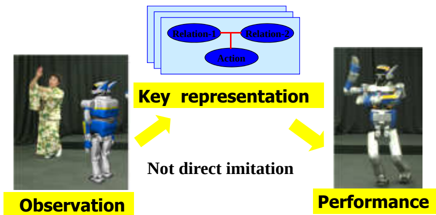
Obtained 3D Sequence



VISION FOR MOTION LEARNING

Learning from observation

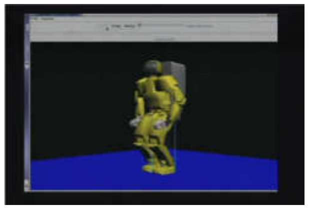
Top-down approach



Joint angles obtained

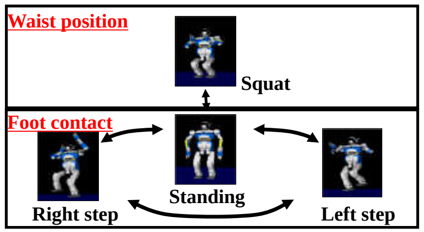
Theoretically, a robot can imitate the same dance???

Issues



AIST dynamic simulator

State transitions (lower body)



Upper body states

Costarring with the dance teacher

Nakaoka 2006, Shiratori 2007

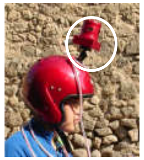
With cooperation of AIST, Kawata

VISION FOR GRAPHICS

Pompeii Walker

Pompeii Walker

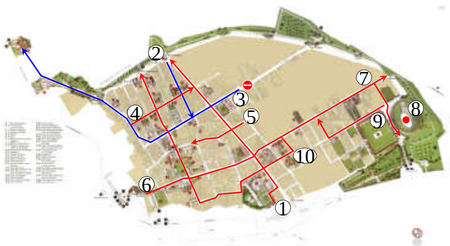
- Data collection by a probe person with a omnidirectional camera



Issues

- **Spatial Structure**
- **Vibration**

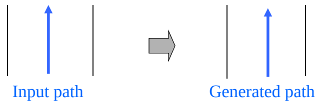
Video + GPS



Stabilization

Issue in Pompeii Walker

- Can display only images along the acquisition path



Can we?

- Generated paths different from the input one?

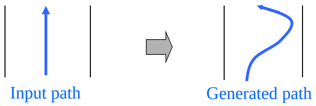
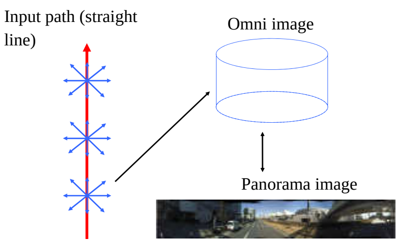
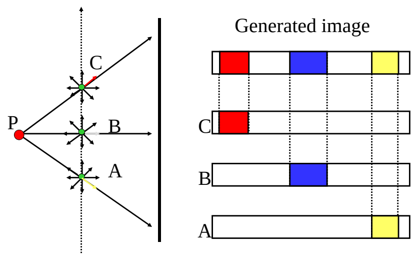


Image-based rendering

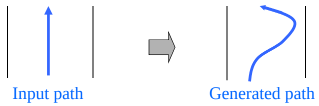


Ray space

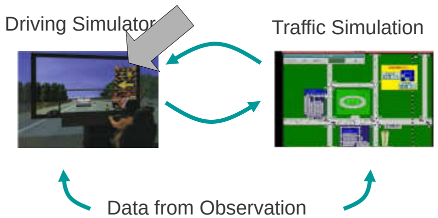


Demonstration

- Generated paths different from the input one

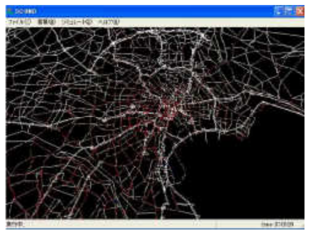


Mixed Reality Driving Environment



Traffic Simulation

(Kuwahara group)



Driving Simulator

(Suda Group)

VISION FOR VIRTUAL REALITY

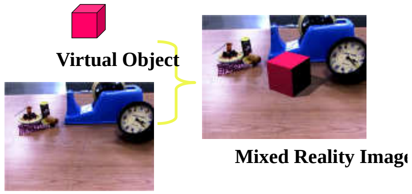
Mixed Reality Display

- Real cite
- From ancient time
- Fusion current with ancient

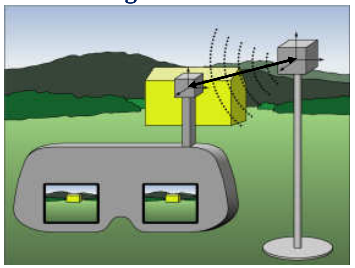
Issues in Mixed Reality

- Geometric consistency
- Photometric consistency

Geometric Consistency

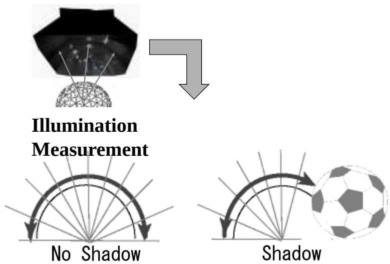


Using a Hardware



Photometric Consistency

Calculation of Shadow



Asukadera, Tyumon Gate

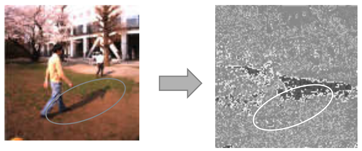
Issue： Occlusion

Occlusion of sight-seers
-Segmentation
-Distance estimation

Occlusion of shadows
-Segmentation
-Recasting

Segmentation

- Black-body assumption
- Illumination invariant

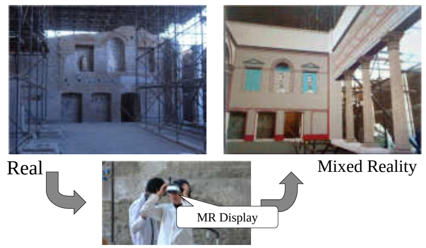


Segmentation based using invariant (no shadow)

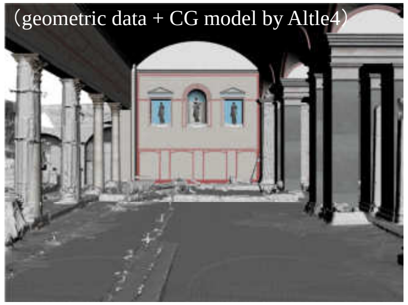
Sonnma Vesuviana

Oct 2008

MR in the Aoyagi site



Geometric model



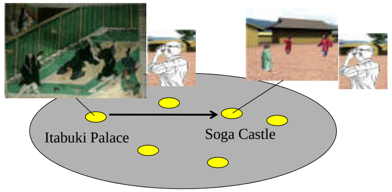
Foro Romano

Dec 2008

Proposed visit

Synchronic display

- Watching events on the site



Tram in Heijo Palace

With ASUKA LAB

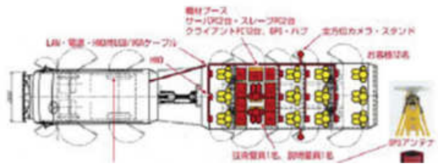
Issue： Location and Direction

- Location： GPS

- Direction： E. Compass＋Gyro

103

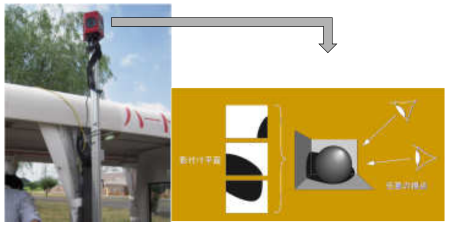
Issue： Multiple Viewers



- Image distribution
- 12 viewers

Multiple Viewers

Issues：Illumination Estimation



Illumination measurement **Real time shadow plane**

Extraction of characters from move archive

[One scene from NHK drama]

Display Video

Anyway, what is Computer Vision?

- Vision is … an information processing task that constructs efficient symbolic descriptions of the world from images. (Marr)
- Vision is … inverse graphics.
- Vision looks easy, but is difficult.
  Vision is … difficult, but is fun. (Kanade)
- Vision is an engineering science to create an alternative of human visual systems on computers（Ikeuchi）

References

- Journals
  - Inter. J. Computer Vision
  - IEEE Trans. Pattern Analysis and Machine Intelligence
  - IEICE D-2
  - IPSJ Trans CVIM
- International conferences
  - Inter. Conf. Computer Vision (ICCV)
  - Computer Vision and Pattern Recognition (CVPR)
  - Asian Conf. Computer Vision (ACCV)
- Special interest groups
  - IPSJ CVIM
  - IEICE PRMU

Schedule

#1	Oct 8th	Introduction	Ikeuchi
#2	Oct 22nd	Object Representation (1)	Zheng
#3	Oct 29th	Object Representation (2)	Zheng
#4	Nov 5th	Time-varying Image Processing (1)	Ono
-----	Nov 12th	No Class (III D Col)	-----------
#5	Nov 19th	Time-varying Image Processing (2)	Ono
#6	Nov 26th	3D Data Vizualization (1)	Okamoto
#7	Dec 3rd	3D Data Vizualization (2)	Okamoto
-----	Dec 10th	No Class (3DV Conference)	----------
#8	Dec 17th	3D Data Processing (1)	Oishi
#9	Jan 7th	3D Data Processing (2)	Oishi
#10	Jan 14th	Patch-based Object Recognition (1)	Kagesawa
#11	Jan 21st	Patch-based Object Recognition (2)	Kagesawa

Sub Areas

- Shape-from-X
  - Shading/binocular Oct 8th Ikeuchi
  - Time varying images Nov 5th & Nov 19th Ono
  - 3D Data Processing Dec 17th & Jan 7th Oishi
- Interpretation
  - Object representation Oct 22nd & Oct 29th Zheng
  - Object recognition Jan 14th & Jan 21st Kagesawa
  - 3D visualization Nov 26th & Dec 3rd Okamoto