# Kanatani's Statistical Optimization for Geometric Computation

# Chapter 11: 3-D Motion Analysis

Olaf Booij

Intelligent Systems Lab Amsterdam
University of Amsterdam, The Netherlands

Kanatani reading club 2-10-2009

# Outline

# 11.0 Meta

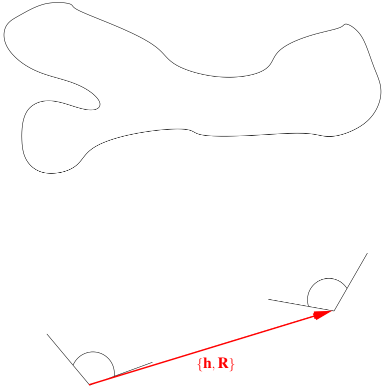
## Typo's

- p332 first line under 11.33
  *data* $x_a$ *and* $x_a$ ->
  *data* $x_a$ *and* $x'_a$
- p335 line 5
  $\mathcal{M}$, $\mathcal{N}^{(1)}$, and $\mathcal{N}^{(1)}$ ->
  $\mathcal{M}$, $\mathcal{N}^{(1)}$, and $\mathcal{N}^{(2)}$
- Throughout the chapter:
  3-D Motion Analysis ->
  3-D Motion and Scene Reconstruction Analysis ->

# 11.1 General Theory

## The problem

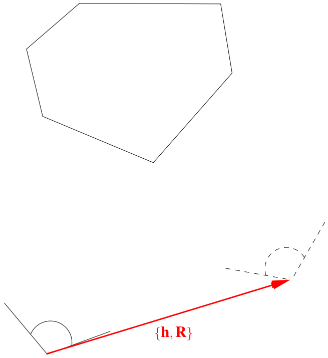
- two camera seeing a non-rigid object
- moving camera seeing a rigid object
- stationary camera seeing a moving rigid object
- moving camera seeing a moving rigid object ...



# 11.1 General Theory

## The problem

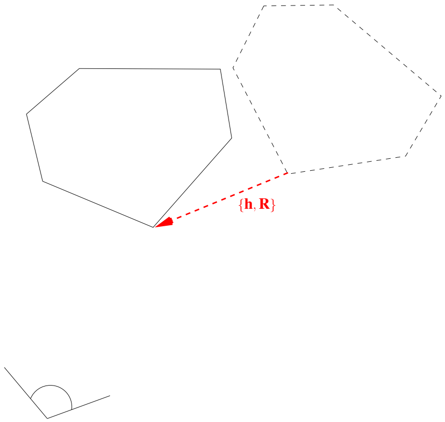
- two camera seeing a non-rigid object
- moving camera seeing a rigid object
- stationary camera seeing a moving rigid object
- moving camera seeing a moving rigid object . . .



# 11.1 General Theory

## The problem

- two camera seeing a non-rigid object
- moving camera seeing a rigid object
- stationary camera seeing a moving rigid object
- moving camera seeing a moving rigid object . . .



# 11.1 General Theory

## The problem

- two camera seeing a non-rigid object
- moving camera seeing a rigid object
- stationary camera seeing a moving rigid object
- moving camera seeing a moving rigid object . . .

# 11.1 General Theory

## input

- noisy image point correspondences:

$$x_\alpha = \bar{x}_\alpha + \Delta x_\alpha$$

$$x'_\alpha = \bar{x}'_\alpha + \Delta x'_\alpha$$

- noise characteristic: $\Delta x_\alpha \in \mathcal{N}(0, V[x_\alpha])$ and $\Delta x'_\alpha \in \mathcal{N}(0, V[x'_\alpha])$
- i.e. non realistic noise assumptions

Extra post-presentation note:
Non realistic, because their are usually outliers as a result of mismatches. Also, usually point correspondences resulted from somewhat different 3d landmarks, because of view point change et al.

# 11.1 General Theory

## Epipolar constraint

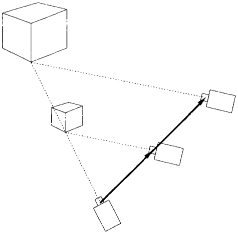
$$|\bar{x}_\alpha, h, R\bar{x}'_\alpha| = 0 \tag{1}$$

## Scale ambiguity

$$|\bar{x}_\alpha, ch, R\bar{x}'_\alpha| = 0 \tag{2}$$

$$c|\bar{x}_\alpha, h, R\bar{x}'_\alpha| = 0 \tag{3}$$

Thus the scale of $h$ can not be determined.



# 11.1 General Theory

## DOF problem

Rotation $R$: +3
Translation $h$: +3
scale ambiguity: -1
net: 5

## DOF correspondence

3d location landmark: -3
2d image 1 location: +2
2d image 2 location: +2
net: 1

## #correspondences needed

$N >= 5$

# 11.1 General Theory

## Optimal estimation of $\{h, R\}$

- $\{h, R\} = min_{h,R} J[h, R]$ ... is given in 11.1.2 ... skipping for now
- non-linear, requires numerical search
- "Rigidity test":

$$J[\hat{h}, \hat{R}] > \chi^2_{N-5,95\%}$$

- "Focus of expansion".
  Just the location of the epipole, right? Extra post-presentation note: indeed

# 11.1 General Theory

## Optimal estimation of $\{h, R\}$

- $\{h, R\} = min_{h,R} J[h, R]$ . . . is given in 11.1.2 . . . skipping for now
- non-linear, requires numerical search
- "Rigidity test":

$$J[\hat{h}, \hat{R}] > \chi^2_{N-5,95\%}$$

- "Focus of expansion".
  Just the location of the epipole, right? Extra post-presentation note: indeed

# 11.1 General Theory

## Optimal estimation of $\{h, R\}$

- $\{h, R\} = min_{h,R} J[h, R]$ ... is given in 11.1.2 ... skipping for now
- non-linear, requires numerical search
- "Rigidity test":

$$J[\hat{h}, \hat{R}] > \chi^2_{N-5,95\%}$$

- "Focus of expansion".
  Just the location of the epipole, right? Extra post-presentation note: indeed

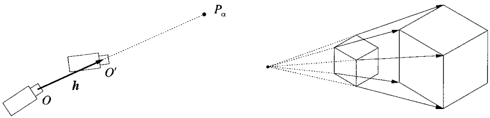
# 11.1 General Theory

## Optimal estimation of $\{h, R\}$

- $\{h, R\} = min_{h,R} J[h, R]$ ...is given in 11.1.2 ...skipping for now
- non-linear, requires numerical search
- "Rigidity test":

$$J[\hat{h}, \hat{R}] > \chi^2_{N-5,95\%}$$

- "Focus of expansion".
  Just the location of the epipole, right? Extra post-presentation note: indeed



# 11.1 General Theory

## Theoretical bound on accuracy

The general idea:

- Determine the covariance of $\{h, R\}$ given $V[x_\alpha]$'s and $V[x'_\alpha]$'s
- results in:

$$\begin{pmatrix} \bar{V}(\hat{h}) & \bar{V}(\hat{h}, \hat{R}) \\ \bar{V}(\hat{R}, \hat{h}) & \bar{V}(\hat{R}) \end{pmatrix} = \left( \sum_\alpha W_\alpha(\bar{h}, \bar{R}) \begin{pmatrix} \bar{a}_\alpha \\ \bar{b}_\alpha \end{pmatrix} \begin{pmatrix} \bar{a}_\alpha \\ \bar{b}_\alpha \end{pmatrix}^T \right)^-$$

where $a_\alpha = x_\alpha \times Rx'_\alpha$ and $b_\alpha = (x_\alpha, Rx'_\alpha)h - (h, Rx'_\alpha)x_\alpha$

## Practical bound on accuracy?

- replace all $\bar{}$s by $\hat{}$s and $a$ by $P_{\hat{h}}a$ to get a practical covariance measure of the motion (?)
- Question is, how useful it is, with all the linearization.

# 11.1 General Theory

## side note: Mystic derivation of bound

In Equation 11.29 $\begin{pmatrix} \Delta h \\ \Delta\Omega \end{pmatrix} \begin{pmatrix} \Delta h \\ \Delta\Omega \end{pmatrix}^T$ is expanded using 11.26. The idea is to compute the $\sum_\alpha(\ldots)\sum_\beta(\ldots)$.

It would (for me...) be more clear if

- $\bar{W}_\beta$ was written out: $W_\beta(\bar{h}, \bar{R})$
- $\begin{pmatrix} \bar{a}_\alpha \\ \bar{b}_\alpha \end{pmatrix} \begin{pmatrix} \bar{a}_\beta \\ \bar{b}_\beta \end{pmatrix}^T$ was used
- $\sum_\alpha(\ldots)\sum_\beta(\ldots)$ instead of $\sum_{\alpha,\beta}$.

Maybe it's just me....

# 11.2 Linearization and Renormalization

## The essential matrix $G$

$$|\bar{x}_\alpha, h, R\bar{x}'_\alpha| = 0 \tag{4}$$

rewrite:

$$(\bar{x}_\alpha, \quad h \times R \quad \bar{x}'_\alpha) = 0 \tag{5}$$
$$(\bar{x}_\alpha, \quad G \quad \bar{x}'_\alpha) = 0 \tag{6}$$

better know as:

$$\bar{x}_\alpha^T E \bar{x}'_\alpha = 0 \tag{7}$$

which is related to fundamental matrix $F$, which incorporates some “linear” camera calibration parameters:

$$(\bar{x}_\alpha, \quad K(h \times R)K'^T \quad \bar{x}'_\alpha) = 0 \tag{8}$$
$$(\bar{x}_\alpha, \quad F \quad \bar{x}'_\alpha) = 0 \tag{9}$$

# 11.2 Linearization and Renormalization

## Use $G$ to estimate $h$ and $R$

Two steps:

- estimate $G$ such that $(\bar{x}_\alpha, G\bar{x}'_\alpha) = 0$ (this Section)
- decompose $G$ in $h$ and $R$ (next Section)

## Estimating $G$

- $G$ has 9 elements, but scale ambiguity: 8 DOF
- Thus minimum nr of correspondences = 8
- ...eight-point-algorithms

## 11.2 Linearization and Renormalization

### Linear estimation of $G$

- first rewrite $(\bar{x}_\alpha, G\bar{x}'_\alpha) = 0$ into an $Mg = 0$ problem:

$$\begin{bmatrix} X_1X'_1 & \cdots & X_NX'_N \\ X_1Y'_1 & \cdots & X_NY'_N \\ X_1Z'_1 & \cdots & X_NZ'_N \\ Y_1X'_1 & \cdots & Y_NX'_N \\ Y_1Y'_1 & \cdots & Y_NY'_N \\ Y_1Z'_1 & \cdots & Y_NZ'_N \\ Z_1X'_1 & \cdots & Z_NX'_N \\ Z_1Y'_1 & \cdots & Z_NY'_N \\ Z_1Z'_1 & \cdots & Z_NZ'_N \end{bmatrix}^T \begin{bmatrix} g_{11} \\ g_{12} \\ g_{13} \\ g_{21} \\ g_{22} \\ g_{23} \\ g_{31} \\ g_{32} \\ g_{33} \end{bmatrix} = 0, \qquad (10)$$

  in which $(X, Y, Z)^T$ are the coordinates of $x$.
- This is similar to 11.7 and 11.39.
- The eigen-vector $g^*$ associated with the smallest eigenvalue of $M^TM$ minimizes $Mg^*$

# 11.2 Linearization and Renormalization

## Iterative re-weighting

- Minimizing the residual is not what we want
- We have to iteratively weight it using $W$ given in 11.12 and 11.41
- Disregarding the noise-variance ($V[x] = I$) this is:

$$W_\alpha = \frac{1}{||G^T x_\alpha||^2 + ||G x'_\alpha||^2 + g^T g}$$

- This is very similar (but not the same (?)) to square "Sampson Weng" weights:

$$Wi^{sw}_\alpha = \frac{1}{||G^T x_\alpha||^2 + ||G x'_\alpha||^2}$$

Extra post-presentation note:
Sampson-Weng weights are determined by taking the partial derivatives of the algebraic errors with respect to the pixel locations: $\frac{\partial(\hat{x}_\alpha, G\hat{x}'_\alpha)}{\partial \hat{x}_\alpha, \hat{x}'_\alpha}$.
If $V[x] = I$ and $\epsilon = 0$ then Sampson-Weng weights are equivalent to Kanatani's.
However, it is unclear what $\epsilon = 0$ (the average overal scale of the pixel error) would mean...
Anyways, Kanatani does take into account non-isotropic noise, which is nice.

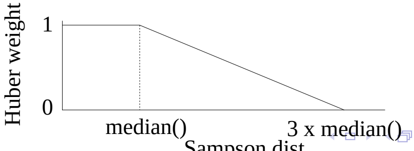
# 11.2 Linearization and Renormalization

## Iterative renormalization

- Kanatani shows this method (including reweighting) is statistically biased.
- Thus: renormalization as explained in Chap 9, by compensating for the bias and estimating the noise (p335).
- Kanatani shows in 2007 that renormalization for motion estimation works better than HEIV.... ( Extra post-presentation note: in "Performance evaluation of iterative geometric fitting algorithms")

## Adding robustness....

- Perhaps some of the correspondences resulted from mismatches.
- Check by computing Sampson distance = residual * weights
- Use robust weighting scheme, eg: Huber:



# 11.3 Optimal Correction and Decomposition

## ...Step 2: from $G$ to $h$ and $R$

Two possible tracks:

- make $G$ decomposable and use "non-robust" decomposition (Horn style)
- decompose $G$ using "robust" decomposition (H&Z style) ( Extra post-presentation note: not H&Z style, they also first make it decomposable.)

more about this later....

## Making $G$ decomposable

- apply svd: $USV^T = \text{svd}(G)$
- $G' = U\text{diag}(1, 1, 0)V^T$

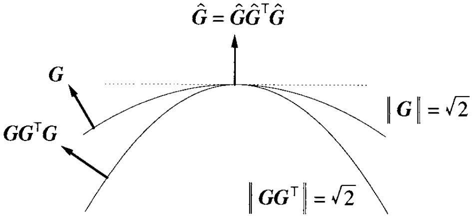
But Kanatani gives an extension also taking $V[G]$ into account.

Extra post-presentation note: see also the work of Ondrej Chum on Oriented epipolar constraint (also termed Ch(e)irality constraint).

# 11.3 Optimal Correction and Decomposition

## $9 - 3 \neq 5$

- Strange: $G$ has nine elements and 3 constraints (Eq 11.59), but only 5 DOF.




- Hartley&Zisserman pointing out the two equal eigen values, resulting in 1 DOF in the svd:

$$G = U \begin{pmatrix} \cos(\phi) & -\sin(\phi) & 0 \\ \sin(\phi) & \cos(\phi) & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \cos(\phi) & \sin(\phi) & 0 \\ -\sin(\phi) & \cos(\phi) & 0 \\ 0 & 0 & 1 \end{pmatrix} V^T$$

# 11.3 Optimal Correction and Decomposition

## "Robust" decomposition

- Uses two svds...
- In general 4 solutions (due to modeling light-rays as lines)
- Kanatani removes 2 by forcing Z-coordinates > 0
- Another one is removed on p342 of Sec 4 after reconstruction...
- This is no good for omnidirectional cameras (see also not 11 p358)

## "Non-robust" decomposition

- faster...
- H&Z use svd for this

# 11.3 Optimal Correction and Decomposition

## "Robust"-track or "Non-robust"-track

- Optimal correction for decomposability seems... more optimal
- Experiments using simple H&Z decomposition shows minor improvement....
- Why not force decomposability in the iterative reweighting scheme?

Extra post-presentation note: There was some discussion about the result of these two tracks, i.e.: if they would result in different $h$ and $R$'s. I think they would be the same...

## Missing in this section

A covariance estimate of $h$ and $R$ given $V[G]$...
(Eg 11.31 should be used)

# 11.4 Reliability of 3-D Reconstruction

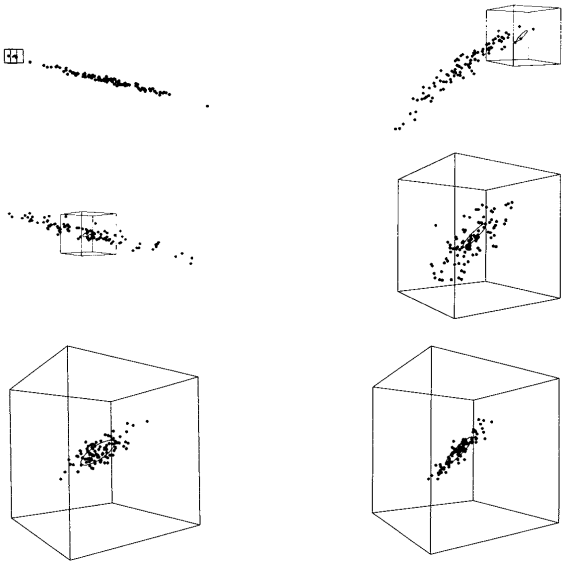
## Reconstruction and its variance

- Nothing much to add (anyone?)
- Is this similar to the "optimal triangulation method" from H&Z ?

## Reconstruction for better motion

- By reconstructing mismatches can be determined
- If during iterative reweighting set their weights to 0
- For Ransac use it to
  - ignore hypotheses
  - remove support

# 11.4 Reliability of 3-D Reconstruction



# 11.5 Critical Surfaces

## Note on critical surfaces

- CV-people get a kick out of planar surfaces. Extra post-presentation note: This is most probably because estimating homographies is more straightforward than epipolar geometry estimation.
- Actually surfaces are never planar in real life, ambiguities should be expressed in the uncertainty, right?
- Kanatani says so on p367
- What is a "false" essential matrix?

## Different critical-categories

- weak: only ambiguity in $G$
- strong: also ambiguity in $h$ and/or $R$

# 11.6 3-D Reconstruction from Planar Surface Motion

## Homography $A$

- If planar surface, then estimate homography
- homography: a projective transformation from 2d to 2d

## Estimation

- Kanatani gives separate Homography algorithm
- Homography should be used if nearly planar (bij twijfel niet inhalen...)
  Extra post-presentation note: This is confirmed by Isaac, who experienced bad essential matrix estimation of images taken from the front of buildings/houses.

# 11.6 3-D Reconstruction from Planar Surface Motion

## Homography DOFS

- camera position +3
- camera rotation +3
- plane position +3
- scale ambiguity -1
- net: 8

## DOF correspondence

3d location landmark on plane: -2
2d image 1 location: +2
2d image 2 location: +2
net: 2

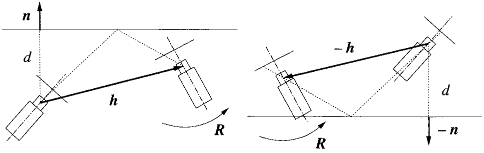
## #correspondences needed

$N >= 4$

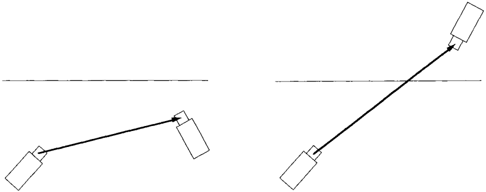
# 11.6 3-D Reconstruction from Planar Surface Motion

## From $A$ to $h$ and $R$

- $A$ already decomposable (because same degrees of freedom)
- But more ambiguities (7 pages on this...)
  - The same 4 as $G$ stemming from rays as lines

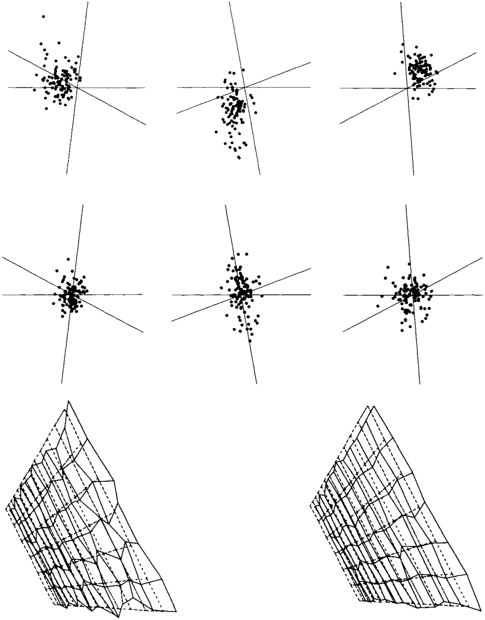



  - + an ambiguity for different sides of the plane



  - last one can not be resolved. Extra post-presentation note: yes it can: force determined to be +1 (p357).

# 11.6 3-D Reconstruction from Planar Surface Motion



# 11.7 Camera Rotation and Information

## Rotation only

- Looks like planar surface: points on plane at infinity
- $R$ can be computed using $G$-track and $A$-track (?) Extra post-presentation note: indeed

## DOF problem

Rotation $R$: +3
net: 3

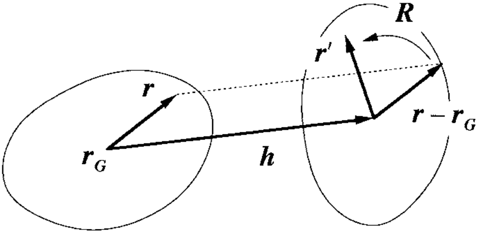
## DOF correspondence

3d location landmark on plane at infinity: -2
2d image 1 location: +2
2d image 2 location: +2
net: 2

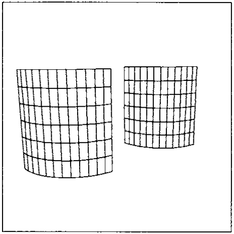
## #correspondences needed

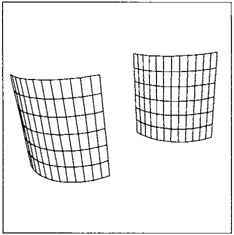
$N >= 1.5$ (i.e. 2 overdetermines)

# Extra, copied figs I did not use

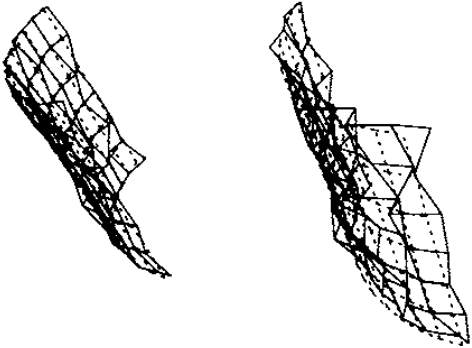


# Extra, copied figs I did not use

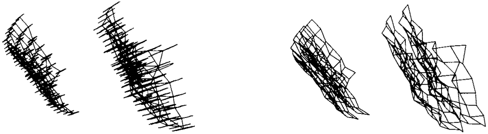




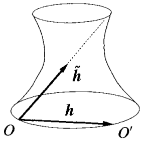
# Extra, copied figs I did not use

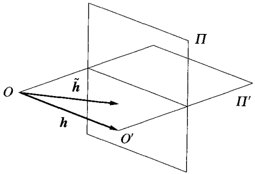


# Extra, copied figs I did not use

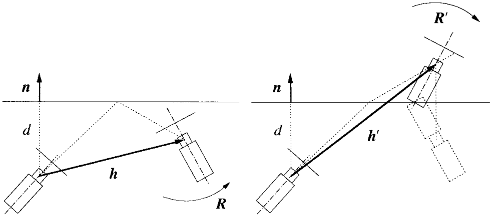


# Extra, copied figs I did not use

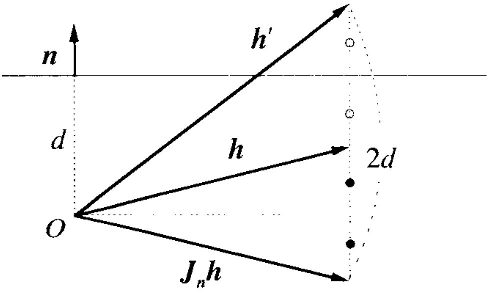


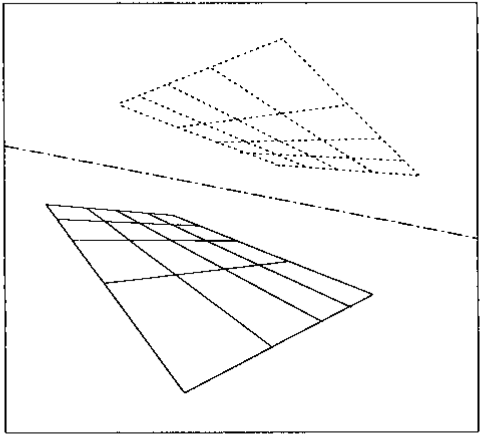
# Extra, copied figs I did not use



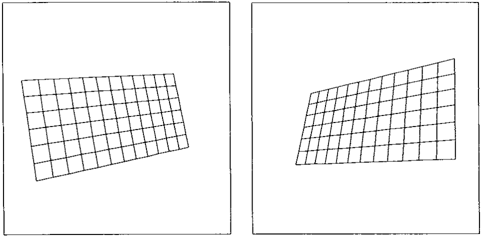
# Extra, copied figs I did not use



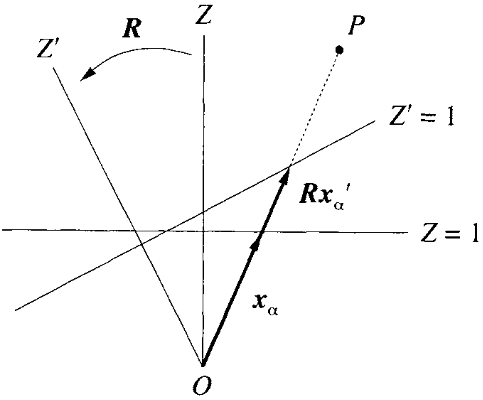
# Extra, copied figs I did not use



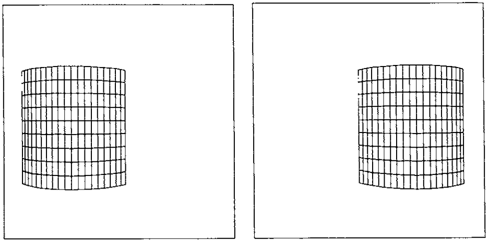
# Extra, copied figs I did not use



# Extra, copied figs I did not use



# Extra, copied figs I did not use



# Extra, copied figs I did not use