Toward 3D Topography: Using Curvature-based Geometry Measuring Machine

Z.Q. John Lu (John.lu@nist.gov)
Information Technology Laboratory
National Institute of Standards and Technology (NIST)
Gaithersburg, Ma4/23/2007ryland

Prepared for NIST Workshop on 3D and 2D Content Representation, Analysis and Retrieval , April 23-24, 2007

I. General Questions

1. **What is shape?**
   - **"Shape" for us means "what is left when the effects associated with translation, scaling and rotation are filtered away". D.G. Kendall (1977) , H. Le and D. G. Kendall (1993, Ann. Stat., Vol. 21, No.3, 1225-1271.)**

2. **Size-and-shape**
   - **Geometrical information that remains when location and rotational effects are filtered out from an object. I.L. Dryden and K.V. Mardia (1998, Wiley book)**

I. General Questions: Statistical Shape Analysis

- Landmarks: Points of correspondence on each object that matches between and within populations.
- Three types:
  - Anatomical landmarks: by experts which correspond between organisms in some biologically meaningful way, e.g. homologous parts
  - Mathematical landmarks: according to some math/geomtical property of the figure, e.g. high curvature, extreme points
  - Pseudo-landmarks: constructed points around the outline and between anatomical or mathematical landmarks, e.g. continuous curves or surfaces by a large number of equally spaced points

I. General Questions: Shape Analysis (cont.)

- Traditional methods: multivariate morphometrics using multivariate statistics (e.g. C.R. Rao's work, Mahalanobis distance)
- Geometrical methods: deformation and transformation, shape is inherently non-Euclidean-----first proposed by D'Arcy Thompson (1917).
  - Geometrical shape analysis in 70s: D.G. Kendall, F.L. Bookstein, mainly on landmark data
  - Pattern theory since the 70s: developed by Ulf Grenander and colleagues using deformable templates, algebraic groups (Lie Groups), Markov graphs, metric theory, and stochastic process simulation. Latest advances such as computational anatomy summarized in: Pattern Theory, from Representation to Inference (with Michael Miller, 2007, OUP)

II. Current practice (cs approach)

- Feature extraction:
  - Representation of the infinite-dimensional shape space by decomposition using some orthogonal basis (spherical harmonics, PCA, wavelets) with a few numbers.
- Define distance on vectors: dissimilarity or similarity measures
- Clustering and classification: data mining, statistical learning, or retrieval

What "Distance" Metrics to Choose

Distribution Functions: To impose a distance metric on high-dimensional multinomial distributions: $(\pi_{i1}, \ldots, \pi_{ip})$, i=1,2. For example,

$$g(\vec{\pi}_1, \vec{\pi}_2) = \tfrac{1}{2}\sum_{i=1}^{p} (\pi_{1i} - \pi_{2i})^2$$

–Diversity connection: this distance may seem very naïve, but it is related to the Gini-Simpson index: probability of mismatching, but ……

$$1 - \sum_{i=1}^{p} \pi_i^2$$

– Is this Euclidean distance appropriate? Composional Data Analysis?

–Generalizations due to C.R. Rao (1982) and others

Differential metrics on distributions

- Geodesic distance between two probabilities for multinomial distribution: $\vec{\pi}_1, \vec{\pi}_2$

$$g(\vec{\pi}_1, \vec{\pi}_2) = 2\cos^{-1}\left(\sum_1^p \sqrt{\pi_{1i}\pi_{2i}}\right)$$

- Information metrics: Kullback-Leibler distance:

$$d(\vec{x}, \vec{y}) = KL(\vec{y} \mid \vec{x}) + KL(\vec{x} \mid \vec{y}) = \sum_1^p (x_i - y_i) \log \frac{x_i}{y_i}$$

NIST
National Institute of Standards and Technology

INFORMATION TECHNOLOGY LABORATORY

Statistical Issues in Shape Inference

1. How to define the "mean shape", and how to quantify uncertainty on shape measurements
2. How to test whether two shapes are same within measurement errors, or comparable/ interoperable

    1. How to define the mean curve?
    2. How to decide significance?
    3. ANOVA to establish repeatability/reproducibility

NIST National Institute of Standards and Technology

ITL INFORMATION TECHNOLOGY LABORATORY

General regression model

- Relating a single attribute to high-dimensional inputs, predictors, or measurands

  $$Y=m(X)+\varepsilon$$

  then try to fit this model to a set of measured data (X_i, Y_i), $i=1,2,\ldots,n$.
- No assumption of m, except that m is a smooth function of X.
- What about the accuracy (consistency) of estimating m?

NIST
National Institute of Standards and Technology

INFORMATION TECHNOLOGY LABORATORY

Review: Kernel regression:

- Smoothing as "local averages": kernel estimation:
- Kernel regression: Nadaraya (1964), Watson (1964)

$$\hat{m}(x)=\Sigma_{i=1}^{n} K(||X_i - \mathbf{x}|| / h)\ Y_i \ / \ \Sigma_{i=1}^{n} K(||X_i - \mathbf{x}|| / h)$$

- Reasoning: weighting function is a monotone function of distance between predictor data vectors: usually Euclidean distance

NIST
National Institute of Standards and Technology

ITL Information Technology Laboratory

A New Kind of Theory

- In order to accommodate high-dimensional data, need the concept of "intrinsic dimension": the most important features lie a much lower dimension than data space

- A singular design model: predictor variables do not have a "joint density".

Singular design model

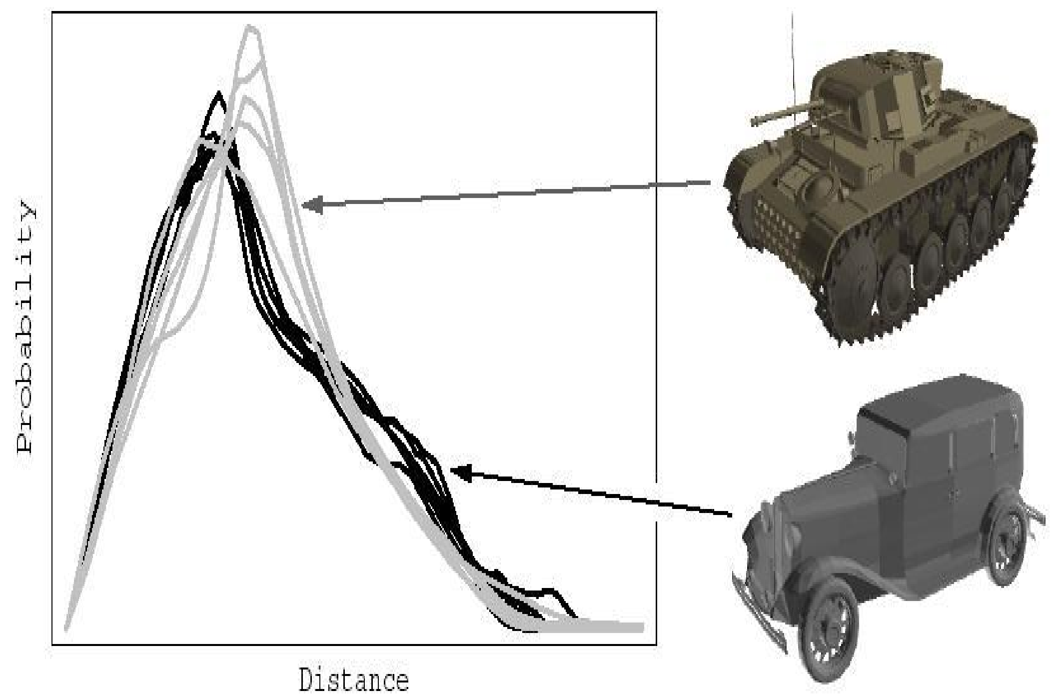
- A theoretical setup for singular design: no design density exsiting: singular design measure
- Notion of intrinsic dimensionality: e.g. number of independent variables, pointwise fractal dimension:

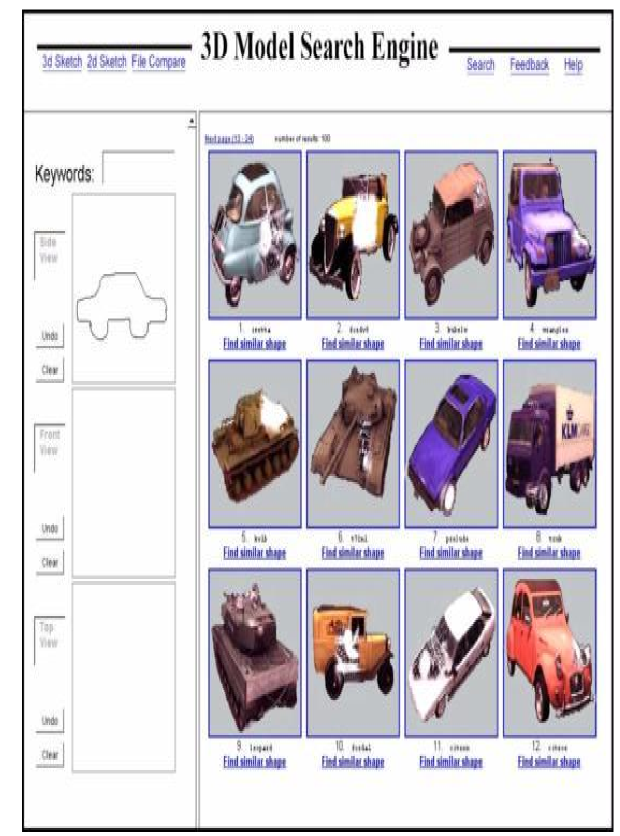
  $$P(||X-\mathbf{x}|| <= h) \sim h^d \text{ as } h \to 0$$

  where $d <= p$ ($d << p$ very likely)
- First proposed in Z.Q. Lu (1999): Nonparametric with Singular Design, *Journal of Multivariate Analysis*
- Morale: **no matter how many number of variables, as long as the underlying intrinsic dimension is low, the statistical accuracy is still good**.

Princeton Shape Retrieval and Analysis Group

Shape distribution





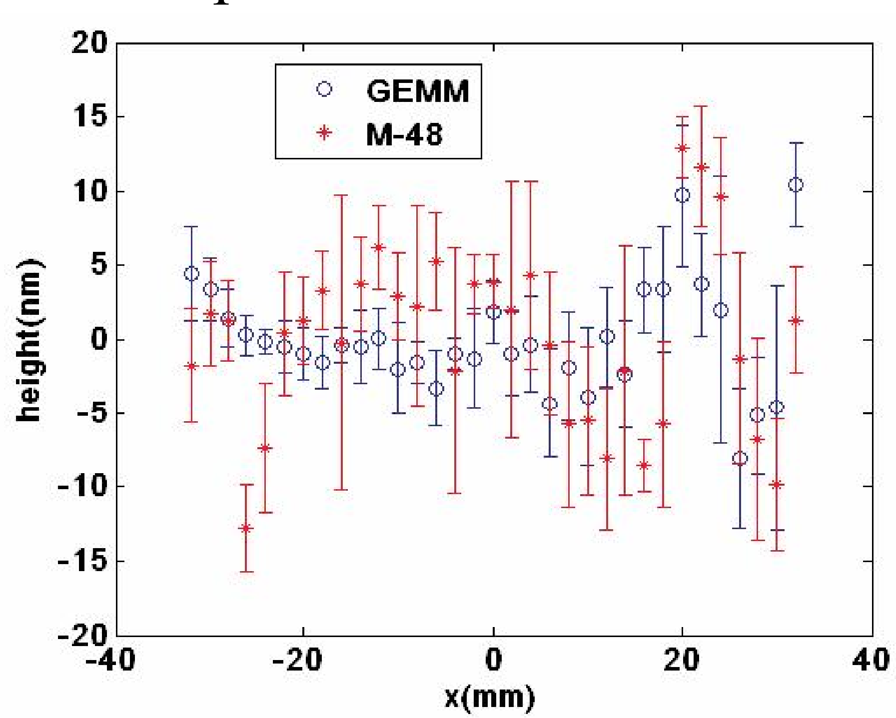
A Statistical Testing Problem

Volume 111, Number 5, September-October 2006
Journal of Research of the National Institute of Standards and Technology

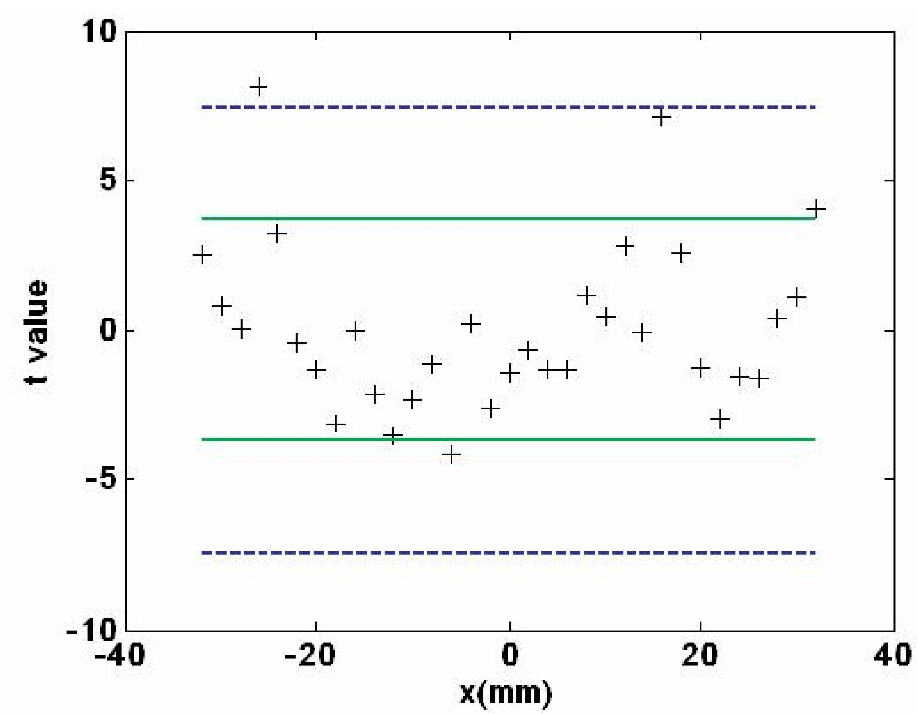
[J. Res. Natl. Inst. Stand. Technol. 111, 373-384 (2006)]

Form-Profiling of Optics Using the Geometry Measuring Machine and the M-48 CMM at NIST

Compare GEMM and CMM

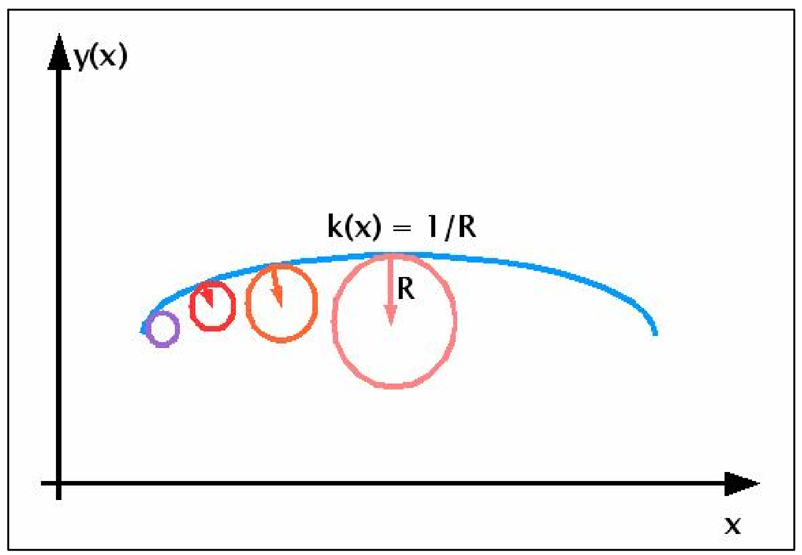


Statistical test of significance



III. Curvature as a Shape Descriptor

Curvature for curves in 2D space



$$\frac{d^2 y(x)}{dx^2} = k(x)\left[1 + \left(\frac{dy(x)}{dx}\right)^2\right]^{\frac{3}{2}}$$

There are three types of curvature definitions for a surface lying in 3D

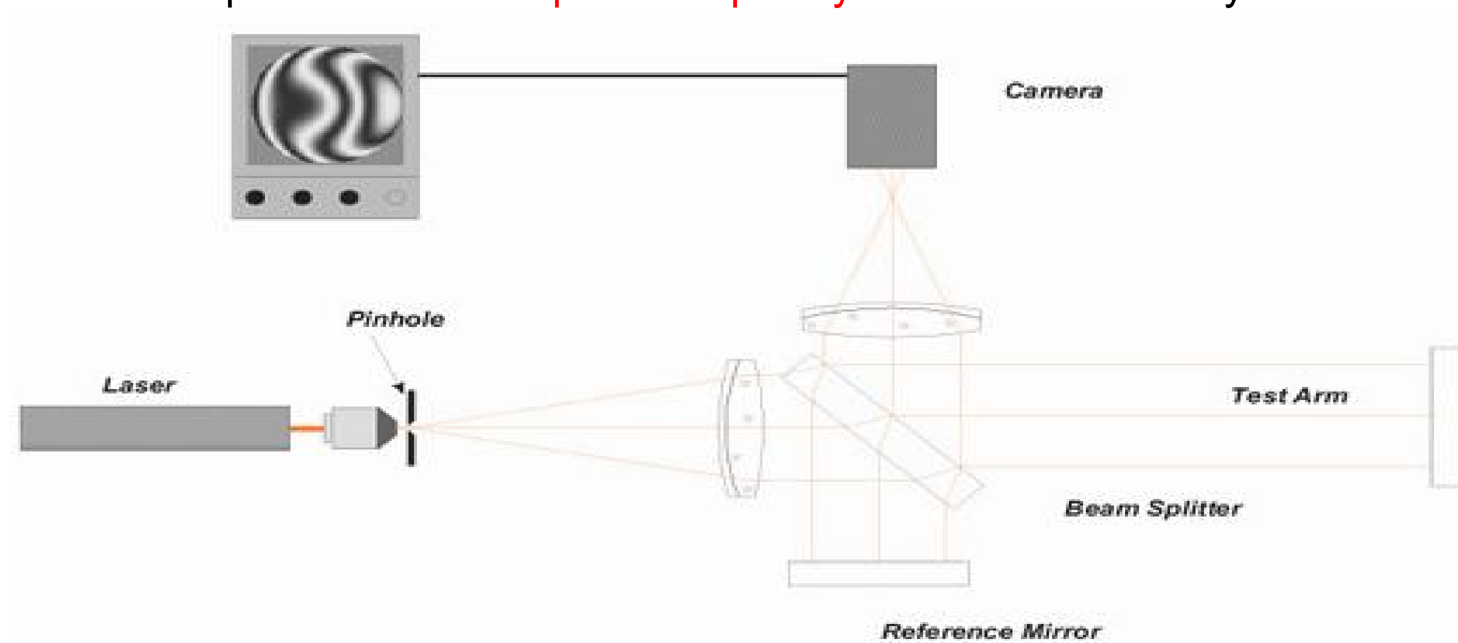
$$K = \frac{z_{xx} z_{yy} - z_{xy}^2}{\left(1 + z_x^2 + z_y^2\right)^2},$$

$$H = \frac{1}{2}\frac{(1+z_x^2)z_{yy} - 2z_x z_y z_{xy} + (1+z_y^2)z_{xx}}{(1+z_x^2+z_y^2)^{3/2}}$$

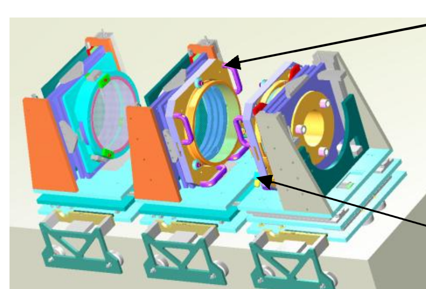
$$k_{1,2} = \frac{1}{2}\left(z_{xx} + z_{yy}\right) \pm \sqrt{\frac{1}{4}\left(z_{xx} - z_{yy}\right)^2 + z_{xy}^2}.$$

Metrology of Optical Surfaces

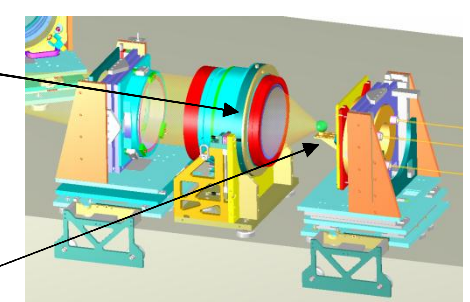
Phase - measuring interferometers are the most accurate metrology tools for measuring the surface form of flats and spheres with low spatial frequency content. Uncertainty ~ 1nm.



Light fields reflected by test- and reference surfaces are observed. Deviation of test part form from reference part form is calculated form the interference of the two fields.



Flat measurement



Sphere measurement

Two set ups of the **E**xtremely Accurate **Calib**ration **I**nterferomete**r** (XCALIBIR)

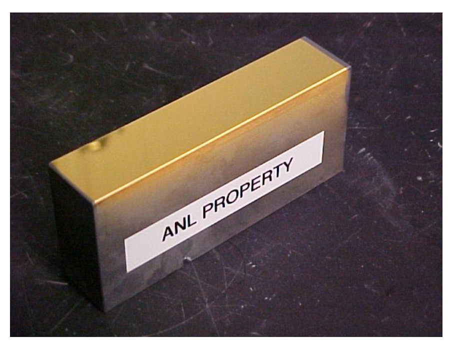
NIST National Institute of Standards and Technology

The Asphere Interferometry Problem

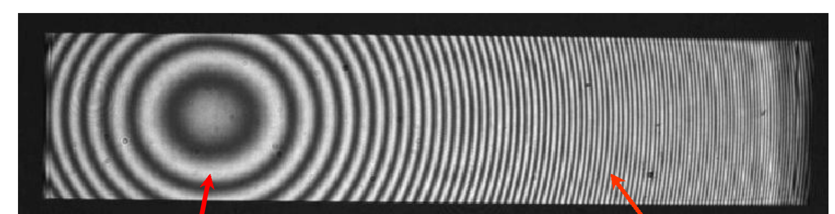
ITL Information Technology Laboratory

Metrology of Optical Free-Form and Asphere Surface using full aperture Interferometers have problems :
- Insufficient dynamic range,
- Impossible to realize common path condition.

Example:



An ellipical torus or free form surfaced X-ray optic, which is used in a Kirkpatrick- Baez imaging system.

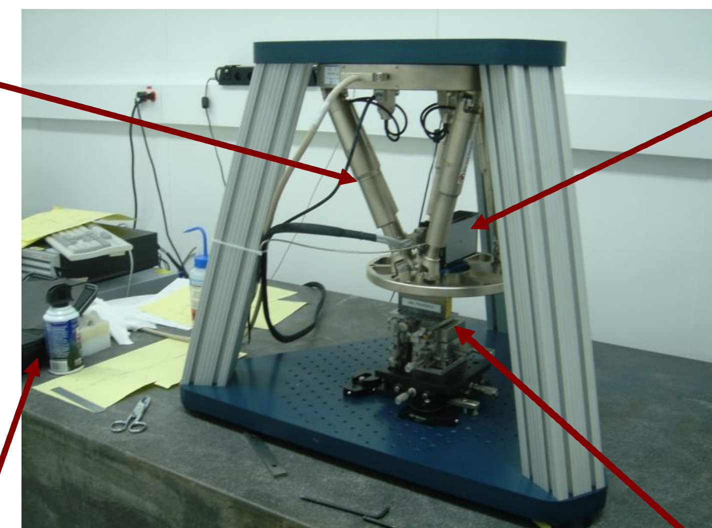


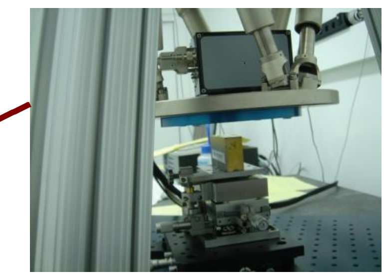
Non-zero fringe density **Insufficient dynamic range**

XCALIBIR interferogram for the X-ray mirror against flat reference.

The NIST-Geometry Measuring Machine (GEMM)

–6-axis stage moves
–curvature sensor and
–orients it normal to
–test surface.

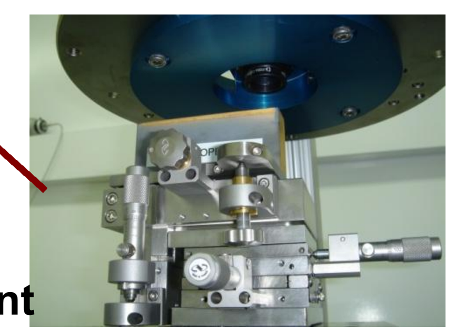




–A compact interferometer
–Is used as the sensor.

–Compressed Gas
–Duster

–Test part
–and mount



Measurements of GEMM: local topographic images

$$\rho(u,v) = \tfrac{1}{2}\left(\mathbf{x}_{11}\cdot\mathbf{N}\,u^2 + 2\,\mathbf{x}_{12}\cdot\mathbf{N}\,uv + \mathbf{x}_{22}\cdot\mathbf{N}\,v^2\right) +$$
$$\frac{1}{6}\left(\frac{\partial^3\mathbf{x}}{\partial u^3}\cdot\mathbf{N}\,u^3 + 3\,\frac{\partial^3\mathbf{x}}{\partial^2 u\,\partial v}\cdot\mathbf{N}\,u^2 v + 3\,\frac{\partial^3\mathbf{x}}{\partial u\,\partial^2 v}\cdot\mathbf{N}\,u v^2\right.$$
$$\left. + \frac{\partial^3\mathbf{x}}{\partial v^3}\cdot\mathbf{N}\,v^3\right) + \cdots$$

Image 10

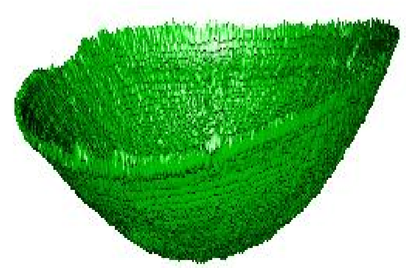


Image 20

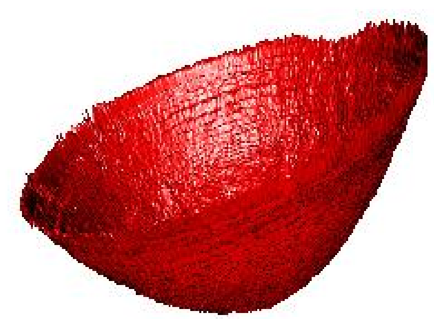
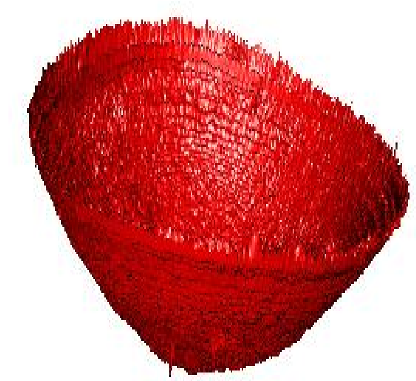
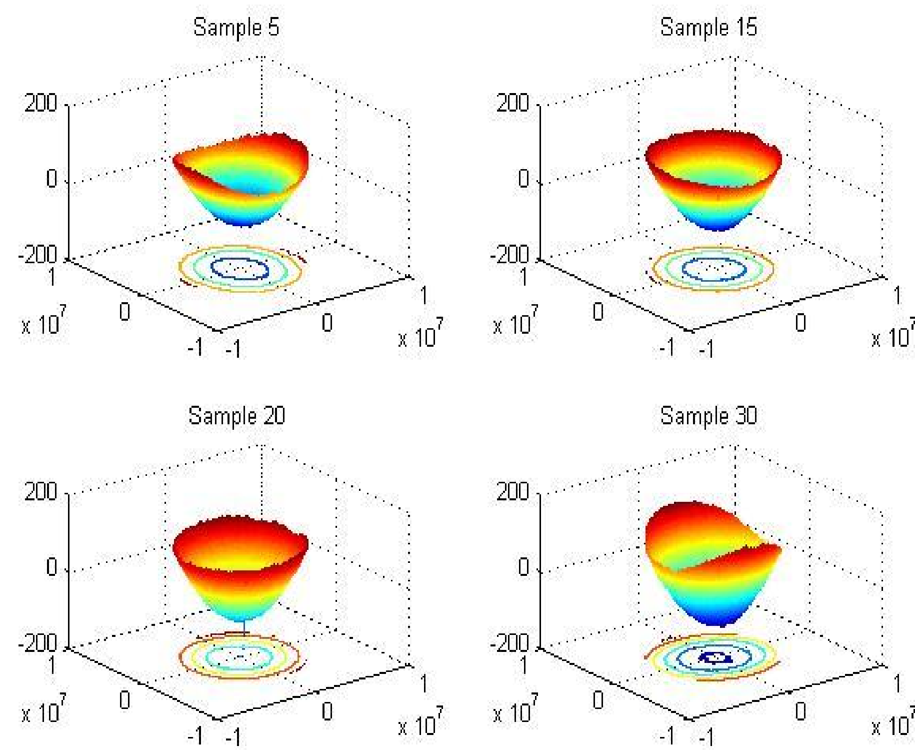


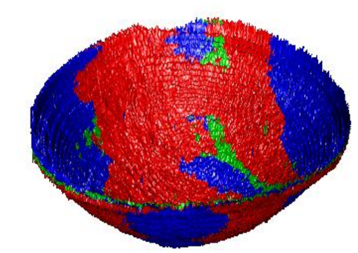
Image 25



More display of GEMM data

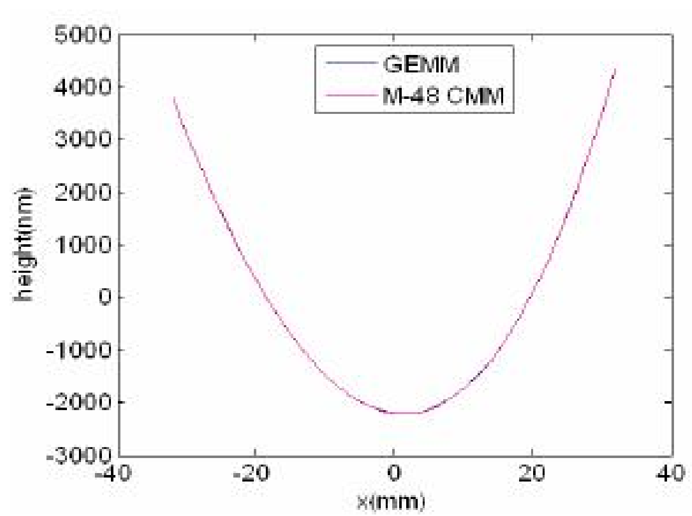


Overlaying three images: 05, 10, 15



2D Topography by GEMM

Gold mirror test



Reconstructed profiles

Differential equation approach

$$\frac{d^2 z(x)}{dx^2} = K(x)\left(1+\left(\frac{dz(x)}{dx}\right)^2\right)^{3/2}. \qquad (2)$$

When the curvature $K(x)$ is measured, Eq. (1) must be solved to determine the profile. This can be accomplished using one of the standard methods for solving differential equations. Alternatively, an integration procedure described by Elster et al. [16] can be used to solve Eq. (1), which is now briefly reviewed.

Let $P(x) = z'(x)$.

Then $P'(x) = K(x)[1 + P(x)^2]^{3/2}$.

Thus $\dfrac{P'(x)}{\left(1+P(x)^2\right)^{3/2}} = K(x)$.

So $\displaystyle\int \frac{dP}{\left(1+P(x)^2\right)^{3/2}} = \int K(x)dx$,

but $\displaystyle\int \frac{dP}{\left(1+P(x)^2\right)^{3/2}} = \frac{P(x)}{\sqrt{1+P(x)^2}}$.

Now let $\psi(x) = \int K(x)dx$.

We then have $\dfrac{P(x)}{\sqrt{1+P(x)^2}} = \psi(x)$.

And finally $P(x) = \dfrac{\psi(x)}{\sqrt{1-\psi(x)^2}} = z'(x)$. (3)

Volume 111, Number 5, September-October 2006
Journal of Research of the National Institute of Standards and Technology

[J. Res. Natl. Inst. Stand. Technol. 111, 373-384 (2006)]

NIST
National Institute of Standards and Technology

ITL Information Technology Laboratory

3D Topography: Curvature Estimation

We propose to use local polynomial regression: e.g, for local quadratic fit, we minimize:

$$\sum_i \left[z_i - a - b^T \begin{pmatrix} x_i - x_0 \\ y_i - y_0 \end{pmatrix} - \frac{1}{2} \begin{pmatrix} x_i - x_0 \\ y_i - y_0 \end{pmatrix}^T \begin{pmatrix} h_{11} & 2h_{12} \\ 0 & h_{22} \end{pmatrix} \begin{pmatrix} x_i - x_0 \\ y_i - y_0 \end{pmatrix} \right]^2 K\left(\frac{x_i - x_0}{h}, \frac{y_i - y_0}{h} \right),$$

for a,b, h_{ij}'s, where b *and* h_{ij}*'s* estimate the partial derivatives at (x_0,y_0).

Related References: Fan and Gijbel (1996), Ruppert and Wand (1994) on regression

Lu (1994, 1996, ***Journal of Multivariate Analysis*****) on first-order partial derivative estimation**

However, if there is strong nonlinearity, we need to use higher order polynomial such as 4th order polynomial be used.

3D Topography: Reconstruction

1. The need for smoothing in curvature measurements (typically noiser than data)

2. 3D reconstruction based on PDE may not be the best solution

3. We propose to use a statistical approach which handles noise data better, via smoothing spline/reproducing kernel (adapting Grace Wahba approach, 1990)

NIST
National Institute of Standards and Technology

Summary

ITL Information Technology Laboratory

- I wish to convey that there are many interesting problems, at least to me, in 3d and 2d shape analysis, cutting across many disciplines, which raise interesting metrology challenges, and standard developments demand novel and challenging high-dimensional statistical research.
- GEMM for topographic measurements of aspheric objects with nanoscale precision, comparable to CMM, but avoids contact
- Interdisciplinary research takes time, energy, patience, and is rewarding for the initiated participants.

Thanks!

Nadia Machkour-Deshayes, CNRS, France,

Charles Hagwood (NIST/ITL),

Johannes Soons (NIST/MEL)