Haptic Access to Virtual Models of Real Data for Training the Visually Impaired

K. Moustakas, G. Nikolakis,
D. Tzovaras and M. G. Strintzis

Informatics and Telematics Institute / Centre for Research and Technology Hellas

Outline

- Introduction - Problem formulation
- Real-time 3D scene representation
  - Structure from motion
  - 3D model generation
    - Parametric model recovery
    - Superquadric approximation
  - Superquadric modeling of the virtual hand
- Experiments and applications
  - 3D haptic representation for the blind
- Conclusions-Discussion

Introduction

- The interest of the global scientific community on haptics enabled virtual environments has been increased during the latest years because:
  - Haptic interaction provides the user with a strong feel of realism
  - Applications for disabled people can be developed to help them overcome their difficulties
  - Ease of use
  - Speed of communication and interaction

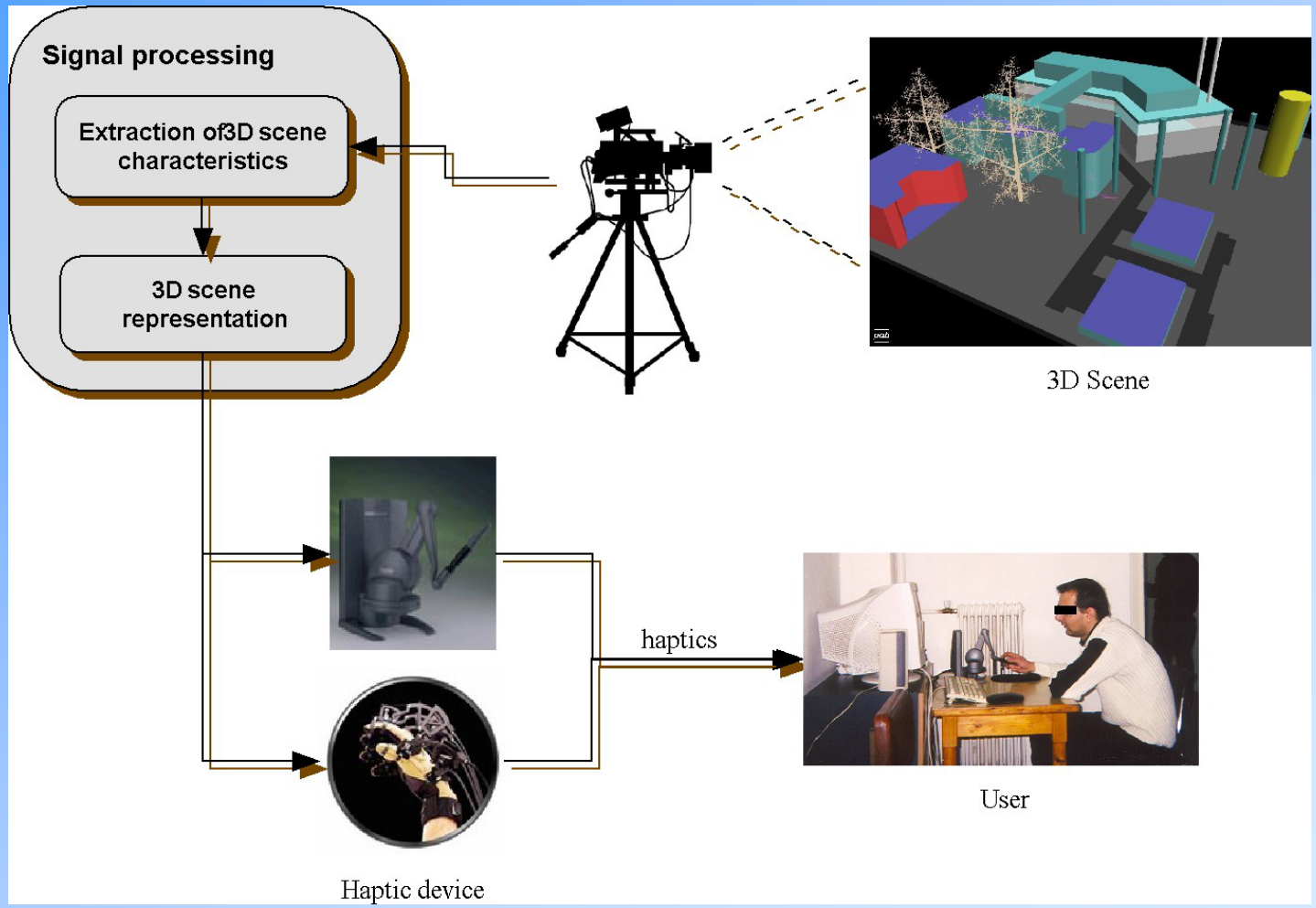
Haptic interaction

- Haptic representations of 3D scenes increase the realism of the HCI
- Despite the fact that the major modalities for HCI are video and audio, VR applications are far from being immersive without the use of haptics
- For some people (visually impaired) it is one of the major means of interacting with their environment

Overview of the developed system

- Input: 2D monoscopic video captured from a single camera
- Output:
  - 3D visual representation.
  - Haptic representation of the observed scene
- System consists of:
  - Structure from motion (SfM) extraction
  - 3D geometry reconstruction

Overview of the developed system

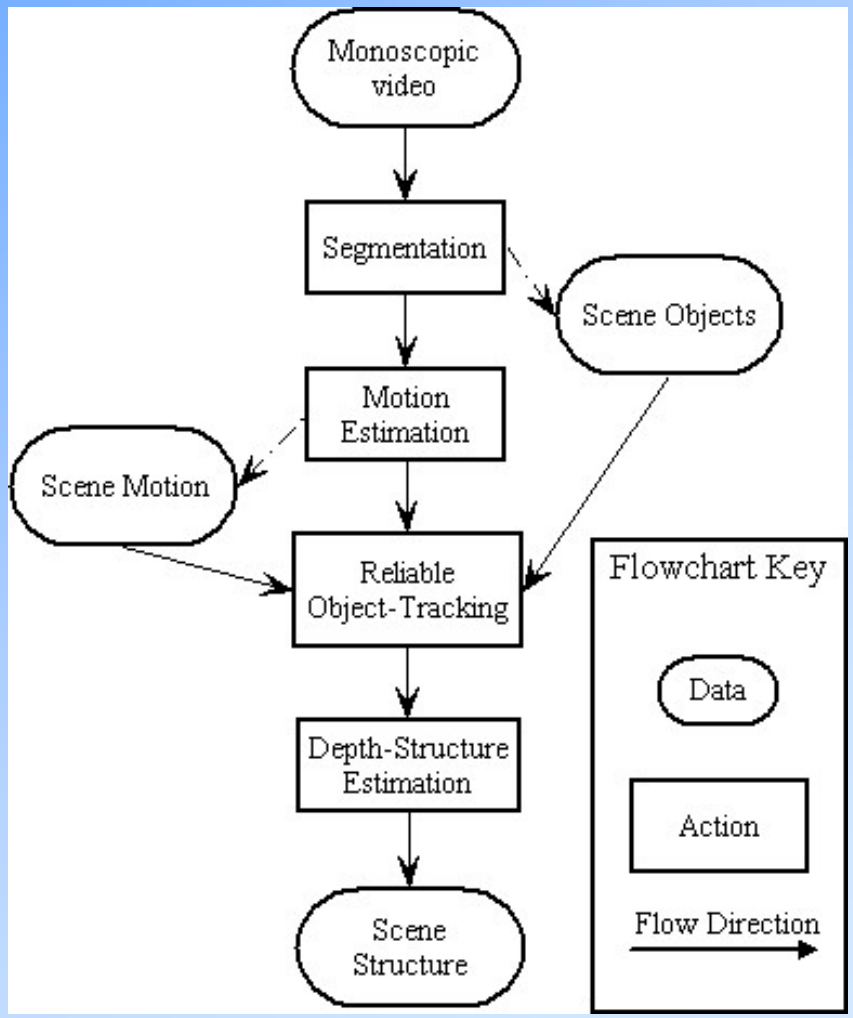


Overview of the developed system

- Step 1: SfM extraction from the monoscopic video
- Step 2:
  - Model parameter estimation
  - 3D scene generation
- Step 3: Haptic representation of the 3D scene.

Structure from motion

- Feature based motion estimation
- Extended Kalman Filter-based recursive feature point depth estimator
- Efficient object tracking
- Bayesian framework for occlusion handling



Model parameter estimation

- If the shape of the model is approximately known, which is the case for most specialized applications, parameters like translation, rotation, scaling, deformation, can be recovered from the SfM data, using least squares methods.
- If the mesh is of unknown shape a dense depth map of the scene is created and transformed into a mesh (terrain) utilizing Delaunay triangulation

3D model generation

- The resulting 3D structure data can be used:
  - in raw format, thus generating an image 3D mesh.
  - to estimate the parameters of existing parametric models if there exists knowledge on the objects composing the scene.
  - In specific tasks like the ones designed for the blind, there exists usually information about the objects in the scene.

3D model generation

- In cases where the objects are convex and relatively simple, superquadrics can be used to model them.
- Superquadrics have been excessively used to model range data.
- They are used to model the tower scene in the present application.

Superquadric approximation

- A superquadric is defined from the following equation:

$$F(x,y,z)=\left(\left(\frac{x}{a_1}\right)^{\frac{2}{\varepsilon_2}}+\left(\frac{y}{a_2}\right)^{\frac{2}{\varepsilon_2}}\right)^{\frac{\varepsilon_2}{\varepsilon_1}}+\left(\frac{z}{a_3}\right)^{\frac{2}{\varepsilon_1}}=1$$

- Parameters $\alpha_1,\alpha_2,\alpha_3,\varepsilon_1,\varepsilon_2$ have to be defined in order to minimize the error:

$$MSE=\sum_{i=1}^{N}\sqrt{a_1a_2a_3}\left(F(x_i,y_i,z_i)-1\right)^2$$

for the N recovered 3D points.

Virtual hand segmentation

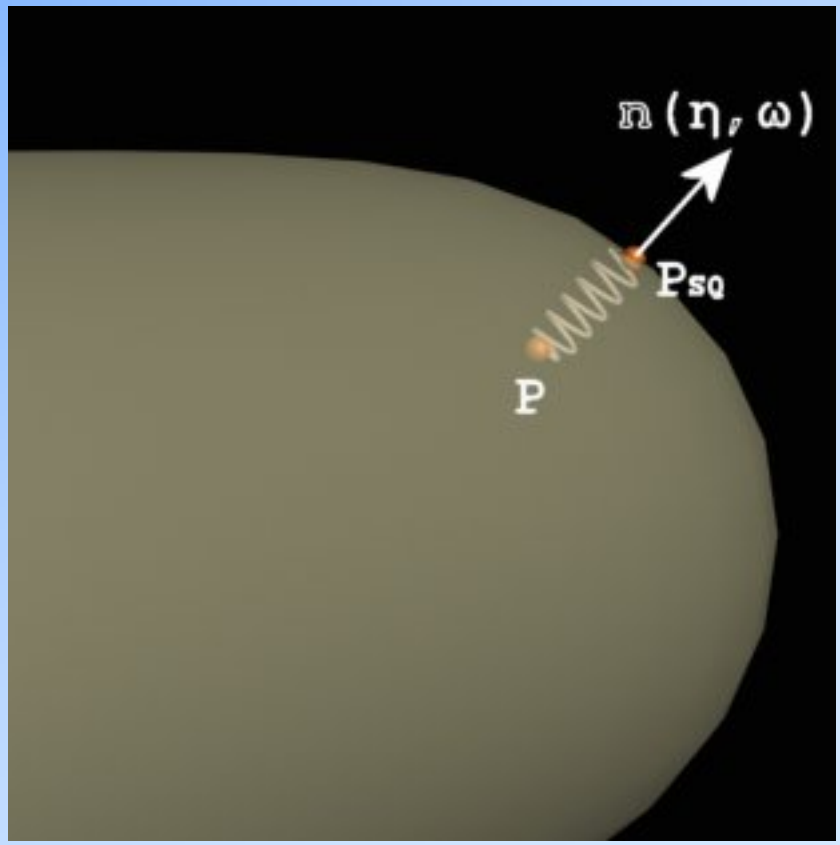
- The virtual hand is segmented into 16 rigid components
- Each segment is modeled using a superquadric
- Advantages:
  - Speed
  - Collision detection with the whole hand
  - Easy and fast calculation of force feedback if needed

Collision Detection

- Real time collision detection algorithm
- Collision detection is performed between the whole virtual hand and the scene objects and only between the fingertips.
- Collision detection takes advantage of the fact that the virtual hand can be efficiently modeled using implicit surfaces (superquadrics)

Force Feedback

- Force feedback is calculated directly from the superquadric modeling of the virtual hand using their analytical formulae.
- The resulting force feedback is smooth with no discontinuities.



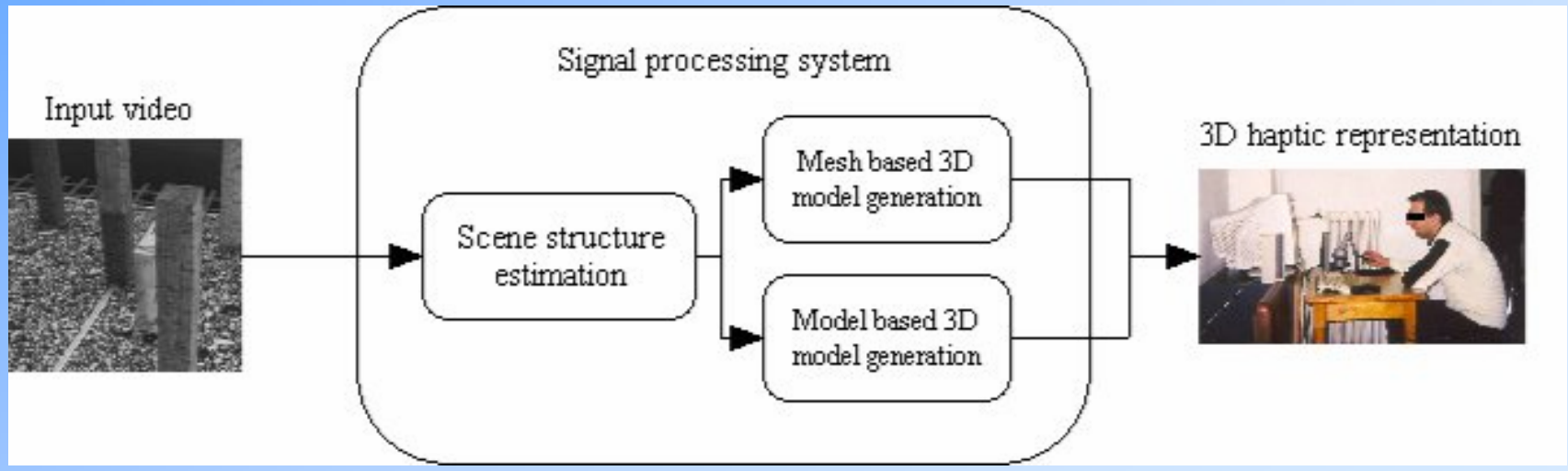
Haptic representation

- The extracted 3D scene used as input for the two haptic devices:
  - **Phantom**: 6 DOF for motion and 3 DOF for force feedback.

  - **CyberGrasp**: 5 DOF for force feedback (1 for each finger)

Application

- 3D haptic representation for the blind. The visually impaired user examines the 3D virtual representation of a real scene using haptic devices.

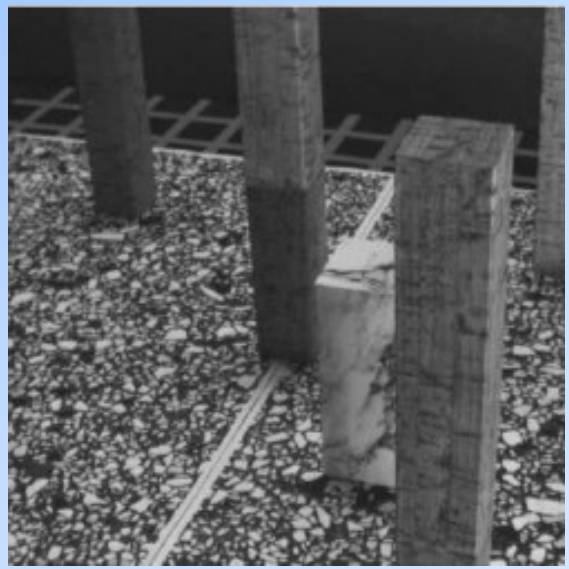


3D haptic representation for the blind

- The scene is captured using a standard monoscopic camera.
- SfM methods are utilized to estimate scene structure parameters.
- The 3D model is generated either from existing parametric models or using the raw SfM mesh.
- The resulting model is fed onto the haptic interaction devices.

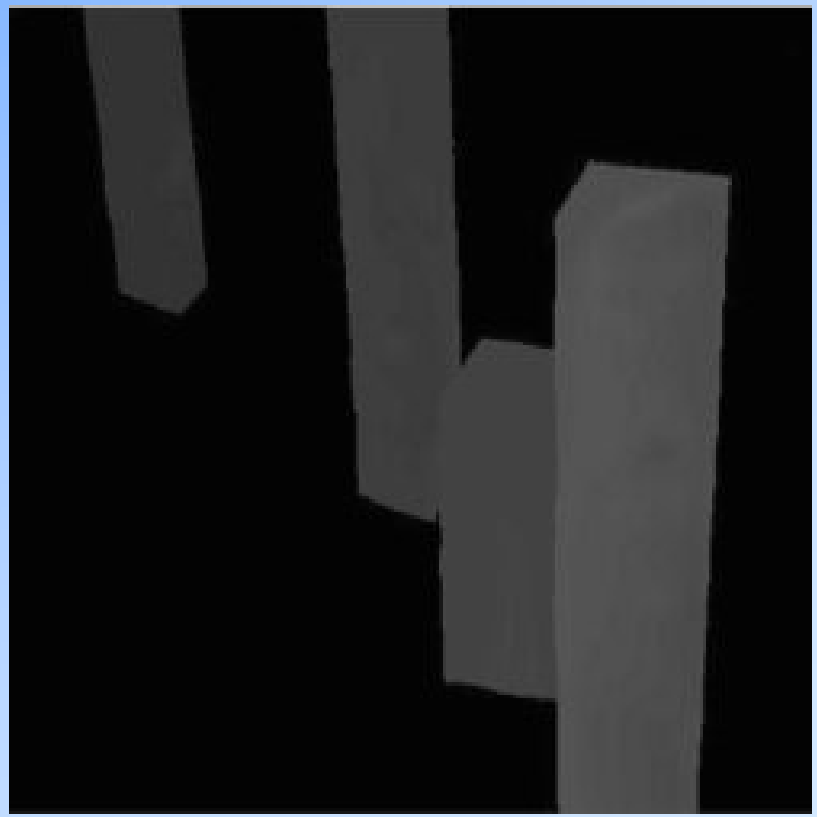
Example: tower scene

- The tower scene consists of four main parallelepipeda moving mainly across the horizontal direction.

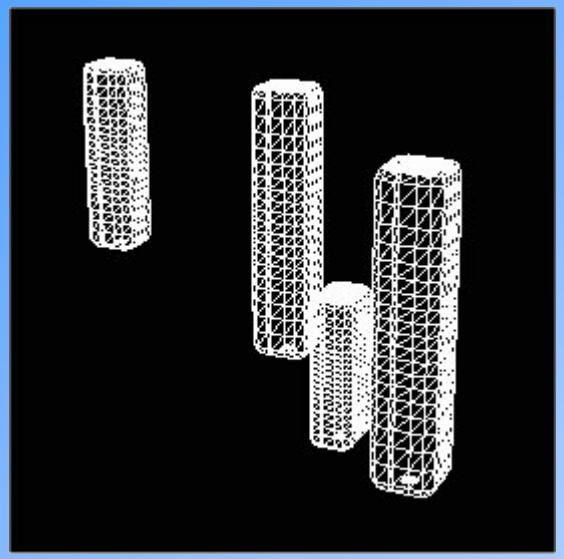


Structure reconstruction

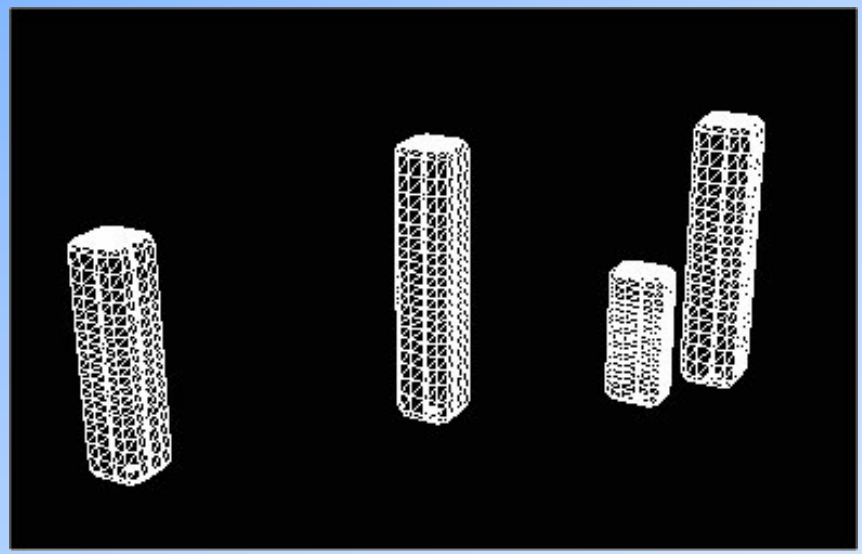
- After SfM is performed the resulting dense depth map is generated



Tower scene 3D model



View 1



View 2

Generation of 3D map models for the visually impaired

- A camera tracks a real map model of an area (indoor or outdoor).
- The equivalent 3D virtual model is produced in real time and fed onto the system for haptic interaction.
- The visually impaired examine the 3D scene using either the Phantom or the CyberGrasp haptic device.

Generation of 3D map models for the visually impaired

Generation of 3D map models for the visually impaired

- 90% of the users succeeded in identifying the area, while 95% characterized the test as useful or very useful.
- Users did not face any usability difficulty, especially when they were introduced with a short explanation of the technology and after running some exercises to practice the new software.

Conclusions

- A system is developed, which extracts 3D information from a monoscopic video and generates a 3D model suitable for haptic interaction.
- Very efficient if information about the structure of the scene is known a priori.
- **Grand challenge:** Dynamic real time haptic interaction with video/animation.

THANK YOU!

INFORMATICS & TELEMATICS INSTITUTE

1st km. Thermi-Panorama Road

PO BOX 361, 57001 THERMI THESSALONIKI, GREECE

TEL: +30 2310 464160 FAX: +30 2310 464164

http://www.iti.gr

Konstantinos Moustakas

Email: moustak@iti.gr

Dr. Dimitrios Tzovaras

Email: tzovaras@iti.gr