3D Acquisition System for the Measurement of the Foot

Cédric Marchessoux, Noël Richard, Christine Fernandez
IRCOM SIC Laboratory, Poitiers University
BP 30170 Futuroscope, France
marchessoux@sic.univ-poitiers.fr

Abstract

This study concerns the conception of a 3D scanner dedicated to the human foot for its 3D reconstruction for biomedical applications. The view of this article has been to show the research made since two years. This system must allow to extract metrology of human foot. Indeed, the precision of 1mm has to allow the quality of the manufacture of adapted prostheses to the patient feet. This article is directed towards two axes : on the one hand, the explanation of the encountered problems and objectives defining the specifications and on the other hand, the solutions brought and choices for the acquisition system. The main goal of the system is to be able to measure various pathologies of met foot. Therefore, the solutions brought must make it possible to obtain a completely automatic system. They include the choice of the extraction method of points of the foot surface and the modelling to rebuild surface starting from the cloud of points. It must be also calibrated to supply the metrology of the foot. Moreover, the system is built with several color cameras. These various choices allow to extract the metrology of the foot in the desired resolution.

1. Introduction

For several years, thanks to the development of new technologies, a significant number of 3D scanners have been designed above all, for medical or industrial applications. A few has been developed for the digitalization of the human foot. Nevertheless the price remains very expensive.

The aim of this study is to conceive and realize a cheap laboratory prototype scanner dedicated to the human foot. Moreover, this last has to provide the metrology[1] of the foot in order to help the orthopaedic shoe conception.

1.1. Problematic

The orthopaedic shoe manufacturing depends on the foot measures. These ones are done hand-made by the expert. Consequently, there is an incertitude of foot measurement. The developed system has to provide all measurements requiring to design the orthopaedic shoe. The digitalization of the human foot is a delicate problem because of the specifications of our system . In fact, this last has to handle the majority of foot pathologies. It is our main problem but it is not the only one. Moreover, the acquisition time must be inferior to few seconds for each foot. Besides, the system has to be very user-friendly. The wished precision is inferior to 1 mm. The principal consequence that the camera calibration phase and scanning method have to be accurate.

1.2. Related works

Several types of scanning systems exist. These lasts are based on different techniques which are not always resulting into images like telemetry. Others systems are based on the vision. Moreover, the vision systems have been developed in two approaches : ”passive” and ”active”.

The passive vision corresponds to a system without lighting which does not add information whereas the active vision corresponds to a system which does it. The lighting is often a structured light obtained by a laser grid or a line. These lightings have to help the detection of points in the image.

In this two vision approaches, several methods have been developed. Nevertheless, there are two main techniques. On the one hand, the triangulation is based on the theodolite theory. An example of a triangulation is explained in [MBW97]. On the other hand, the explanation and the reproduction of the human visual system result into the stereovision. This technic consists in the matching of points present in different images. This last seems very interesting because it exists several alternatives like shape from silhouette, shape from shading, shape from motion, depth from focus or shape from texture described in [Rou99].

A color image of a human foot contains all information in order to use stereovision alternatives like texture, color

[1] A set of measures allowing to represent accurately the foot surface.

or shape, geometry or photometric properties. The aim of our scanning system is to extract a whole of points of foot surface called : the cloud of points.

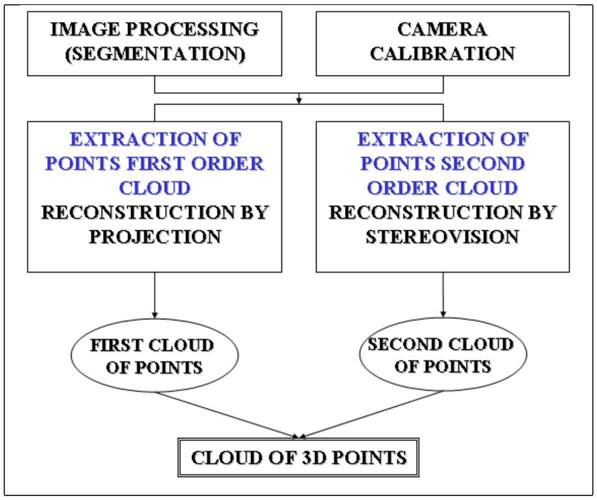


Figure 1: *Synoptic scheme of the acquisition system.*

2. Methodology

The methodology is representing by the synoptic scheme given by the figure 1. In order to handle both simple pathologies and more important ones, the cloud points number is increased thanks to a combination of two methods. The first method is very simple and is presented in a next section. This method is based on different projections. The second one has to increase the points cloud obtained by the first, because of the reasons explained in a previous section, the technique chosen is the stereovision. However, before to put these techniques in place, each camera of the system has to be calibrated. Moreover, each image obtained has to be segmented by an image processing. The different stages of the synoptic scheme are described in the next sections.

2.1. Acquisition system

The acquisition system is composed of digital color cameras which are disposed around the foot. Five color digital images of the foot are obtained.

All the images are obtained with the same conditions. Moreover, all the study images are acquired with the same regulations (focal, object-camera distance, lighting). In order to compare correctly the images between them, these constant parameters allow to measure the reproducibility.

The combination of lighting conditions and background color has to ease the image processing : segmentation step. This stage is explained in a next section.

2.2. Calibration

An acquisition system composed of cameras has to be calibrated. This stage allows to compute the intrinsic and extrinsic parameters (focal length, camera position, rotation angles and translational components). The calibration method used depends on the camera model chosen. R.Y. Tsai proposed a model [Tsa87] and R. Willson has improved this model [Wil94].

Our work is based on researches done by R. Horaud and O. Monga [HM93]. The camera model is the pinhole camera. Their method consists in the determination of a matrix M composed of 12 elements which allow to compute all camera parameters. The matrix represents the transformation method between the scene or world and the image.

2.2.1. Transformation between the world and the image

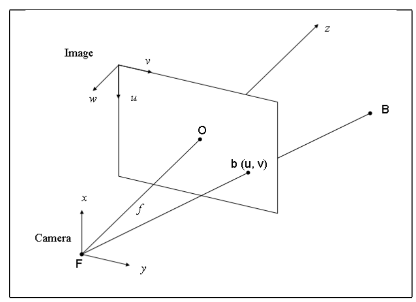


Figure 2: *The representation of acquisition system world.*

This transformation is given by the equation 1.

$$\mathrm{M} = \begin{pmatrix} \alpha_u & 0 & u_0 & 0 \\ 0 & \alpha_v & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (1)$$

A point B of the world with the coordinates X_B, Y_B and Z_B, is projected onto the image in a point b with the coordinates in the camera reference : x_b, y_b and z_b. The b coordinates in the image are : u and v (figure 2).

α_u et α_v represent the horizontal and vertical pixels size. If α_u is different from α_v, the pixels are neither square nor rectangular. In the most of usual cameras, the value of the ratio $\frac{\alpha_u}{\alpha_v}$ is constant and equals to $0,7$.

u_0 et v_0 represent the projection of the optical center onto the image.

The standard intrinsic parameters correspond to the focal (f) and to the coordinates of the center of the image (u_0,v_0). If another projection type like projective is used, in others words if the camera model is different from the chosen model, others intrinsic parameters can be added in order to take account the distortion.

The extrinsic parameters characterize the transformation between the world and the camera. This transformation is characterized by a rotation given by the matrix R ($r_{11}, r_{12}, r_{13}, r_{21}, r_{22}, r_{23}, r_{31}, r_{32}$ and r_{33}) and a translation given by the vector T (t_x,t_y and t_z).

The transformation from a point B to a point b in the image can be wrote as :

$$\begin{pmatrix} su \\ sv \\ s \end{pmatrix} = \begin{pmatrix} m_{11} & m_{12} & m_{13} & m_{24} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \quad (2)$$

A matrix composed of 12 elements has to be determined thanks to the calibration step :

$$\begin{cases} r_3 = m_3 \\ u_0 = m_1.m_3 \\ v_0 = m_2.m_3 \\ \alpha_u = - \parallel m_1 \wedge m_3 \parallel \\ \alpha_v = \parallel m_2 \wedge m_3 \parallel \\ r_1 = \frac{1}{\alpha_u}(m_1 - u_0 m_3) \\ r_2 = \frac{1}{\alpha_v}(m_2 - v_0 m_3) \\ t_x = \frac{1}{\alpha_u}(m_{14} - u_0 m_{34}) \\ t_y = \frac{1}{\alpha_v}(m_{24} - v_0 m_{34}) \\ t_z = m_{34} \end{cases} \quad (3)$$

The calibration consists in the use of three-dimensional points with coordinates known for the detection of these points in the image in order to compute their coordinates in the image. Thanks to the equation 2, the parameters of the matrix M can be computed with the injection of the coordinates of points known. In this way, the parameter m_{34} is approximated to 1 and the method to resolve the linear system is the LU decomposition. Thanks to the equation 3 and the matrix M, the intrinsic and extrinsic parameters can be obtained for each camera. The position of three-dimensional points in the world reference (X_i, Y_i, Z_i) and in the image reference (u_i, v_i) are obtained by the use of a calibration pattern. On the one hand, the position of three-dimensional points in the world reference are known and on the other hand, their position in the image reference is computed.

2.2.2. Calibration pattern

The choice of the calibration is a delicate and a very important problem. Moreover, it depends on the method envisaged.

A calibration technique developed by J.M Lavest in [LVD99], which allows to free oneself from a calibration pattern composed of a whole of three-dimensional points with coordinates known. Nevertheless, this method can not be used for our system. In fact, the system has to be as user-friendly as possible. For the technique described in [LVD99], several images of the pattern in different positions are acquired for each camera.

Therefore, the pattern used is shown by the figure 3. The whole of three-dimensional points with coordinates known is given by the intersections of lines. Nevertheless, the realization of this pattern is very difficult. Moreover, the positions (u_i, v_i) of of the points in the image have to be computed. An image processing has to be applied in order to obtain the positions in the image. The method used is the Hough transform.

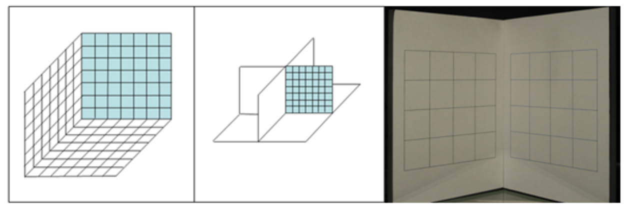


Figure 3: *The calibration pattern used.*

2.2.3. Hough transform

The Hough transform allows to detect lines in an image in order to extract specific points corresponding to the intersections [Mai85]. The image of the lines is the result of the photography of the calibration pattern. The detection of these points in the image allows to calibrate the camera. The Hough transform represents a point by a line in the Hough space.This last has for Y-axis the slope (m) and for X-axis original ordinate (p), $y = mx + p$. In the continuum space, a line with the equation $y = m_1 x + p_1$, is represented by an only point (m_1, p_1) in the Hough space (m, p). In the discrete space, a line is considered as a whole of points. In our case, the images obtained by the digital cameras are discrete. In this case, each line is a whole of points (x, y) whereby in the Hough space, an important number of lines cross. The presence of a line in a discrete image can be considered as a whole of points aligned. The Hough transform of this image computes for each point of the line all the couples (m_i, p_i) corresponding to the lines crossing by these points. There is normally a specific couple with coordinates (m_a, p_a) with $i = a$, which is common to each point of the line in the discrete image. In order to compute the Hough transform, this specific couple (m_a, p_a) has to be found. Moreover, vertical lines are present in the image of the calibration pattern, it would be seem wise to be in polar coordinates (ρ, θ).

The algorithm can be described like this :

- Threshold or grid detection in the pattern.
- Creation of an Hough table with 2 dimensions.
  - Declaration of a table : Hough [ρ, θ]
  - Initialization of Hough [ρ, θ] = "0"
  - $\theta \in [0, \theta_{max}]$, $\theta_{max} = \pi$, $\rho \in [0, \rho_{max}]$
  - $\rho_{max} = \frac{\sqrt{rows^2+columns^2}}{2}$
- Path in the image for the incrementation of the table Hough.
- Search for maximums in the table Hough.
- Reconstruction of the lines in the image.

Eventually, the computation of intersections lines gives the position (u_i, v_i) of points in the image.

2.3. Image processing : segmentation

The stage of calibration is now completely explained. The next stage consists in the description of the process done on images obtained by the acquisition system (figure 1).

In order to obtain a 3D reconstruction of the patient foot and because of 3D techniques envisaged, it is necessary to extract and reproduce accurately its shape. Consequently, a segmentation stage is unavoidable and requires a specific algorithm given that standard algorithms do not provide enough satisfactory results. Moreover, the environment definition (background color, lighting conditions) depends on the envisaged method. This method is described in [MLRF03]. The first step consists in the study of different kinds of feet and then to try several backgrounds in order to find the optimal combination between the acquisition background and the lighting conditions.

Now, the first stage of the synoptic scheme given by the figure 1 has been described.

2.4. 3D reconstruction

2.4.1. Reconstruction by projections

A method called : reconstruction surface by projection is used. This last is described in [MRF02]. Thanks to the acquisition system, five profiles are available to reconstruct the foot. In a first time, we thought that we were able to find several extremum points in each image. Each extremum point is present in several images. Consequently it was easy to find the third point coordinate. For all pathologies, we are able to extract four extremum points in each images. The relative length, width and height of the foot are computed and allow to construct a first volume in which the foot is included (image on the upper left-hand of the figure. This volume is a parallelepiped. Granted that five orthogonal profiles can be used. They can be projected in the volume in which the foot is included (equation 4).

$$I : image, \quad V : volume, \quad Pi : cloud \quad of \quad points,$$
$$\mathrm{R}^{\mathbf{2}} \longrightarrow \mathrm{R}^{\mathbf{3}}, \mathrm{I} \longrightarrow \mathrm{V}' = \mathrm{V} * \mathrm{I},$$
$$V(i,j,k) = I(i,j) * V(i,j,k) \tag{4}$$

The skin surface must be extracted. Every point is studied in a related neighborhood 6, if one of the nearby points is at off, the point is called belonging to the surface (equation 5), only surface points are retained.

$$Pi \in [0,1], \quad i \in S \quad if \quad and \quad only \quad yes \quad \sum_{i \in \mathbf{R}} Pi < 7,$$
$$such \quad as \quad \mathrm{R} : \mathrm{d(i,j)} = 1 \tag{5}$$

2.4.2. Reconstruction by stereovision

The segmentation stage of the 5 images obtained by the acquisition system has to extract the foot pixels from the background pixels and ease the matching between the foot pixels. Thanks to the calibration stage and the equation 6, the matrix M for each camera is known as well as the coordinates of whole of three-dimensional points of the pattern calibration (X_i, Y_i, Z_i) can be computed. The coordinates of these points in the image reference (u_i, v_i) are obtained by the Hough transform.

$$\begin{cases} u' = \frac{m'_{11}X+m'_{12}Y+m'_{13}Z+m'_{14}}{m'_{31}X+m'_{32}Y+m'_{33}Z+m'_{34}} \\ \\ v' = \frac{m'_{21}X+m'_{22}Y+m'_{23}Z+m'_{24}}{m'_{31}X+m'_{32}Y+m'_{33}Z+m'_{34}} \\ \\ u" = \frac{m"_{11}X+m"_{12}Y+m"_{13}Z+m"_{14}}{m"_{31}X+m"_{32}Y+m"_{33}Z+m"_{34}} \\ \\ v" = \frac{m"_{21}X+m"_{22}Y+m"_{23}Z+m"_{24}}{m"_{31}X+m"_{32}Y+m"_{33}Z+m"_{34}} \end{cases} \tag{6}$$

2.4.3. Modelling method

Several surface modelling methods have been developed. The cloud of points obtained by the method is non organized. So, the modelling surface is based on the thesis work of H. Hoppe [Hop94]. His researches allow to work with unorganized cloud of points. The algorithm takes a list of points which are on the surface of the foot. A signed measure of the distance to the surface is computed and sampled on a regular grid. The grid can then be contoured at zero to extract the surface. The default values for neighborhood size and sample spacing should give reasonable results for most uses but can be set if desired.

The aim of our study is to find the foot metrology. Consequently, the foot reconstruction is not the first goal but it is required with the view to visualize our results and validate our method. For the study, the visualization toolkit is used. It's a free available C++ class library for 3D graphics and visualization [SML97].

3. Experiment

In order to validate the methodology described previously, we have experimented the calibration method, the segmentation method on images obtained from the developed acquisition system, and the first method so as to extract a cloud of points.

3.1. Calibration

For the first time, the radial lens distortion was not taken into account [Tsa87]. The quality of the determination of intrinsic and extrinsic parameters depends on the precision of the lines detection in the calibration pattern. The figure 4 shows the result of the Hough transformed used for the lines detection. The lines detected are in a white color. The experiment has shown that longer the segment of line to detect is, better the detection is. Nevertheless, if distortion is present in the image, this last is not taking in account. The solution consists in reducing the distortions by placing the pattern in order to have its projection in the center of the image. Then, the matrix M for each camera is computed and the equation 6 is used so as to do the stereovision.

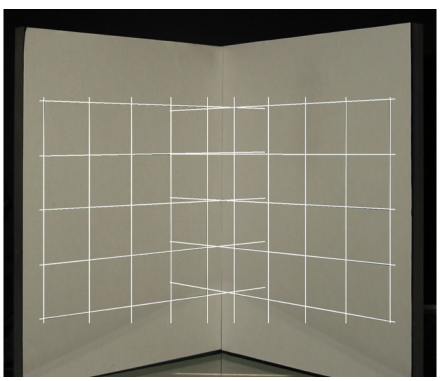


Figure 4: *Lines detection in the pattern in "white color".*

3.2. Image processing : segmentation

The comparison of the *RGB* components of several feet skin has shown a weak occurrence of the blue component. A colorimetric study was done using a set of backgrounds with different referenced blue colors (Pantone®). The study was done in the *HSI* color space allowing the decorrelation of the *RGB* components. A set of images was acquired while combining several feet with different types and different blue backgrounds. So, the study of the hue component obtained shows a perfect uniformity with a sufficiently high saturation component which allows a first segmentation.

In the end, the background color corresponds to the Pantone® *U542* reference given by the figure 5-a. The image obtained for a foot with this background is shown by the figure 5-b, its hue by the figure 5-c. Nevertheless, an additional process is applied on the image thanks to the combination of 2 morphological operators (aperture and geodesic dilatation) described in details in [MLRF03]. Eventually, the result is given by the figure 5-d.

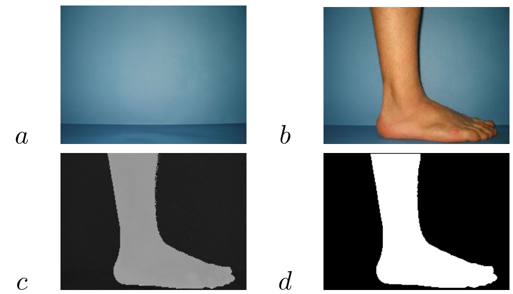


Figure 5: *a : blue background of Pantone® U542 reference, b : background and a foot, c : its hue, d : image obtained by the segmentation method*

The segmentation and calibration steps allow to put into practice the second stage of the synoptic scheme given by the figure 1 and which consists in the extraction of foot surface points.

3.3. Reconstruction method

The reconstruction method has been envisaged thanks to the combination of several ones in order to handle simple and more important pathologies. For the first studies, the human foot was not used. Therefore the researches have been made on a retsina foot made by casting shown by the figure 6.

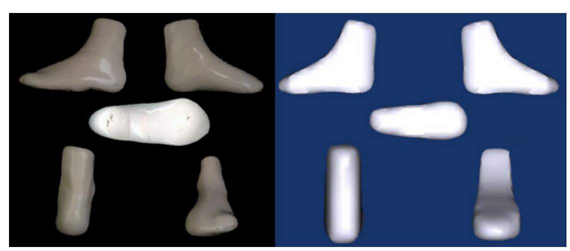


Figure 6: *Example of a surface Reconstruction.*

A summary of the first method reconstruction is given by the figure 7. The first cloud of points is obtained. However by comparing the real object and its reconstruction, some projections have killed too many points of the foot (figure 6). This phenomenon is due to the perspective projection. Indeed, too many points are abolished by the projection of two images aside of the foot. Nevertheless, this method is used for the foot bottom image. This last is obtained thanks to a linear camera. Therefore, only one view is available and shown by the figure 8-a.

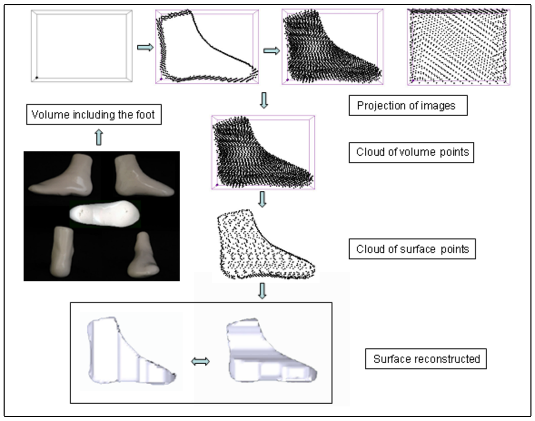
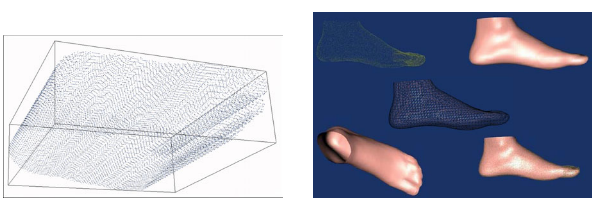


Figure 7: *Surface Reconstruction Method.*

3.4. Modelling method

The modelling method has been experimented with an unorganized cloud of points of a foot obtained with a scanner of high resolution. This last is composed of 20000 points. The result is shown by the figure 8-b.



a b

Figure 8: *a : first cloud of points, b : surface modelling.*

4. Conclusion and future works

The synoptic scheme represented by the figure 1 has been completely described. Nevertheless, the second method to obtain a second cloud of points is in realization. This part corresponds to the future works. The only problem is the matching of points between the images but this step depends on the quality of the segmentation method used. Moreover, the stereovision depends of the method chosen in order to match the points. Consequently this last problem is a delicate problem. The final cloud of points corresponds to the fusion of the first one and the second one.

References

[HM93] R. Horaud and O. Monga. *Vision par ordinateur, outils fondamentaux.* Traité des nouvelles technologies, série informatique. Hermès edition, 1993.

[Hop94] H. Hoppe. *Surface reconstruction from unorganized points.* PhD thesis, Department of Computer Science and Engineering, University of Washington, 1994.

[LVD99] J.M. Lavest, M. Viala, and M. Dhome. Quelle précision pour une mire d'étalonnage. *Traitement du Signal*, 16(3), 1999.

[Mai85] H. Maitre. Un panorama de la transformée de hough. *Traitement du signal*, 2(4):305–317, 1985.

[MBW97] L. Matthies, T. Balch, and B. Wilcox. Fast optical hazard detection for planetary rovers using multiple spot laser triangulation. In *IEEE Conf. on Robotics and Auto.*, 1997.

[MLRF03] C. Marchessoux, M.C. Larabi, N. Richard, and C. Fernandez. Skin segmentation in the *HSI* color space dedicated for medical imaging. In *IS&T/SPIE Digital Photography, PICS*, Rochester, NY, may 2003.

[MRF02] C. Marchessoux, N. Richard, and C. Fernandez. Description of simple method in 3d reconstruction in medical imaging. In *IEEE 1st symposium on 3D Data Processing Visualization Transmission*, Padova, Italy, june 2002.

[Rou99] D. Roussel. *Reconstruction de courbes et de surfaces 3d en stéréo-acquisition.* PhD thesis, Université de Paris XI, january 1999.

[SML97] W.J. Shroeder, K.M . Martin, and B. Lorensen. *The visualization toolkit : an object-oriented approach to 3D graphics.* Prentice Hall-PTR, second edition, 1997.

[Tsa87] R.Y. Tsai. A versatile camera calibration technique for high-accuracy 3d machine vision metrology using off-the-shelf tv cameras and lenses. *IEEE journal of robotics and automation*, RA-3(4):323–343, august 1987.

[Wil94] R. Willson. *Modeling and calibration of zoom lenses.* PhD thesis, Department of Electrical and Computer Engineering, Canergie Mellon University, Pittsburg, USA, january 1994.