

# Tracking Scissors using reflective Lines

Katharina Pentenrieder

16<sup>th</sup> July 2004

## Introduction

Tracking plays an important role in Augmented Reality (AR) applications. One possibility to find and track an object is to attach a fixed arrangement of balls to it and use them to determine its three-dimensional position. Medical instruments like scissors could also be tracked that way but the ball arrangement can be disturbing while working with the instrument. For medical scissors another approach is possible to identify their position and orientation using only the legs of the instrument.

## Proceeding

The aim of this diploma thesis is to track medical scissors using reflective lines. The legs of the instrument are covered by retroreflective material and the pair of scissors is exposed to infrared flashlights. Thus images can be generated where the legs are easily recognizable.

The project can be divided into three major parts of work:

- finding the lines in the infrared images using Hough Transformation
- tracking the lines in a series of images
- determining the three-dimensional position of the instrument using images of two stereo cameras

The order of the latter two parts is not fixed. Still, the question needs to be solved which order is better for our case.

## Hough Transformation

Hough transformation is a very efficient procedure for detecting lines and curves in pictures. Instead of testing the lines formed by all the pairs of figure points the points are transformed into a parameter space. There they vote for all the lines they lie on. Finally the present lines in

the picture should be the ones which gained a lot of votes.

## Theory

Mathematically the transformation can be represented as follows:

$$\rho = x \cdot \cos(\theta) + y \cdot \sin(\theta)$$

Each figure point  $(x, y)$  is transformed into a sinusoidal curve in the parameter plane. Collinear points are represented by curves with a common point of intersection (see Figure 1). So present

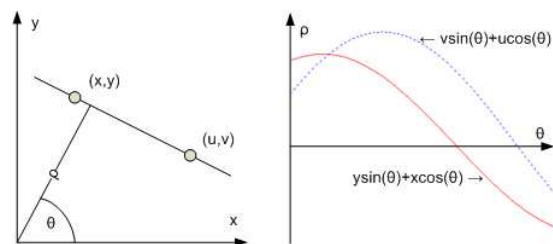


Figure 1: Hough Transformation

lines in the picture should be found by counting the number of intersecting curves (the votes) for each point in the parameter plane, for each pair  $(\rho, \theta)$ . This is done in a so called accumulator array. For the implementation of the hough transformation the parameter plane needs to be discretised with a convenient grid. The accumulator array then represents this grid and collects the votes for each cell. Figure 2 shows an image representation of such an array.

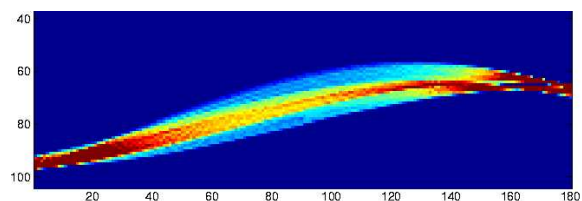


Figure 2: Accumulator Array

## Practical Problems and Possibilities

Literature on the Hough Transformation mentions several general practical problems and for this project an additional specific problem arises:

### Discretisation:

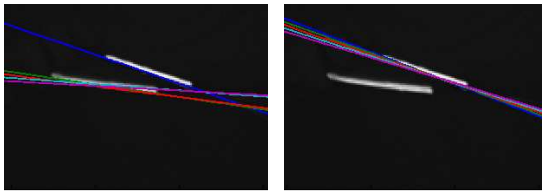
Finding an appropriate grid size is very difficult. Too coarse grids can lead to falsely large votes because quite different lines correspond to the same grid cell. And if the grid is too fine, present lines might not be found because votes end up in different cells.

### Difficulties with noise:

An advantage of the Hough Transformation is that it connects widely separated tokens if they are nearly collinear. But this can also be a weakness. Phantom lines might be detected in regions with uniformly distributed figure points.

### Different line intensities:

Tests with infrared images of the pair of scissors showed the problem of different line intensities. If the number of figure points belonging to the legs of the instrument differ considerably, the Hough Transformation results in either only lines along the "clearer" leg or in too many lines along both legs (see Figure 3).



grey value threshold: 150    grey value threshold: 200

Figure 3: Different line intensities

But what can be done to face these problems? The actual implementation contains various attempts which shall be discussed next:

### Variable cell size:

Variable parameter discretisation ( $\rho$  and  $\theta$ ) allows arbitrary adjustments of the cell size.

### Variable grey value threshold:

Additionally the threshold for grey values can be modified to influence the number of figure points which are included in the Hough Transformation. The effect on the results can be seen in Figure 3.

### Variable back transformation threshold:

This threshold cannot ameliorate the results of the transformation but it is an indicator for the quality of the other chosen variables. Which accumulator cells have votes high enough? These

pairs  $(\rho, \theta)$  shall be transformed back to the original plane (represented as colored lines in Figure 3).

### Grouping of parameter pairs:

Usually a larger number of parameter pairs needs to be considered for back transformation to find all the lines in the original image. But this means that a lot of computed lines will stand for the same image line. To reduce the number of computed lines they can be grouped together by similar values of  $\rho$ .

### Splitted Hough Transformation

All the efforts described above are not sufficient to allow the accurate discovery of both legs of the pair of scissors. Therefore the actual implementation searches for each leg separately. First the whole image is considered in the Hough Transformation and the maximum value in the accumulator array is chosen as representation of the first leg. After back transformation the discovered line is eliminated and another Hough Transformation is done with the resulting image. Again the maximum value is chosen to represent the second (probably less clear) leg of the pair of scissors. This approach allows a result with appropriate accuracy (see Figure 4).

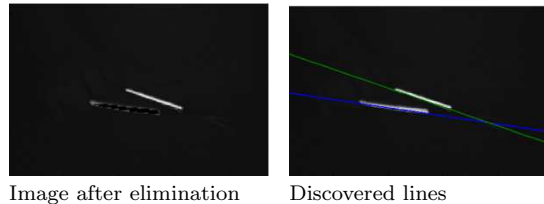


Image after elimination    Discovered lines

Figure 4: Splitted Hough Transformation

## Linetracking

Linetracking can be done either directly after the Hough Transformation or it is postponed until the 3D position of the object is determined. The alternative that will be chosen here will depend on the data provided by the Hough Transformation and the possibilities of calculation in either case.

In the former a method of tracking could be the following (see Blake and Isard [1998]):

A template vector of grey values represents the profile of the line which shall be tracked. It can be calculated from the first image in the series. This template is now moved along the line normal in the next image until an optimum of conformance with the image data is found. Figure 5 shows the principal configuration.

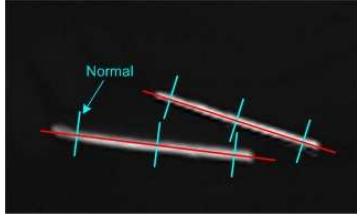


Figure 5: Configuration for line tracking

For a template  $T$  and an normal image  $I$  the point of maximum conformance can be determined by minimizing the mathematical correlation (**correlation matching**):

$$\int_{x=0}^{\delta} I(x - x')T(x)dx \text{ with } 0 \leq x \leq \delta$$

### 3D-Positioning

Given the data from two cameras the pair of scissors - discovered in the images via Hough Transformation - shall now be placed into the 3D-space. To compute its position several steps are necessary. Here only an idea of the theory behind is given, for further details see Zisserman and Hartley [2003].

### Camera Model and camera matrix

The first step is to express the relationship between world points and image points:

The projection of a point in homogenous world coordinates  $\mathbf{X} = (X, Y, Z, 1)^T$  onto the image plane (image point  $\mathbf{x} = (x, y, z)^T$ ) can be represented as a matrix-vector-multiplication

$$\mathbf{x} = P\mathbf{X}$$

where  $P$  is the camera matrix.  $P$  can be computed from point correspondences  $\mathbf{X}_i \leftrightarrow \mathbf{x}_i$  of world coordinate points and camera coordinate points.

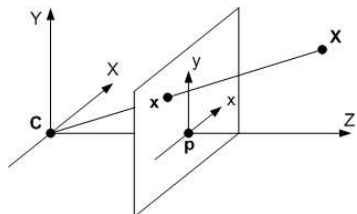


Figure 6: Camera Model

### Epipolar Geometry and fundamental matrix

Epipolar geometry considers two image representations of the world. The relationship between the two images of a world point shall be expressed next:

The world point  $\mathbf{X}$  is represented as  $\mathbf{x}$  and as  $\mathbf{x}'$  on two image planes.  $\mathbf{C}$  and  $\mathbf{C}'$  are the projection centres, the cameras. The line connecting them is called base line and the intersections of the base line with the image planes are called epipoles ( $\mathbf{e}$  and  $\mathbf{e}'$ ).  $l'$ , the epipolar line, is the intersection of the epipolar plane  $\pi$  and the image plane of  $\mathbf{C}'$ . Figure 7 (a) shows the principle structure. For the case of tracking scissors two lines need to be positioned in 3D-space. In epipolar geometry such a line is the intersection of two planes as can be seen in Figure 7 (b).

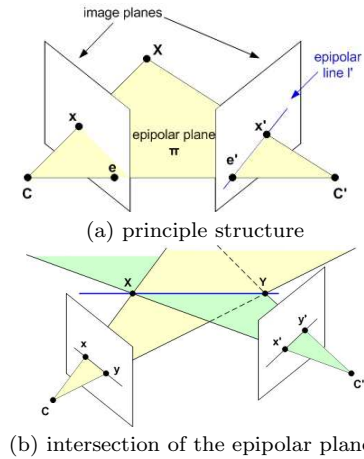


Figure 7: Epipolar geometry

The algebraic representation of the epipolar geometry is the fundamental matrix  $F$ . The basic idea is that there exists a mapping between an image point  $\mathbf{x}$  and the epipolar line on the second image plane  $l'$ . Geometrically  $l'$  connects the epipolar point  $\mathbf{e}'$  and the image point  $\mathbf{x}'$  and there exists a mapping from  $\mathbf{x}$  to  $\mathbf{x}'$ . It can be shown that

$$l' = F\mathbf{x}$$

For the fundamental matrix  $F$  the following condition holds for any pair of correspondences  $\mathbf{x}_i \leftrightarrow \mathbf{x}'_i$ :

$$\mathbf{x}'^T F \mathbf{x} = 0.$$

### 3D-Reconstruction

With the basis described in the previous sections the projective 3D-reconstruction can be done as

follows:

Given enough point correspondences  $\mathbf{x}_i \leftrightarrow \mathbf{x}'_i$  the fundamental matrix  $F$  can be calculated solving the linear equation system

$$\mathbf{x}'_i{}^T F \mathbf{x}_i = 0.$$

As a second step the camera matrices  $P$  and  $P'$  can be computed. Finally for each correspondence  $\mathbf{x}_i \leftrightarrow \mathbf{x}'_i$  the point  $\mathbf{X}_i$  in space is retrieved via Triangulation using the equations

$$\mathbf{x}_i = P\mathbf{X}_i, \quad \mathbf{x}'_i = P'\mathbf{X}_i.$$

## References

- Andrew Blake and Michael Isard. *Active Contours*. Springer Verlag, 1998.
- Richard O. Duda and Peter E. Hart. *Use of the Hough Transformation to detect lines and curves in pictures*. Communication of the ACM (Volume 15, Issue 1), January 1972.
- D. Forsyth and J. Ponce. *Computer Vision - A modern Approach*. Prentice Hall, 2003.
- A. Zisserman and R. Hartley. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2003.