

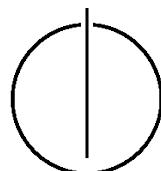
Technischen Universität München
Computational Science and Engineering (Int. Master's Program)

Master's Thesis

Implementation of Contracting Curve Density
Algorithm for Applications in Personal Robotics

Shulei Zhu

1st examiner: Beetz, Michael; Prof.
2nd examiner: Hans-Joachim, Bungartz; Prof. Dr.
Assistant advisor: M.Sc. Pangercic, Dejan
Thesis handed in on: April 1, 2011



I hereby declare that this thesis is entirely the result of my own work except where otherwise indicated. I have only used the resources given in the list of references.

München, den 12. April 2011

Shulei Zhu

Acknowledgments

If someone contributed to the thesis... might be good to thank them here.

Abstract

An abstracts abstracts the thesis!

Contents

Acknowledgements	v
Abstract	vii
1 Introduction	1
1.1 The CCD algorithm and curve fitting problem	1
1.2 Motivation of the CCD Approach	2
1.3 The Contracting Curve Density (CCD) Approach	3
1.3.1 Math description	3
1.3.2 Possible problems	4
1.3.3 Sketch of the CCD algorithm	4
1.4 Overview of the Thesis	5
2 Related work	7
2.1 2-Dimensional and 3-Dimensional deformable model	7
2.2 Statistical models	8
3 Software and Hardware Infrastructure	11
3.1 The Robot Operation System (ROS)	11
3.2 PR2	11
4 Shape-space Models and B-Spline Curves	13
4.1 B-Spline Function	13
4.1.1 Uniform quadratic B-Spline	14
4.1.2 Uniform cubic B-Spline	14
4.2 Parametric B-spline curve	15
4.2.1 Norm for B-spline curves	15
4.3 Shape-space models	15
5 The CCD Algorithm	17
5.1 Pre-processing and model parametrization	17
5.1.1 Pre-processing of input data	17
5.1.2 Contour initialization	17
5.1.3 Model parametrization	18
5.2 Local statistics	19
5.2.1 Collecting local information	19
5.2.2 Learning local statistics	20
5.3 Refine parameters	22
5.4 The CCD tracker	24

Contents

5.5	Summary of the CCD algorithm and its variant	25
5.6	Efficiency and complexity	25
6	Experiments and Results	27
6.1	Contour convergence on still image	27
6.1.1	Segmenation of a ball	28
6.1.2	Segmenation of a 3-D obejct	28
6.1.3	Segmenation of a transparent obejct	32
6.1.4	Fitting a rigid wire frame	32
6.2	Tracking initialization from sift features	33
6.3	Tracking initialization from 3D point cloud	35
7	Conclusion and Future Work	39
	Bibliography	41

1 Introduction

In the field of computer vision and pattern recognition, determining a model from a finite set of samples or a complex image without prior knowledge is an ill-posed problem. Due to the the ill-posedness, in the sense that a unique model may not exist, models specifying a priori knowledge play an very important role in solving many important computer vision problems, such as object tracking, shape recognition, pose estimation segmentation, edge detection, stereo matching and 3-D reconstruction.

Some parametric curve models, also known as deformable models, deformable templates, snakes, or active contours, have been developed to guide the image interpretation process towards particularly likely interpretations(AC 53). These parametric curve models, namely prior knowledge have two main outstanding influences in the development of related algorithms. First is the snakes which represented a fundamentally new framework for delineating an object outline in an image. This framework attempts to minimize an energy associated to the current contour as a sum of an internal and external energy(wikipedia). The goal of snakes is to balance the prior knowledge with evidence from an image. The second outstanding influence is in the field of pattern recognition and machine learning where statistical distributions play key roles. There treatment of prior knowledge about shape is put into a probabilistic context. Any shape is regarded as the result of applying some distortion to an ideal prototype shape, and an appropriate distribution governs the nature and extent of the distortion. For both kinds of method, the common aim is to strengthen the visual interpretation of shape via the stabilising influence of prior expectations of the shapes that are likely to be seen.(AC 53) This paper concerns the Contracting Curve Density (CCD) algorithm whose basic step is solving a curve fitting problem in a probabilistic context.

1.1 The CCD algorithm and curve fitting problem

Curve fitting is such a process that makes a practical deformable templates (model) attach to the feature in a image. It in essence involves using algorithms to find the parameter values for a given parametric curve model. A lot of important and challenging computer vision tasks can be formulated as variants of the curve-fitting problem. The nature of curve fitting problem is taken into account from the outset, usually it is quite unnecessary to examine an entire image. Hence, there is no need to organise or group features in the image, and the analysis can be restricted to a relatively narrow ROI (region of interest), so it is usually expected to be computationally cheaper.

The CCD algorithm can be described as follows. Given one or multiple images as input data and a parametric curve model with a prior distribution of the model parameters, through curve fitting process we estimate the model parameters which determine the approximation of the posterior distribution in order to make the curve model best match

the image data. Figure 1 depicts a curve-fitting problem and the corresponding solution obtained by the CCD algorithm.

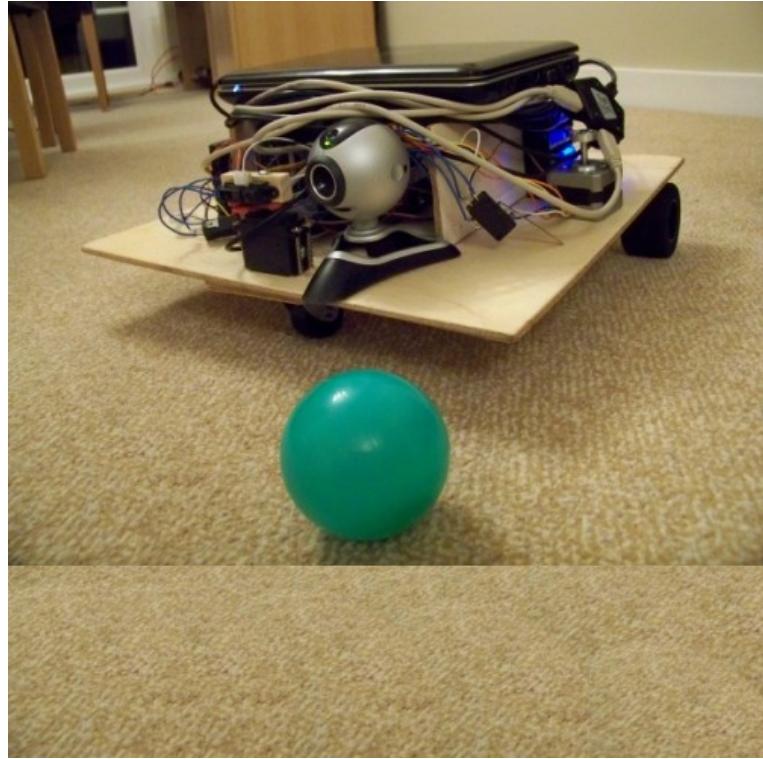


Figure 1.1

1.2 Motivation of the CCD Approach

Curve fitting problem and its variant have a wide range of applications in the field of robotics, medical processing, user interface, surveillance and biometrics (hanek's paper). In order to be widely applicable to practical computer vision problems, following requirements, such as Robustness, accuracy, efficiency, any-time property and versatility, should be considered when a novel approach is designed and implemented. However, in the computer vision community, solving object segmentation and the related object contour tracking problem are always challenging, especially in natural and unconstrained scenes. Due to clutter, shading, texture, and highlights it is very difficult to segment an object from an inhomogeneous background. Furthermore, some physical conditions, such as the illumination or surface properties, will influence the efficiency and stability of related approaches. It is necessary and important to develop a method to determine adequate segmentation criteria in advance or even and a single criteria that is applicable for all parts of an object boundary.

Among the currently available methodologies, the CCD algorithm is an interest one. It is developed and presented in as a state-of-the-art improvement over other advanced techniques such as the condensation algorithm (panin 2006). In the CCD approach, curve

fitting problem is put in the context of probabilistic framework and converted to a regression one in the field of pattern recognition. It is especially helpful to touch the nature of curving fitting problem. By introducing local image statistics, the algorithm can cope with highly inhomogeneous image regions and provide therefore locally adapted criteria for separating the adjacent image regions. The use of blurred curve models as efficient means for iteratively optimizing the posterior density over possible model parameters. These blurred curve models enable the algorithm to trade-off two conflicting objectives, namely heaving a large area of convergence and achieving high accuracy. The CCD algorithm achieves a high level of robustness and subpixel accuracy even in the presence of severe texture, shading, clutter, partial occlusion, and strong changes of illumination.(modify)

1.3 The Contracting Curve Density (CCD) Approach

1.3.1 Math description

In the field of pattern recognition, the key concept is that of uncertainty. In a image, the uncertainty arises both through noise on measurements, as well as through the nature of objects. Probability theory provides a consistent framework for the quantification and manipulation of uncertainty. In this section, we view curving fitting problem from a probabilistic perspective and talk us towards to Bayesian treatment.

We assume a parametric curve model is governed by a prior distribution over the model parameters Φ (usually a D -dimensional vector). For simplicity, let us consider a Gaussian distribution of the form.

$$p(\Phi|\mathbf{m}_\Phi, \Sigma_\Phi) = \mathcal{N}(\Phi|\mathbf{m}_\Phi, \Sigma_\Phi) = \frac{1}{(2\pi)^{D/2}} \frac{1}{|\Sigma_\Phi|^{1/2}} \exp \left\{ -\frac{1}{2} (\Phi - \Sigma_\Phi)^T \Sigma_\Phi^{-1} (\Phi - \Sigma_\Phi) \right\} \quad (1.1)$$

where the D -dimensional vector \mathbf{m}_Φ is called the mean, the $D \times D$ matrix Σ_Φ is called the covariance, and $|\Sigma_\Phi|$ denotes the determinant of Σ_Φ

We now use local image pixels \mathcal{L} as the training data to determine the likelihood function. If the data are assumed to be drawn independently from the distribution, then the likelihood function is given by

$$p(\mathcal{L}|\Phi) = \prod_{n=1}^N \mathcal{N}(l_n|\Phi) \quad (1.2)$$

where n denotes the number of groups of local discrete image pixels. The goal in the curve fitting problem is to be able to determine the posterior distribution for model parameters Φ .

In this paper, by pixel value we denote the vector containing the local single-or multi-channel image data associated to a pixel. In our experiments, we directly use the sensed RGB values as pixel values. However, other types of local features computed in a pre-processing step may also be used, e.g. texture descriptors or color values in other color spaces

Using Bayes' theorem, the posterior distribution for Φ is proportional to the product of the prior distribution and the likelihood function

$$p(\Phi|\mathcal{L}) \propto p(\mathcal{L}|\Phi)p(\Phi|\mathbf{m}_\Phi, \Sigma_\Phi) \quad (1.3)$$

We can now determine $\hat{\Phi}$ by finding the most probable value of Φ given the data, in other words by maximizing the posterior distribution. This technique is called maximum posterior, or simply MAP.

$$\hat{\Phi} = \arg \max_{\Phi} p(\Phi|\mathcal{L}) \quad (1.4)$$

1.3.2 Possible problems

In order to implement the MAP estimation above, we have to address four three questions.

- How to create a parametric curve model as prior knowledge , and at same time determine the distribution? There are a variety of forms active shape models available, such as principally snakes, deformable templates and dynamic contours. For curve fitting problem, deformable template is a good choice to match the image data. Then we have to choose suitable number of parameters to control the strength of prior assumptions. If there are too many parameters in the model, it is easy to control the contour accurately at the expense of complexity and computational cost. Balance among these aspects are non-trivial.
- How can the local statistics in the image evaluated? First, it is necessary to establish the relation between model parameters Φ and the image data \mathcal{L} . This is a especially challenging in the presence of clutter and strong texture if we consider the factors, such as illumination, surface properties , characteristics of the camera and shadow.
- How can the fit be optimized efficiently? Generally MAP and some other nonlinear optimization algorithms may find multiple local maxima and are not guaranteed to find the largest of these maxima.

For the first question, we will use parametric B-Spline curve as our prior knowledge. It is proved that this kind of curve model is flexible enough and can well represent the contour of object. The parametric B-Spline curve will be described in the chapter 3. The second question is especially challenging. In order to simplify the problem, it is necessary to make some assumptions about the image data. we will describe such assumptions in more detail. For the last question, different optimization methods used for curve-fitting will be discussed in this thesis.

1.3.3 Sketch of the CCD algorithm

In this section, the basic steps of the CCD algorithm will be sketched.

Given an input image as training data, we first choose an initial contour for an object or feature which will be fitted or tracked, and some initial values for the means, covariances. Then we alternatively between the following two updates that we shall call the learning local statistics step and the refining parameters step, for reasons that will become apparent shortly. In the step of learning local statistics, for each pixel ν in the vicinity of the expected curve, two sets of local statistics are computed, one set for each side of the curve. The

local statistics are obtained from pixels that are close to pixel nu and most likely lie on the corresponding side of the curve, according to the current estimate of the model parameters. The resulting local statistics represent an expectation of "what the two sides of the curve look like".(Hanek's paper)

The CCD algorithm has some interesting similarities to the Expectation-Maximization (EM) algorithm (Dempster et al., 1977), which is often used for clustering-based image segmentation, a subset of region-based image segmentation (Hermes et al., 2002; Belongie et al., 1998). The first step computes local statistics defining an expectation of the pixel values (E-step). The second step maximizes this expectation (M-step). The CCD algorithm differs mainly by: 1.) using local statistics, 2.) exploiting a curve model and optimizing model parameters rather than pixel classifications.

1.4 Overview of the Thesis

The remainder of this paper is organized as follows. In Chapter 2, related work on PR2, Robot Operation System (ROS), active contour, some curve-fitting ,image segmentation and tracking algorithms will be introduced. Details on shape-space models and parametric B-Spline curves are discussed in chapter 3. In chapter 4 , we give a brief introduction of software and hardware infrastructure used in the implementation of the CCD approach. We will concern the OpenCV library, ROS and PR2, and chapter 5 explain the CCD approach in detail. This will be followed by a description of the application of the CCD approach in robotics. In chapter 7, experiments and results are given. In addition, we evaluate the performance of the CCD algorithm and the CCD tracker in terms of robustness, accuracy, and runtime. In chapter 8, the work of this thesis is concluded with the future work for improvement.

1 Introduction

2 Related work

As mentioned in the previous chapter, the CCD approach introduced in this chapter is a model-based computer vision algorithm. The topics of model-based approach in the field of computer vision have been addressed in numerous research papers. In this chapter we review selected publications related to the topics covered in this thesis and try to reveal the revolution process. The articles have been carefully selected based on the following criteria:

- Articles on 2-Dimensional and 3-Dimensional deformable models, such as snakes and gradient vector flow deformable models[33].
- Articles on applying statistical knowledge to the models

There are possible overlaps in two categories. We will discuss the advantages and limitations of these methods related to the work we aim to solve. In section ?? we will discuss the related work belonging to the first category. This is followed by the introduction of previous work on statistical models.

2.1 2-Dimensional and 3-Dimensional deformable model

Many traditional segmentation methods are effected by the assumption that the images studied in computer vision usually is self-contained, namely, the information need for a successful segmentation could be extracted from the images.

In 1980s, a paradigm named *Active Vison*[1] escape this bind and push the vision process in a more goal-directed fashion. After that, a notably successful departure, the *Snake*, is proposed in a seminar work conducted by Kass. The original paper is [15], which spawned many variations and extensions including the use of Fourier parameterisation[27], incorporation of topologically adaptable model[22] and its application [21], incorporation of discrete dynamic contour model[19]. A realization of snakes using B-splines was developed in [3]. In two dimensions, The snake has a variation named active shape model[5], which is a discrete version of this approach. Gradient Vector Flow[32], or GVF, is an extension developed based on new type of external field. Another paper written by the same authors [33], concerning convergence properties of deformable models. In three dimension, a good deal of research works has been conducted on matching three-dimensional models, both on rigid [12] and deformable[30].

Besides the segmentation, dynamic models, such as snake and its variations, are greatly used in application of object tracking. A real-time tracking system based on deformable models, such as snakes, is developed in [31]. It proved that the active shape and motion estimators are able to deal very effectively with the complex motions of nonrigid objects. Furthermore, the combination of active models and Kalman filter theory is also a popular approach to tracking, some work about this can be found in [25].

[2] is a wonderful book about the topics of geometric and probabilistic models for shapes and their dynamics.

2.2 Statistical models

Pattern recognition theory is a general statistical framework which is important in the study of model-based approaches. This is started from the 1970s and 1980s when a new interpretation of image was proposed in the statistical community. Analyzing the model problems in probabilistic context has two great advantages. The first is that it approaches the nature of the problems, the ranges of shapes are defined by a probability, this provides another viewpoint for curve fitting problem in the field of computer vision. Another advantage is that after we convert a problem in pattern recognition, there are abundant tools and skills to deal with such problem. The CCD approach is a method developed in the probabilistic context.

In [17] and [16], an elegant use of statistical models for the segmentation of medical images is designed. The resulting segmentation system consists of the building of statistical models and the automatic segmentation of new image data sets by restricting elastic deformation of models. The work in [26] and [18] also exploit the prior knowledge from the perspective of probability, furthermore, the statistical shape models enforce the prior probabilities on objects by designing complicated energy function. In this thesis, we assume that the shape priors have a Gaussian form in shape-space. In the case of a norm-squared density over quadratic spline space, the prior is a Gaussian Markov Random Field (MRF)[2], which is used widely for modelling prior distributions for curves[29].

However, defining a prior distribution for shape is only part of the problem. Prior knowledge only control the feature interpretation in a image, a successful segmentation or curve fitting requires to obtaining the posterior distribution, which takes account of what shape is actually likely to be present in a particular image. Inspired by the iterative algorithm k-means[7] and Expectation-Maximization (EM)[6] algorithm, especially the later one, a novel curve fitting method the ccd algorithms is proposed in [11]. Bayesian treatment is used in this algorithm in incorporation of maximum a posteriori probability (MAP) [28] estimate instead of complicated cost function. Not like the EM algorithm, the CCD use local statistics instead of global statistics. Moreover, blurred curve models is proposed as efficient means for iteratively optimizing. The algorithm can be used in object localization and object tracking. As an example of applications of the CCD approach, an efficient, robust and fully automatic real-time system for 3D object pose tracking in image sequences is presented in [23] and [24]. Multiocular contracting curve density algorithm (MOCCD) [9] is an extension of the CCD approach. In the paper, it is integrated into tracking system to tracking human body.

Both the snake and CCD can be used to build naive tracking system. However, they are limited by the performance and stability problems. Several methods achieve a speed-up propagate a Gaussian distribution of the model parameters over time, such as tracking based Kalman filter in [4]. The method is limited by the range of probability distributions they presented. The condensation algorithm (Conditional Density Propagation) [14] is proved to lead to a marked improvement in tracking performance. Another feature of this method is that it only consider the pixels on some perpendiculars. The CCD tracker

[10] uses only pixels on some perpendiculars like the condensation algorithm, but focuses on the vicinity of the contour. This means the CCD tracker can save time and improve performance.

2 Related work

3 Software and Hardware Infrastructure

In this chapter, we will introduce the software and hardware infrastructure used in this thesis. In [3.1](#) we introduce the Robo Operation System.

3.1 The Robot Operation System (ROS)

3.2 PR2

3.3 OpenCV library

4 Shape-space Models and B-Spline Curves

In the CCD algorithms, A parametric curve model is responsible for restricting Region of Interest (ROI) and providing prior knowledge. Throughout this thesis, visual curves are represented in terms of parametric spline curves which is widely applied in computer graphics.

In this chapter, we introduce a deformable 2-D model called shape-space models. In the first section, we start by explaining spline functions and their construction. Details on how parametric curves are constructed from spline functions are discussed in the later section. This forms the basis for the CCD approach.

4.1 B-Spline Function

In the mathematical, every spline function of a given degree, smoothness, and domain partition, can be represented as a linear combination of B-splines of that same degree and smoothness, and over that same partition(wikipedia). In terms of this statement, we can construct a spline curve parametrized by spline functions that are expressed as linear combinations of B-splines.

Given a nondecreasing vector of real values $U = [u_0, \dots, u_{m-1}]$ called knot vector, with u_i is called the knot, a B-spline of degree n is a parametric curve

$$\mathbf{C}(u) = \sum_{i=0}^{m-n-2} P_i B_{i,n}(u), u \in [u_n, u_{m-n-1}] \quad (4.1)$$

where P_i are called control points, $B_{i,n}$ are called basic function. For $n=0,1,\dots,m-2$, the $m-n-1$ basis B-splines of degree n can be defined as

$$\begin{aligned} B_{i,0}(u) &= \begin{cases} 1 & \text{if } u_i \leq u < u_{i+1}, \\ 0 & \text{otherwise} \end{cases}, \quad i = 0, \dots, m-2 \\ B_{i,n}(u) &= \frac{u - u_i}{u_{i+n} - u_i} B_{i,n-1}(u) + \frac{u_{i+n+1} - u}{u_{i+n+1} - u_{i+1}} B_{i+1,n-1}(u), \quad i = 0, \dots, m-n-2. \end{aligned} \quad (4.2)$$

Furthermore, we can write the B-spline compactly in matrix notation as

$$\mathbf{C}(u) = \begin{pmatrix} \mathbf{B}(u)^T & 0 \\ 0 & \mathbf{B}(u)^T \end{pmatrix} \mathbf{P} \quad (4.3)$$

where vector \mathbf{P} denotes the axis components of control points

$$\mathbf{P} = \begin{pmatrix} P_x & P_y \end{pmatrix}^T \quad \text{where} \quad P_x = \begin{pmatrix} P_0^x \\ \vdots \\ \dots \\ P_{N-1}^x \end{pmatrix} \quad (4.4)$$

Where N denotes the number of control points. $\mathbf{B}(u)$ is a vector of basis function

$$\mathbf{B}(u) = (B_0(u), \dots, B_{N-1}(u))^T \quad (4.5)$$

The B-spline is said to be uniform when the knots are equidistant, otherwise non-uniform. In this thesis, for simplicity, we only consider the uniform case.

As two special examples of the B-Spline, we consider the matrix form and derivative of uniform quadratic and cubic B-Spline respectively.

4.1.1 Uniform quadratic B-Spline

Because the knot-vector is equidistant, the blending function can easily be computed. In each segment, the blending function shares the same form as following

$$\mathbf{b}_{i,2}(u) = \begin{cases} \frac{1}{2}u^2 \\ -u^2 + u + \frac{1}{2} \\ \frac{1}{2}(1-u)^2 \end{cases} \quad (4.6)$$

Its derivative is

$$\mathbf{b}'_{i,2}(u) = \begin{cases} u \\ -2u + 1 \\ u - 1 \end{cases} \quad (4.7)$$

The matrix-form of uniform quadratic B-Spline is given by

$$\mathbf{B}_i(u) = [u^2 \quad u \quad 1] \frac{1}{2} \begin{bmatrix} 1 & -2 & 1 \\ -2 & 2 & 0 \\ 1 & 1 & 0 \end{bmatrix} \quad (4.8)$$

4.1.2 Uniform cubic B-Spline

Uniform cubic B-Spline is most commonly used form of B-Spline, The blending function is

$$\mathbf{b}_{i,3}(u) = \begin{cases} \frac{1}{6}(-u^3 + 3u^2 - 3u + 1) \\ \frac{1}{6}(3u^3 - 6u^2 + 4) \\ \frac{1}{6}(-3u^3 + 3u^2 + 3u + 1) \\ \frac{1}{6}u^3 \end{cases} \quad (4.9)$$

Its derivative can be written as

$$\mathbf{b}'_{i,3}(u) = \begin{cases} \frac{1}{2}(-u^2 + 2u - 1) \\ \frac{1}{2}(3u^2 - 4u) \\ \frac{1}{2}(-3u^2 + 2u + 1) \\ \frac{1}{2}u^2 \end{cases} \quad (4.10)$$

Its matrix-form is given by

$$\mathbf{B}_i(u) = [u^3 \ u^2 \ u \ 1] \frac{1}{6} \begin{bmatrix} -1 & 3 & -3 & 1 \\ 3 & -6 & 3 & 0 \\ -3 & 0 & 3 & 0 \\ 1 & 4 & 1 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{p}_{i-1} \\ \mathbf{p}_i \\ \mathbf{p}_{i+1} \\ \mathbf{p}_{i+2} \end{bmatrix} \quad (4.11)$$

4.2 Parametric B-spline curve

In a 2-D image, a curve is composed of many discrete points which are identified by its coordinates. For those points on a curve, their axis components x, y are particular functions of knots value u , known as splines. Hence, we can define a B-spline curve as

$$\mathbf{C}(u) = (x(u), y(u))^T = \mathbf{U}(u)\mathbf{P} \quad (4.12)$$

where $\mathbf{U}(u)$ denotes the matrix mapping control point vector \mathbf{P} to image curve $\mathbf{C}(u)$

$$\mathbf{U}(u) = \begin{pmatrix} \mathbf{B}(u)^T & 0 \\ 0 & \mathbf{B}(u)^T \end{pmatrix} \quad (4.13)$$

4.2.1 Norm for B-spline curves

Norm and inner product of curves are useful in curve approximation. In an image plan, using the Euclidean distance measure, we can define the norm for curves as

$$\|\mathbf{C}\|^2 = \mathbf{C}^T \mathcal{U} \mathbf{C} \quad (4.14)$$

where \mathcal{U} represents the metric matrix for curves, it can be defined as

$$\mathcal{U} = \begin{pmatrix} \mathcal{B}^T & 0 \\ 0 & \mathcal{B}^T \end{pmatrix} \quad (4.15)$$

\mathcal{B} denotes the metric matrix for a given class of B-spline function

$$\mathcal{B} = \frac{1}{L} \int_0^L \mathbf{B}(u) \mathbf{B}(u)^T du \quad (4.16)$$

4.3 Shape-space models

The shape-space \mathcal{S} is a linear parameterisation of the set of allowed deformations of a base curve(AC). From the perspective of mathematics, A shape-space $\mathcal{S} = \mathcal{S}(\mathbf{A}, \mathbf{P})$ is a linear mapping of a shape-space vector $\Phi \in \mathbb{R}^{N_\Phi}$ to a spline vector $\mathbf{C} \in \mathbb{R}^{Np}$:

$$\mathbf{C} = \mathbf{AX} + \mathbf{P} \quad (4.17)$$

where T is a $Np \times N_\Phi$ shape-matrix. The constant P_0 is a deformable template curve. (AC), Usually, the dimension N_Φ of the shape-space vector is clearly smaller than the dimension of the spline-vector Np . This makes the problem simplicity and computationally cheaper.

In this chapter, we consider a important shape space named planar affine. The planar affine has 6 degree of freedom: horizontal translation, vertical translation, rotation, horizontal scale, vertical scale and diagonal scale. The planar affine group can be viewed as the class of all linear transformations that can be applied to a template curve $\mathbf{C}(AC)$.

$$\mathbf{C}(\mathbf{u}) = \mathbf{T} + M\mathbf{C}_0(u) \quad (4.18)$$

where $T = (T_1, T_2)^T$ represents the two-dimensional translation vector, and M is 2×2 matrix to control the rotation and scaling, so that T and M together can represent the 6 degree of freedom of planar Affine space. We can write shape-matrix \mathbf{A} as

$$A = \begin{bmatrix} 1 & 0 & P_0^x & 0 & 0 & P_0^y \\ 0 & 1 & 0 & P_0^y & P_0^x & 0 \end{bmatrix} \quad (4.19)$$

The model parameters can be interpreted as

$$\Phi = (T_1, T_2, M_{11} - 1, M_{22} - 1, M_{21}, M_{12}) \quad (4.20)$$

The planar affine shape-space discussed above can be extended for three-dimensional affine shape-space, which is more suitable for a non-planar object. The new space-model needs 8 degrees of freedom to make up of the six-parameter planar affine space and a two-parameter extension. The transformation matrix for the three-dimensional case can be written as:

$$A = \begin{bmatrix} 1 & 0 & P_0^x & 0 & 0 & P_0^y & P_0^z & 0 \\ 0 & 1 & 0 & P_0^y & P_0^x & 0 & 0 & 0 & P_0^z \end{bmatrix} \quad (4.21)$$

the three-dimensional affine shape-space components have the following interpretation:

$$\Phi = (T_1, T_2, M_{11} - 1, M_{22} - 1, M_{21}, M_{12}, \nu_1, \nu_2) \quad (4.22)$$

where ν_1, ν_2 are the two additional basis elements.

5 The CCD Algorithm

In this chapter, we describe the underlying principles of the Contracting Curve Density (CCD) algorithm, which is shown to over-perform in many challenging problems in the field of computer vision and robotics(paper list). Generally, as mentioned in section 1.3.3, the algorithm includes three steps: initialization, learning local statistics and model parameters refinement. We describe these steps in the following sections respectively.

5.1 Pre-processing and model parametrization

The input of the CCD algorithm is image data (one or multiple images) and parametric curve model. Hence, initialization step comprises pre-processing of input data and model parametrization.

5.1.1 Pre-processing of input data

Before we start to processing step, it is required to improve the quality of the images though the CCD algorithm is proved robust for noise and clutter information in images. Sometimes pre-processing make a problem easier to solve because it greatly reduces the variability , and is helpful to extract features. Furthermore, pre-processing might be performed in order to speed up computation. In our implementation, the idea case is that input data is noiseless, shadowless and have sharp edges or boundaries between objects. There are a group of operations corresponding to this aim.

- **Noise reduction.** Many linear and non-linear algorithms can help to remove noises, such as mean filer, Gaussian blur, Bilinear interpolation and so on.
- **Contrast improvement.** Usually, we can use a window to adjust brightness and contrast by selecting a range of input values which are mapped to the full output range of intensities.
- **Sharpening and detail enhancement.** Scale space is one of methods to highlight and enhance fine details in a image.

OpenCV provides a series of implementation of these operations. Note that for different input data, according to the object properties, illumination and other physical conditions, different operations should be taken.

5.1.2 Contour initialization

After the pre-processing of input data, let's start to discuss the initialization of model contour. It is followed by the process of model parametrization.

In our implementation, we model the contour as a continuous, differentiable and uniform quadratic or cubic B-Spline in \mathbb{R}^2 , which is discussed in chapter 3. Given a ROI (region of interest), the first step is generating sufficient control points $\mathbf{P} = P_0, \dots, P_{N_p-1}$ to do justice to the complexity of the object shape (add a image here).

Now there are two major methods to generate control points: manual initialization and intelligent initialization. The former is given as a simple form by hand. It is easy to use and control. However, there are some problems and limitations in case that the cost of manual operation is large, or vision deviation due to similar brightness between object and background. Therefore, a few new intelligent initialization methods are proposed to extract the contour. In this paper, a initial contour estimation method is described and implemented by employing the well-known SIFT algorithm (source) for keypoint detection and matching. Chapter 6.1(?) be discussed this in detail.

5.1.3 Model parametrization

By applying the (uniform quadratic or cubic) B-spline interpolation to the control points, a new curve grouped by a sequence of equidistant distributed points is generated (add image). The B-spline curve is defined in (Formula 4.12) and is composed of a sequence of points $\mathbf{C} = \{C_0, \dots, C_k\}, k = N_C-1$. N_c denotes the number of sample points in the parametric curve. Because the parametric curve is continuous and differentiable, we can compute the normal $\vec{n} = \{\vec{n}_0, \dots, \vec{n}_k\}$ and tangent vector $\vec{t} = \{\vec{t}_0, \dots, \vec{t}_k\}$ conveniently.

In the planar affine shape-space \mathcal{S} . The curve, a hypothetical initial estimate, can be compactly represented using a vector with 6 real elements Φ , namely model parameters. From the perspective of probability, the hypothetical curve gives a uncertainty, the solution of the curving problem is therefore no longer just a single curve, but a whole family of possible curves. The Gaussian probability density for these possible curves in shape-space \mathcal{S} is:

$$p(\Phi) \propto \exp \left\{ -\frac{1}{2} (\Phi - \Sigma_\Phi)^T \Sigma_\Phi^{-1} (\Phi - \Sigma_\Phi) \right\} \quad (5.1)$$

As claimed in chapter 1, Σ_Φ is a 6×6 covariance matrix, which measure the variability of how much two group of model parameters change together. The information matrix Σ_Φ^{-1} can be written as

$$\Sigma_\Phi^{-1} = \frac{N_\Phi}{\rho_0^2} \mathbf{A}^T \mathcal{U} \mathbf{A} \quad (5.2)$$

where ρ_0^2 denotes the mean-square displacement along the entire curve (AC). \mathbf{A} is the shape-matrix, and \mathcal{U} is metric matrix for curves. N_Φ represents the number of model parameters, for our case, it is 6. Note ρ_0^2 is a real value and can be computed easily as

$$\rho_0^2 = \text{tr}(\Sigma_\Phi) \quad (5.3)$$

where $\text{tr}(\cdot)$ denotes the trace of a matrix.

The pre-processed input data and parametric curve model have been prepared. In the following sections, the iterative procedure of the CCD algorithm will be described in detail.

5.2 Local statistics

5.2.1 Collecting local information

As claimed in chapter 1.1, one of advantages of the curve-fitting problem is that the problem can be restricted to a ROI, and is expected to reduce the computational cost. Therefore, we first define the region which contains pixels in the vicinity of expected image curve. Consider complexity and the expense of computing, it is practicable that choose those pixels actually required for interpolation along normals of the curve. This is proved useful and important for image perception and real-time tracking system. In the implementation of this thesis, processing is limited to a segment of each normal within a search region. A fixed distance h along the normal segment are choose according hypothetical uncertainty of parametric curve. In the special case of the norm-squared prior in spline space, a reasonable search segment is

$$h = \sqrt{2\rho_0} = \sqrt{\text{tr}(\Sigma_{\Phi} \mathbf{A}^T \mathcal{U} \mathbf{A})} \quad (5.4)$$

Usually, h denotes the size of *window* used for computing local statistics. In the begining of interative procedure, the value is relatively big and only describe the vicinity of the image curve roughly due to the high uncertainty. The uncertainty is reduced after some iteration steps, as a result, the h becomes smaller and smaller. After determining the length of search segment, a set of points located on these segments can be collected and evaluated. Note the parametric model curve is not required to be closed, but whatever case it encloses a limited area. We only plan to analyze the pixels located in the vicinity of the contour. Therefore, we should pay attention to limit the search distance on the *internal* side in order to avoid crossing the opposite boundary to sample pixels from the wrong area (ICVS06). In order to decrease the computational expenses, it is advisable to uniformly sample those pixels on both segments in the vicinity of the contour. In the other hand, we should avoid to only collect a small number of pixels, thus it makes no sense from the perspective of statistics. Let's denote the sample distance using δh , then the an overall number of spaced sample points L (2L for both sides in all) can be given by

$$L = \lfloor \frac{h}{\delta h} \rfloor \quad (5.5)$$

As mentioned before, the goal of the algorithm is assignment each pixel $(v_{k,l}, k \in [0, \dots, N_C - 1], l \in [0, 2L - 1])$ to either side of the contour, which is determined by the curve distribution. We first compute the probabilistic assignments \mathbf{a}_v for each pixel $v_{k,l}$

$$\mathbf{a}_v = (a_{v,1}, a_{v,2})^T \quad (5.6)$$

where $a_{v,1}$ describes to which extent a pixel v is expected to be influenced by side 1 of the curve, and $a_{v,2}$ is equivalent for side 2 given by $a_{v,2} = 1 - a_{v,1}$. For arbitrary curve C , it is difficult to give a closed form of a_v . In the following, a efficient approximation of the assignment is derived.

We use $d_{k,l}$ to denote the **signed** distance between pixel $v_{k,l}$ and a given curve C (a image). $d_{k,l}$ can be approximated by

$$d_{k,l} = \vec{n}_k \cdot (\mathbf{v}_{k,l} - C_k) \quad (5.7)$$

where $\mathbf{v}_{k,l} = \begin{pmatrix} x_{k,l} \\ y_{k,l} \end{pmatrix}$ is the axis components of pixel $v_{k,l}$, and C_k is the equivalent for point C on the curve given by $\begin{pmatrix} x_k \\ y_k \end{pmatrix}$. Now consider that the curve is distorted by a Gaussian distribution $p(\Phi)$, therefore, the displacement $d_{k,l}$ is also a Gaussian distribution, $p(d_{k,l}) \sim \mathcal{N}(d_{k,l}|m_d, \sigma)$, where m_d and σ are mean and covariance of the distribution. Covariance *sigma* holds

$$\sigma = \vec{n}_k \cdot \mathbf{J}_k \cdot \Sigma_\Phi \cdot \mathbf{J}_k^T \cdot \vec{n}_K^T \quad (5.8)$$

where \mathbf{J}_k is the Jacobian of curve \mathbf{C} . *sigma* can be take as the uncertainty of the curve along the normal introduced by the covariance Σ_Φ . Therefore, the probability that a point lies on side 1 of the curve can be evaluated by

$$a_{v,1} = \frac{1}{2} \operatorname{erf}\left(\frac{d_{k,l}}{\sqrt{2}\sigma} + 1\right) \quad (5.9)$$

Where $\operatorname{erf}(\cdot)$ is the error function of a distribution. This approximation process is called **fuzzy** (or **smooth**) assignment. The accuracy of this assignment will increase as the uncertainty of curve governed by covariance Σ_Φ decreases.

With this assignment and following the suggested rule in (hanek's paper), we now start to assign two suitable weighting functions ω_1, ω_2 to the pixels $v_{k,l}$ along the normal for the two sides of the curve. the weighting functions are defined as

$$\omega_{1/2}(d_{k,l}) = C \left(\frac{a_{1/2}(d_{k,l}) - \gamma_1}{1 - \gamma_2} \right)^4 [e^{-d_{k,l}/(2\hat{\sigma}^2)} - e^{-\gamma_2}]^+ \quad (5.10)$$

where γ_1 holds 0.5 (disjoint weight assignment) and γ_2 holds 4 for the truncated Gaussian in (5). In addition, the standard deviation is chosen in order to cover the specified distance h , which yields

$$\hat{\sigma} = \max \left[\frac{h}{\sqrt{(2\gamma_2)}}, \gamma_4 \right], \sigma = \frac{1}{\gamma_3} \hat{\sigma} \quad (5.11)$$

with the two additional constants $\gamma_3 = 6$ (linear dependence between σ and $\hat{\sigma}$) and $\gamma_4 = 4$ (minimum weighting window width) introduced in (ICVS06).

In the implementation of this thesis, $2 \cdot L \cdot N_C$ distance ($d_{k,l}$), fuzzy assignment ($a_{v,1}(d_{k,l})$) and weight function ($\omega_1(d_{k,l}), \omega_2(d_{k,l})$) are evaluated offline and saved in a big array. Now we have restrict our analysis the region of interest in the limited area, then collect all statistic information required to learn local statistics. In the next part, we will evaluated the local statistics.

5.2.2 Learning local statistics

For simplicity, in current implementation, we only consider raw RGB statistics. With the pixel coordinates, assignment and weight function, local mean vectors $\mathbf{m}_{v,s}$ and local covariance matrices Σ_Φ will be derived for each side $s \in 1, 2$ of the curve. We first calculate

the zero, first and second order weighted moments $M_{k,s}^0(d_{k,l,s})$, $\mathbf{M}_{k,s}^1(d_{k,l,s})$ and $\mathbf{M}_{k,s}^2(d_{k,l,s})$

$$M_{k,s}^{(0)}(d_{k,l,s}) = \sum_{l=0}^{2L-1} \omega_s(d_{k,l}) \quad (5.12)$$

$$\mathbf{M}_{k,s}^{(1)}(d_{k,l,s}) = \sum_{l=0}^{2L-1} \omega_s(d_{k,l}) \mathbf{I}_{k,l} \quad (5.13)$$

$$\mathbf{M}_{k,s}^{(2)}(d_{k,l,s}) = \sum_{l=0}^{2L-1} \omega_s(d_{k,l}) \mathbf{I}_{k,l} \mathbf{I}_{k,l}^T \quad (5.14)$$

Where \mathbf{I} is just the pixel raw RGB value, its elements' value are between 0 and 255. Then local mean vectors $\mathbf{m}_{v,s}$ and local covariance matrices Σ_Φ are obtained by

$$\mathbf{m}_{k,s} = \frac{\mathbf{M}_{k,s}^{(1)}}{M_{k,s}^{(0)}} \quad (5.15)$$

$$\Sigma_{k,s} = \frac{\mathbf{M}_{k,s}^{(2)}}{M_{k,s}^{(0)}} - \mathbf{m}_{k,s} \mathbf{m}_{k,s}^T + \kappa \mathbf{I} \quad (5.16)$$

In function (????) κI means an identity matrix scaled by κ in order to avoid numerical singularity because later it is required to calculate the inverse matrix of $\Sigma_{k,s}$. In our experiments, we choose κ to be quite small, $\kappa = 0.5$. It is shown that this is efficient to avoid numerical problems in the process of iteration.

With the local mean vectors $\mathbf{m}_{v,s}$ and local covariance matrices Σ_Φ , we can compute the likelihood function $p(\mathbf{I}_{k,l} | \mathbf{m}_{v,1}, \mathbf{m}_{v,2}, \Sigma_{v,1}, \Sigma_{v,2})$ for each pixel $v_{k,l}$. In terms of observation model, the likelihood function describes how probable the observed data set is for different settings of the parameter vector. Hence, we first establish the relation between image data $\mathbf{I}_{k,l}$ and the model parameter Φ . Here we model the pixel value $\hat{\mathbf{m}}_{k,l}$ and $\hat{\Sigma}_{k,l}$ for all pixels $v_{k,l}$ in the vicinity of curve as the linear combination of $\mathbf{m}_{v,1}$ and $\mathbf{m}_{v,2}$,

$$\hat{\mathbf{m}}_{k,l} = a_{v,1}(d_{k,l}) \mathbf{m}_{v,1} + a_{v,2}(d_{k,l}) \mathbf{m}_{v,2} \quad (5.17)$$

If we define $\hat{\Sigma}_{k,l}$ using the rule as $\hat{\Sigma}_{k,l} = a_{v,1}(d_{k,l}) \Sigma_{v,1} + a_{v,2}(d_{k,l}) \Sigma_{v,2}$, the computational cost in the procedure of parameters refinement which is discussed in next section will be dramatically high. Instead, we decide to model the covariance matrix $\hat{\Sigma}_{k,l}$ following the rule in (?????????), but we do not consider it as the function about $d_{k,l}$,

$$\hat{\Sigma}_{k,l} = a_{v,1}(d_{k,l}) \Sigma_{v,1} + a_{v,2}(d_{k,l}) \Sigma_{v,2} \quad (5.18)$$

Now for each observed pixel $\mathbf{I}_{k,l}$, the likelihood function is given by

$$p(\mathbf{I}_{k,l} | \mathbf{m}_{v,1}, \mathbf{m}_{v,2}, \Sigma_{v,1}, \Sigma_{v,2}) = p(\mathbf{I}_{k,l} | \hat{\mathbf{m}}_{k,l}, \hat{\Sigma}_{k,l}) \quad (5.19)$$

However, we hope to the likelihood for all pixels in the vicinity of the curve. If we consider the coupling or other complex relation among different group of pixels, the problem will become dramatically complicated and the cost of computing will be very expensive. We

can avoid these problem if we assume pixels are drawn independently from the same distribution (this is not true), namely independent and identically distributed (i.i.d)(bishop). Thus we can model the likelihood function as

$$p(\mathbf{I}_V | \hat{\mathbf{m}}_V, \hat{\Sigma}_V) = \prod_l \prod_k p(\mathbf{I}_{k,l} | \hat{\mathbf{m}}_{k,l}, \hat{\Sigma}_{k,l}) \quad (5.20)$$

The index V indicates quantities for all pixels v in V . Note we only take into account those pixels which are in the vicinity V of the curve. Pixels outside V are not considered.

We have derived the likelihood function of observed pixels. Together with the input data and prior knowledge, now we can go into parameters refinement stage by applying Bayesian's theorem.

5.3 Refine parameters

With the likelihood function in (5.20) and prior distribution in (5.1), we can estimate the $\hat{\Phi}$ using MAP which is based the Bayesian treatment. Firstly, the posterior distribution holds

$$\begin{aligned} p(\Phi | \mathbf{I}_V) &\propto p(\mathbf{I}_V | \hat{\mathbf{m}}_V(\Phi), \hat{\Sigma}_V(\Phi)) p(\Phi | \mathbf{m}_\Phi, \Sigma_\Phi) \\ &= \frac{1}{(2\pi)^{1/2}} \frac{1}{|\Sigma_\Phi|} \exp\left\{-\frac{1}{2}(\phi - m_\Phi)^T \Sigma_\Phi^{-1} (\phi - m_\Phi)\right\} \cdot \\ &\quad \prod_{k=0}^{N_c-1} \prod_{l=0}^{2L-1} \frac{1}{(2\pi)^{1/2}} \frac{1}{|\hat{\Sigma}_{k,l}|} \exp\left\{-\frac{1}{2}[\mathbf{I}_{k,l} - \hat{\mathbf{m}}_{k,l}(a_{v,1})]^T \hat{\Sigma}_{k,l}^{-1} [\mathbf{I}_{k,l} - \hat{\mathbf{m}}_{k,l}(a_{v,1})]\right\} \end{aligned} \quad (5.21)$$

Let's using a function \mathcal{Q} to denote the logarithm function of right part in (5.21)

$$\begin{aligned} \mathcal{Q} &= -2 \ln \{p(\mathbf{I}_V | \hat{\mathbf{m}}_V(\Phi), \hat{\Sigma}_V(\Phi)) p(\Phi | \mathbf{m}_\Phi, \Sigma_\Phi)\} \\ &= \ln(2\pi) + 2 \ln |\Sigma_\Phi| + \Phi^T \Sigma_\Phi^{-1} \Phi + 2LN_C \cdot \ln 2\pi + \\ &\quad 2 \sum_{k=0}^{N_c-1} \sum_{l=0}^{2L-1} \ln |\hat{\Sigma}_{k,l}| + \sum_{k=0}^{N_c-1} \sum_{l=0}^{2L-1} \{[\mathbf{I}_{k,l} - \hat{\mathbf{m}}_{k,l}(a_{v,1})]^T \hat{\Sigma}_{k,l}^{-1} [\mathbf{I}_{k,l} - \hat{\mathbf{m}}_{k,l}(a_{v,1})]\} \end{aligned} \quad (5.22)$$

We interpret the estimate $\hat{\Phi}$ of the model parameters Φ as the mean \mathbf{m}_Φ of a Gaussian approximation of the posterior distribution, and $\hat{\Phi}$ can be evaluated as

$$\hat{\Phi} = \mathbf{m}_\Phi \arg \max_{\Phi} \mathcal{Q} \quad (5.23)$$

For the estimate \mathbf{m}_Φ optimizing \mathcal{Q} , the Gauss-Newton approximation to the Hessian matrix can be adopted, first the partial derivatives of \mathcal{Q} is computed as

$$\nabla_{\Phi} \{\mathcal{Q}(\Phi)\} = 2\{\Sigma_\Phi^{-1}\}^T \Phi - \sum_{k=0}^{N_c-1} \sum_{l=0}^{2L-1} \{\mathcal{J}_{a_{v,1}}^T \hat{\Sigma}_{k,l}^{-1} [\mathbf{I}_{k,l} - \hat{\mathbf{m}}_{k,l}(a_{v,1})]\} \quad (5.24)$$

with

$$\mathcal{J}_{a_{v,1}} = (\mathbf{m}_{k,1} - \mathbf{m}_{k,2}) (\nabla_\phi a_{v,1}(d_{k,l}))^T \quad (5.25)$$

because $d_{k,l}$ is a function with respect to Φ , there $a_{v,1}(d_{k,l})$ holds

$$\nabla_{\Phi} a_{v,1}(d_{k,l}) = \frac{\partial a_{v,1}(d_{k,l})}{\partial d_{k,l}} \left(\frac{\partial d_{k,l}}{\partial x_{k,l}} \mathbf{J}_{\Phi}(\mathbf{v}_{k,l}(x)) + \frac{\partial d_{k,l}}{\partial y_{k,l}} \mathbf{J}_{\Phi}(\mathbf{v}_{k,l}(y)) \right) \quad (5.26)$$

Where $\mathbf{v}_{k,l}(x)$ and $\mathbf{v}_{k,l}(y)$ are the axis components of pixel $\mathbf{v}_{k,l}$, and $d_{k,l}$ is given by

$$d_{k,l} = (x_{k,l} - x_k)n_k(x) + (y_{k,l} - y_k)n_k(y) \quad (5.27)$$

with $n_k(x)$ and $n_k(y)$ are the components of normal vector of curve point C_k . Moreover, we have

$$\nabla_{\Phi} a_{v,1}(d_{k,l}) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left\{ -\frac{d_{k,l}^2}{2\sigma^2} \right\} (n_k(x)\mathbf{J}_{\Phi}(\mathbf{v}_{k,l}(x)) + n_k(y)\mathbf{J}_{\Phi}(\mathbf{v}_{k,l}(y))) \quad (5.28)$$

In terms of the properties of B-spline curve and planar affine model-space. The pixel coordinate of \mathbf{v}_{kl} can be written as

$$\mathbf{v}_{k,l}(x) = \mathbf{U}_k^T \mathbf{A}_x \Phi + \mathbf{U}_k P_0(y) + \Delta_h n_k(x) \quad (5.29)$$

$$= \Phi_0 \sum_i^n U_{k,i} + (1 + \Phi_2) \sum_i^n U_{k,i} * x_{k,i} + \Phi_5 \sum_i^n U_{k,i} y_{k,i} + \Delta_h n_k(x) \quad (5.30)$$

$$\mathbf{v}_{k,l}(y) = \mathbf{U}_k^T \mathbf{A}_x \Phi + \mathbf{U}_k P_0(y) + \Delta_h n_k(y) \quad (5.31)$$

$$= \Phi_0 \sum_i^n U_{k,i} + (1 + \Phi_2) \sum_i^n U_{k,i} * x_{k,i} + \Phi_5 \sum_i^n U_{k,i} y_{k,i} + \Delta_h n_k(x) \quad (5.32)$$

Now we can compute \mathbf{J}_{Φ} by the following formula

$$\mathbf{J}_{\Phi}(\mathbf{v}_{k,l}) = \begin{bmatrix} \frac{\partial \mathbf{v}_{k,l}(x)}{\partial \Phi_0} & \frac{\partial \mathbf{v}_{k,l}(x)}{\partial \Phi_1} & \frac{\partial \mathbf{v}_{k,l}(x)}{\partial \Phi_2} & \frac{\partial \mathbf{v}_{k,l}(x)}{\partial \Phi_3} & \frac{\partial \mathbf{v}_{k,l}(x)}{\partial \Phi_4} & \frac{\partial \mathbf{v}_{k,l}(x)}{\partial \Phi_5} \\ \frac{\partial \mathbf{v}_{k,l}(y)}{\partial \Phi_0} & \frac{\partial \mathbf{v}_{k,l}(y)}{\partial \Phi_1} & \frac{\partial \mathbf{v}_{k,l}(y)}{\partial \Phi_2} & \frac{\partial \mathbf{v}_{k,l}(y)}{\partial \Phi_3} & \frac{\partial \mathbf{v}_{k,l}(y)}{\partial \Phi_4} & \frac{\partial \mathbf{v}_{k,l}(y)}{\partial \Phi_5} \end{bmatrix} \quad (5.33)$$

Afterwards, the Gauss-Newton approximation to the Hessian matrix is given by

$$\mathcal{H}_{\Phi} \mathcal{Q} = \Sigma_{\Phi}^{-1} + \sum_{k=0}^{N_c-1} \sum_{l=0}^{2L-1} \left\{ \mathcal{J}_{a_{v,1}}^T \hat{\Sigma}_{k,l}^{-1} \mathcal{J}_{a_{v,1}} \right\} \quad (5.34)$$

The overall gradient and Hessian matrices for the optimization are obtained by adding the prior cost function derivatives, and the Newton optimization step can finally be performed as

$$\begin{aligned} \mathbf{m}_{\Phi}^{new} &= \mathbf{m}_{\Phi} - (\mathcal{H}_{\Phi} \mathcal{Q})^{-1} \nabla_{\Phi} \mathcal{Q} \\ \Sigma_{\Phi}^{new} &= c \Sigma_{\Phi} - (1 - c) (\mathcal{H}_{\Phi} \mathcal{Q})^{-1} \end{aligned} \quad (5.35)$$

with $c = \frac{1}{4}$ in our implementation. Note the covariance matrix is updated as well by an exponential decay rule. Coefficient c specifies the maximum decrease of the covariance within one iteration step[Hanek's paper]. If c is very large, due the slow reduction of covariance the convergence process will be very slow. On the other hand, if c is very small, it might converge to a wrong solution.

We can investigate the iteration process of the CCD algorithm by comparing with the counterpart of k-means or EM algorithms, both of which also include two stages. First it is required to choose some initial values for the parameters governing the posterior distribution. Then in the E (expectation) step, minimizing some cost function while keeping the parameters fixed. In the M (maximization) step, calculate the new values of the parameters while keeping the expectation values fixed. This two-stage optimization is then repeated until convergence.

The two stages of the CCD algorithm are iterated until the convergence criterion is satisfied. After each iteration step, there are two model parameters vector \mathbf{m}_Φ^{old} and \mathbf{m}_Φ^{new} , which correspond two spline curves \mathbf{C}^{old} and \mathbf{C}^{new} respectively. Which kind of convergence criterion should we choose? Two options, one is using the norm of difference between \mathbf{m}_Φ^{old} and \mathbf{m}_Φ^{new} , another is by measuring the displacement between two spline curves, \mathbf{C}^{old} and \mathbf{C}^{new} . Experiments show that the measure of curve difference using the norm is sensitive to parametrization, and computation of the normal curve-displacement measure $D(\mathbf{C}^{new}, \mathbf{C}^{old})$ is feasible if we are content with local, rather than global minimization. The normal curve-displacement between two points on $\mathbf{C}^{new}(u)$ and $\mathbf{C}^{old}(u)$ is given by

$$D(u) \approx [\mathbf{C}^{new}(u) - \mathbf{C}^{old}(u)] \cdot \mathbf{n}(u) \quad (5.36)$$

Where the total displacement $\mathbf{C}^{new}(u) - \mathbf{C}^{old}(u)$ at a point represents the sum of components along the curve tangent and normal respectively. The tangential component approximate the displacement along the curve \mathbf{C}^{old} , which reflects the variation of parameterisation between curves . If the tangential component is eliminated, the normal component will represents purely the distance between curves. (AC) The use of normal displacement is a standard technique in Computer Vision and can be explained intuitively. Traversing all points on curves, the curve-displacement holds

$$D(\mathbf{C}^{new}, \mathbf{C}^{old}) = \sum_{i=0}^{N_C-1} [\mathbf{C}_i^{new} - \mathbf{C}_i^{old}] \cdot \mathbf{n}_i \quad (5.37)$$

In the implementation of this thesis, we use the $(\mathbf{C}^{new}, \mathbf{C}^{old})$ as the convergence criterion of the iteration. Moreover, the curve-displacement can also be used to detect and delete rogue data or outliers.(AC)

Finally, the best recorded estimate \mathbf{m}_Φ^{best} and Σ_Φ^{best} will be given as output the CCD algorithm.

5.4 The CCD tracker

We have derived the CCD algorithm in the previous section. Generally, it is used to cope with the curve-fitting problem in a single image. Like some other model-based segmentation algorithms, the CCD algorithm can also be used to handle object tracking problem in a sequence of images captured at successive time-steps. Some distinguishing characteristics of make it stable and efficient.

In this section we describe a naive tracking algorithm based the CCD algorithm.

A sequence of images for example, video data, consist of neighboring 2-D slices of a higher-dimensional volume. Usually tow neighboring frames have similar information

and features and might only be distinguished by investigation in detail. Hence, a naive CCD tracking algorithm can be obtained by applying the CCD algorithm to each frame independently, and the output of parameters obtained from the previous frame can be used in the next frame of image. However, such a tracker would be of limited use because the expense of computation will be intolerable in a practical environment which require high performance and efficiency. Furthermore, without exploiting statistical dependencies between successive images, the tracker can not work if a radical change or big perturbation happened while switching into a new frame. A real-time CCD tracker exploiting the coherence of curve motion and the temporal coherence of pixel values is proposed in (Hanek's thesis).

5.5 Summary of the CCD algorithm and its variant

In this section, for sake of clarity, a sketch of the naive CCD tracker is below given. Note we only consider the manual implementation.

5.6 Efficiency and complexity

Table 5.1: Algorithm: the naive Contracting Curve Density (CCD) tracker

1. Sequence image data $\mathbf{I}(t)$, t denotes the time-step

Input: 2. A set of control points: $\mathbf{P} = \{P_0, \dots, P_{N_{cp}}\}$, N_{cp} is the number of control points.

Output:

1. estimated model parameters vectors \mathbf{m}_Φ^{best}
2. corresponding covariance matrix Σ_Φ^{best}

1. Apply some pre-processing methods to the image data to $\mathbf{I}(0)$, 0 denotes the first frame

Initialization:

2. Compute the B-spline curve $\mathbf{C}_0(u)$ and corresponding normal and tangent vectors
3. Initialize $\Sigma_\Phi(0)$ as zero vector
4. Compute the $\Sigma_\Phi(0)$

if the time-step $t > 0$, execute the *initialization* step for the frame of image.

Main loop: Firstly do pre-processing, then initialize $\Sigma_\Phi(t)$ as $\Sigma_\Phi^{best}(t-1)$, and generate a new B-spline curve using $\Sigma_\Phi(t)$, last compute the $\Sigma_\Phi(0)$. After these initialization steps, go into loop of two stages of the CCD algorithm.

1. Compute the uncertainty norm segment length h in terms of current estimate

2. Collect the pixels $v_{k,l}$ according to the h and assign weight to these pixels.

3. Compute the local statistics $\mathbf{m}_{k,s}$ and $\Sigma_{k,s}$

4. Evaluate the observe model parameters $\hat{\mathbf{m}}_{k,l}$ and $\hat{\Sigma}_{k,l}$

5. Compute the gradient $\nabla_\Phi \mathcal{Q}$ and the Hessian $\mathcal{H}_\Phi \mathcal{Q}$

6. Apply optimization algorithm to the \mathcal{Q} in order to compute the \mathbf{m}_Φ^{new} and Σ_Φ^{new}

7. Compute the curve-displacement $\mathbf{D}(\mathbf{C}^{new}, \mathbf{C}^{old})$, if it is satisfied with the convergence criterion, accept the $\mathbf{m}_\Phi^{best}(t)$ and Σ_Φ^{best} . Otherwise, go to step 1 and until convergence.

if convergence condition is satisfied, $\mathbf{m}_\Phi^{best}(t)$ and $\Sigma_\Phi^{best}(t)$ are obtained. Then read a new frame of image and go to the beginning of main loop.

6 Experiments and Results

So far we have discussed the CCD approach in detail. In this chapter, the experimental evaluation of the CCD algorithm and the naive CCD tracker will be provided. In previous chapter, we have introduced the application of the CCD algorithm and its variant. Experiments have been performed on the PR2. we apply the CCD approach to three different kinds of model fitting and tracking problems:

1. Contour convergence on still image
2. Tracking initialization from sift features
3. Tracking initialization from 3D point cloud

We analyze the performance of the approach in terms of robustness, accuracy and run-time. This is followed the comparison between the CCD approach and other state-of-the-art methods.

6.1 Contour convergence on still image

As mentioned in related work, the CCD algorithm is an effective segmentation method like other model-based segmentation algorithms.

A segmentation of an image $I : \Omega \subset R^2 \rightarrow R$ is the partitioning of its domain into homogeneous regions $\Omega_1, \dots, \Omega_n \subset \Omega$. There are many practical applications of image segmentation, such as medical imaging, object tracking, face recognition, fingerprint recognition, machine vision and so on. In this thesis, we concern its application in object tracking. In many cases segmentation is the bottleneck when trying to track an object. Many segmentation methods are developed, however, there is not general solution to the segmentation problem. In the real world, segmentation problems will often require much domain knowledge before a successful segmentation can be performed. The CCD algorithm is just such a method which is combined with the prior knowledge. By converting a pure segmentation problem to a problem in pattern recognition, a large amount of techniques in the field of pattern recognition are introduced to solve segmentation problems, which are proved to be effective and helpful. Compared with other segmentation methods, such as intensity-based method, cluster method, region-based method and compression-based methods, the CCD algorithm has following advantages:

- restrict segmentation problem to a limited explicit region, it is helpful to decrease the computation cost.
- probabilistic representation of the variation of the registered samples could be easily given, thus statistical inference between the model and the image will be applied. All these are helpful to improve the segmentation results.

- The CCD algorithm achieves sub-pixel accuracy and high robustness because only a relatively small fraction of the pixels is taken into account in the end of iteration.

In the following, we apply the proposed CCD algorithm to 4 kind of objects:

1. segmentation of a ball
2. segmentation of 3-D object
3. segmentation of a transparent object
4. fitting a deformable models

6.1.1 Segmentation of a ball

Like other model-based method, we model a contour of the ball as the prior knowledge. Due to existing general contour for spherical object, besides manual initialization, we can use some built-in function in OpenCV library to generate the contour with given center and radius. According the definition of prior distribution, it is required to guarantee that the hypothesis is in the range of the variation of registered sample, namely, the initial contour can not be too far away from the target.

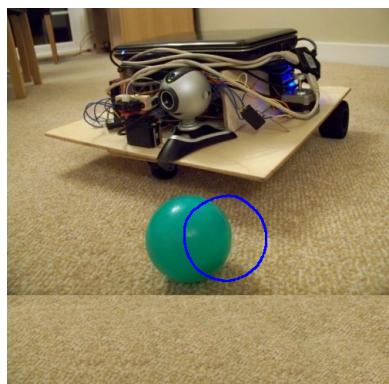
In figure 6.1, we intentionally set a hypothesis which is far away from the ball, but after about 15 iteration, we successfully fit the contour the edge of the ball. For the sake of a ground truth we need investigate the pixels used in the iterations. The figure 6.2 show that only a small number of pixels is taken into account after the contour starts to converge. However, these pixels contain the necessary information for refining the fit. Because the pixels used to compute became smaller and smaller, at last, it is supposed that only a small number of pixels near the boundary are considered. The CCD algorithm can achieve high sub-pixel accuracy.

6.1.2 Segmentation of a 3-D obejct

In the experiments of segmentation of a spherical object, a planar affine space of contour is generated. It works very well because the contour of the ball is always a circle, there is not parallax effect for such objects. In this case, a 6-dimension model parameters vector can cope with basic geometrical transformation of a object. However, clearly it can not be expected to suffice for non-planar surfaces. Imagine such a scenario, we generate the planar affine space contour of a non-planar cube in the fist view. In the subsequent view the outline of the cube does not lie in the old space. Due the parallax effect the a mismatch in the fitted contour will be expected to detect. This can be evidenced in the by the fitting experiments in figure 6.3.

The new 3-dimensional affine shape-space with 8 degree of freedom, made up of the 6-parameter planar affine space and a two-parameter extension, is introduced to cope with such problem. Two added components account for the depth variation that is not visible in the template view. If tacking the two new parameters to the shape matrix, the form of shape matrix for 3-dimensional case can be given by

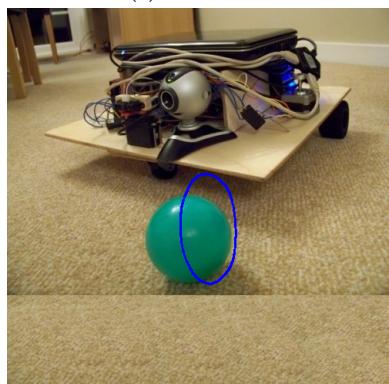
$$A = \begin{bmatrix} 1 & 0 & P_0^x & 0 & 0 & P_0^y & P_0^z & 0 \\ 0 & 1 & 0 & P_0^y & P_0^x & 0 & 0 & P_0^z \end{bmatrix} \quad (6.1)$$



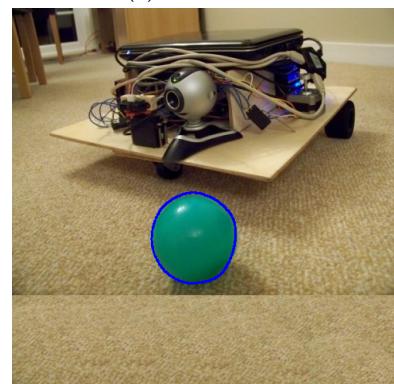
(a) iteration 1



(b) iteration 3



(c) iteration 7



(d) iteration 16

Figure 6.1: segmentation of a ball

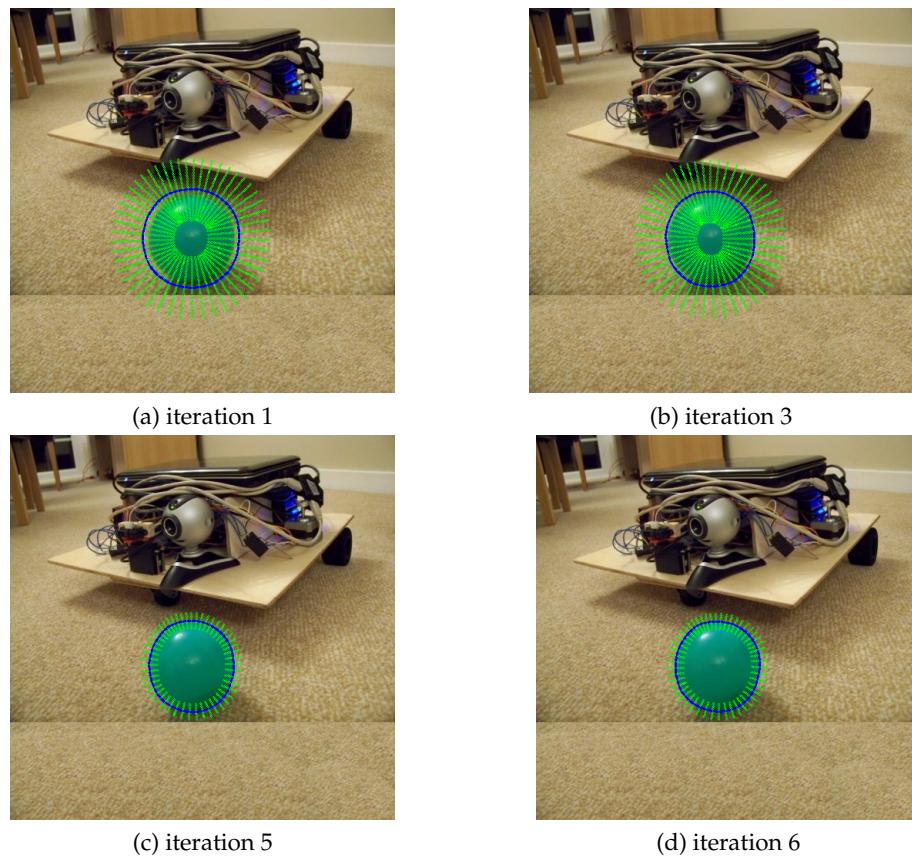


Figure 6.2: pixels used in each iteration

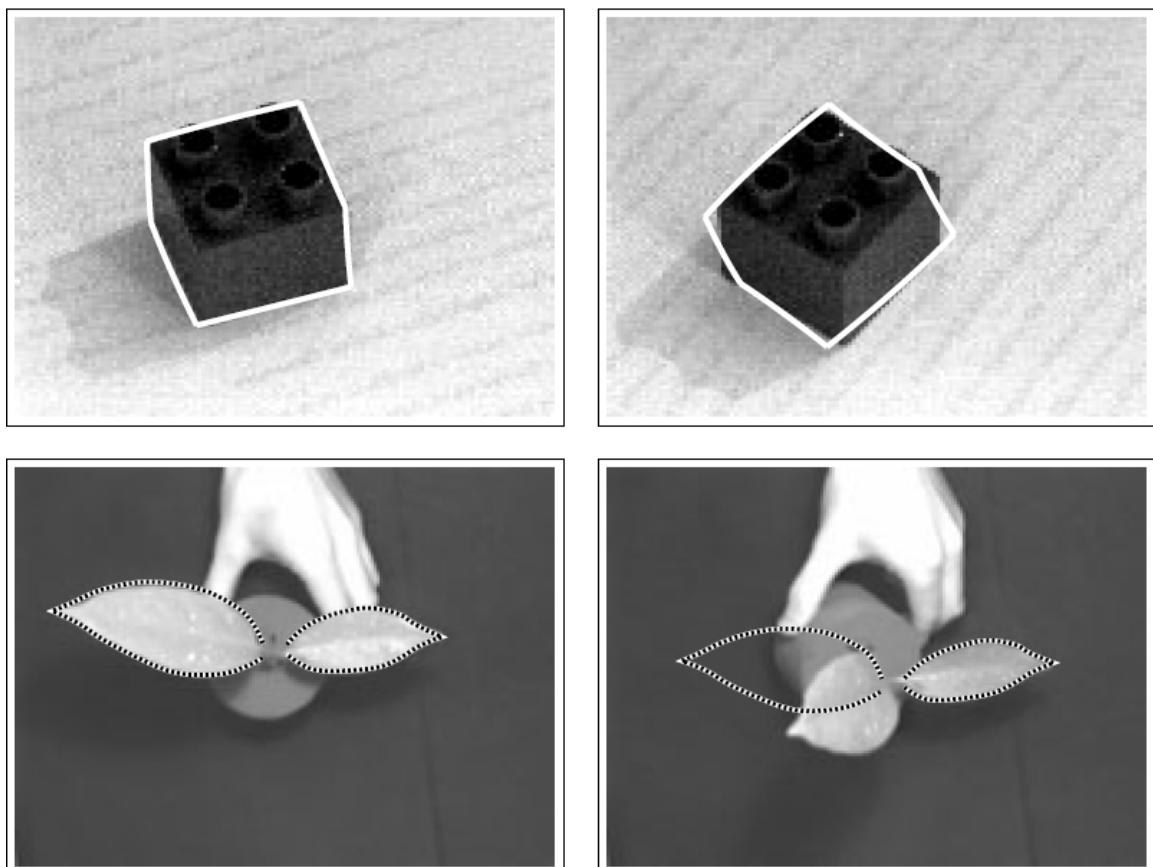


Figure 6.3: Planar space can not encompass a general non-planar contour

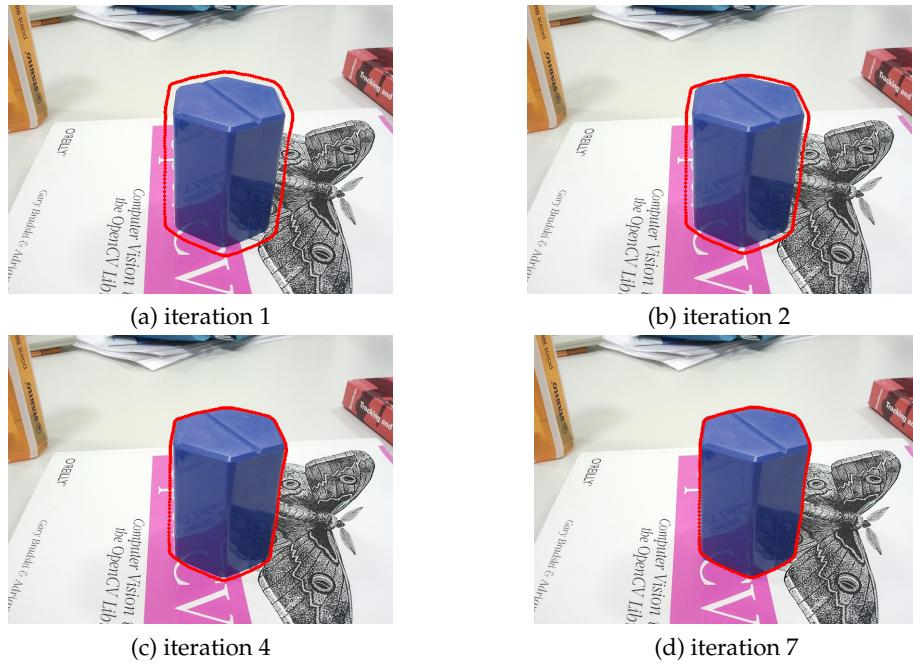


Figure 6.4: segmentation of a non-planar object

The three-dimensional affine shape-space model parameters have the following interpretation

$$\Phi = (T_1, T_2, M_{11} - 1, M_{22} - 1, M_{21}, M_{12}, \nu_1, \nu_2) \quad (6.2)$$

The expanded space now encompasses the outlines of all views of the three-dimensional outline, evidenced by figure 6.4

6.1.3 Segmentation of a transparent object

In this thesis, local statistical information is collected in terms of pixel intensity in the vicinity of a contour, therefore, the CCD algorithm works not well for transparent. In this experiment, the target is a cola bottle whose upper part is nearly transparent and the lower part is opaque. The result in 6.5 shows that the CCD algorithm can segment the lower part but can not encompass the contour in the upper part completely. However, because the model only have 6 (or 8 for 3-dimensional object) degree of freedom, the convergence of other part will effect the fitting in the transparent part such that we can detect that the convergence happened in the transparent part.

6.1.4 Fitting a rigid wire frame

The next example shows that the CCD algorithm is not only able to exploit object contours, but it can also use internal model edges separating multiple, possibly dependent, image regions. We fit a rigid wire frame model to the image depicted in Figure 5.12a. The degrees of freedom are the 8 pose parameters. The CCD algorithm reliably estimates the pose of

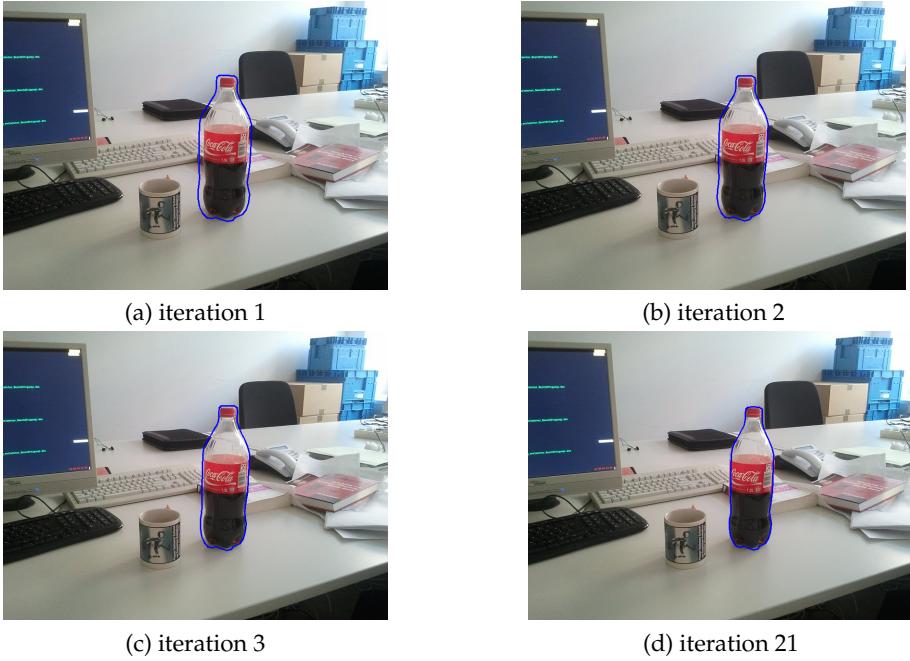


Figure 6.5: segmentation of a transparent bottle

the box despite the partial occlusion. Again, a gradient-based edge detector (Steger, 2000) yields multiple false positive and false negative errors, see Figure 6.6

6.2 Tracking initialization from sift features

Scale-invariant feature transform (or SIFT) is an algorithm to detect and describe local features in images. It is widely used to solve many problems in the field of computer vision, such as object recognition, robotic mapping and navigation, video tracking, and match moving. There are four key stages in the sift algorithm[20]:

1. **Scale-invariant feature detection:** an image is transformed into a collection of vectors which are scale-invariant. By applying a difference-of-Gaussian function to a series of smoothed and resampled images, potential interest can be identified.
2. **Keypoint localization and indexing:** keypoints are selected at given candidate location, then store SIFT keys and identifying matching keys from the new image
3. **Orientation assignment:** based on local image gradient directions each keypoint is assigned one or more orientations.
4. **Keypoint descriptor:** measure the local image gradients at the selected scale in the region around each keypoint

The SIFT keypoints and features are local and based on the appearance of the object at particular interest points. The SIFT algorithm has following features and advantages:

6 Experiments and Results

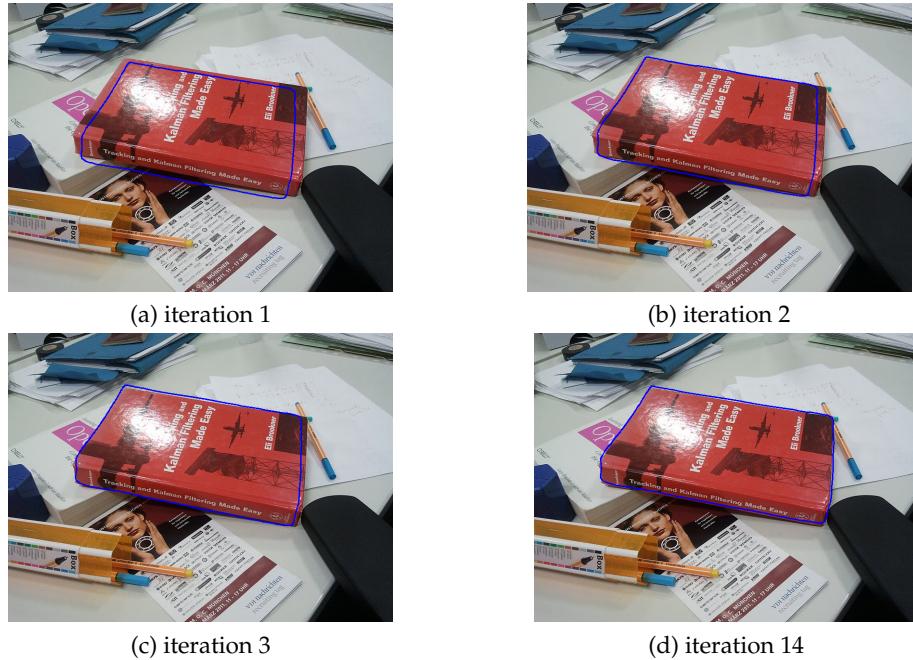


Figure 6.6: Fitting a rigid wire frame

- invariant to image scale and rotation, robust to changes in illumination and minor changes in viewpoint
- highly distinctive, easy to extract, allow for correct object identification and easy to match against a database of local features.
- need few features (as few as 3) from an object to compute its location and pose. Computation cost is moderate on modern computer hardware.

Based on these features, SIFT is usable for object recognition, the steps are given below.

- extract SIFT features from a series of input template image, then store these features into a database.
- given a new input image, by using the algorithm, the features in the image are matched to the SIFT feature database obtained from the training images.

In this thesis, we plan to use the SIFT algorithm to initialize the contour of an object. Assume we have marked the boundary points in the training images. In order project these boundary points to a image being studied, it is essential to estimate the homography between the images. The SIFT is helpful for realize this. However, SIFT matching in the step 2 above might leads to lots of "false" matches. In order to evaluate the homography, firstly we have to exclude the mismatches.

RANSAC is a good approach for this and ideal partner of SIFT. It is a iterative process by randomly select enough matches to fit homography. It contains following stages [8].

- Randomly select a sample of matched points and instantiate the model from this subset

- Determine the set of data points that are within a distance threshold of the homography. The set is the consensus set of the sample and defines the inliers.
- If the size of the set (e.g. the number of inliers) is greater than some threshold, re-estimate the homography using all the matched points and terminate. Otherwise, select a new subset and repeat the above.
- After some trials the largest consensus set is selected, and the homography is re-estimated using all the points in the subset.

One fast way to compute the homography in a least squares sense is to use the Normalized Direct Linear Transform (normalized DLT)[13]. The Normalized DLT algorithm computes a homography for a projective transformation by using at least 4 point correspondences and then minimizing corresponding norm.

After obtaining the homography between the template image and the image being studied, the contour of object can be easily projected into the target image to obtain the estimate of the contour position. This is illustrated in Figure 6.7 on page 36 and 6.8 on page 37. Note that because SIFT is only invariant for minor changes in view points, the method does not always work for non-planar object which have different viewpoint from the one in the template image. Hence, those objects in our examples are planar.

6.3 Tracking initialization from 3D point cloud

A point cloud is a set of vertices in a three-dimensional coordinate system. These vertices are usually defined by X, Y, and Z coordinates, and typically are intended to be representative of the external surface of an object. Point clouds are most often created by 3D scanners. These devices measure in an automatic way a large number of points on the surface of an object, and often output a point cloud as a data file. The point cloud represents the set of points that the device has measured. As the result of a 3D scanning process point clouds are used for many purposes, including to create 3D CAD models for manufactured parts, metrology/quality inspection, and a multitude of visualization, animation, rendering and mass customization applications. While point clouds can be directly rendered and inspected,[1] usually point clouds themselves are generally not directly usable in most 3D applications, and therefore are usually converted to polygon or triangle mesh models, NURBS surface models, or CAD models through a process commonly referred to as surface reconstruction. There are many techniques for converting a point cloud to a 3D surface. Some approaches, like Delaunay triangulation, alpha shapes and ball pivoting, build a network of triangles over the existing vertices of the point cloud, while other approaches convert the point cloud into a volumetric distance field and reconstruct the implicit surface so defined through a marching cubes algorithm.[2] One application in which point clouds are directly usable is industrial metrology or inspection. The point cloud of a manufactured part can be aligned to a CAD model (or even another point cloud), and compared to check for differences. These differences can be displayed as color maps that give a visual indicator of the deviation between the manufactured part and the CAD model. Geometric dimensions and tolerances can also be extracted directly from the point cloud. Point clouds can also be used to represent volumetric data used for example in medical



Figure 6.7: sift contour initialization



Figure 6.8: sift contour initialization

6 Experiments and Results

imaging. Using point clouds multi-sampling and data compression is achieved.[3] In geographic information system, point clouds are one of the sources to make digital elevation model of the terrain.[4] The point clouds are also employed in order to generate 3D model of urban environment, e.g.[5]

employed in order to generate 3D model of urban environment e.g.

7 Conclusion and Future Work

Bibliography

- [1] J. Aloimonos, I. Weiss, and A. Bandyopadhyay. Active vision. *International Journal of Computer Vision*, 1(4):333–356, 1988.
- [2] A. Blake, M. Isard, et al. *Active contours*, volume 2. Springer London, 1998.
- [3] P. Brigger, J. Hoeg, and M. Unser. B-spline snakes: A flexible tool for parametric contour detection. *Image Processing, IEEE Transactions on*, 9(9):1484–1496, 2000.
- [4] E. Brookner and J. Wiley. *Tracking and Kalman filtering made easy*. Wiley Online Library, 1998.
- [5] T.F. Cootes, C.J. Taylor, D.H. Cooper, J. Graham, et al. Active shape models-their training and application. *Computer vision and image understanding*, 61(1):38–59, 1995.
- [6] A.P. Dempster, N.M. Laird, and D.B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1):1–38, 1977.
- [7] C. Ding and X. He. K-means clustering via principal component analysis. In *Proceedings of the twenty-first international conference on Machine learning*, page 29. ACM, 2004.
- [8] M.A. Fischler and R.C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981.
- [9] M. Hahn et al. Tracking of human body parts using the multiocular contracting curve density algorithm. 2007.
- [10] R. Hanek. *Fitting parametric curve models to images using local self-adapting separation criteria*. PhD thesis, Technische Universit
"at M
"unchen, Universit
"atsbibliothek, 2004.
- [11] R. Hanek and M. Beetz. The contracting curve density algorithm: Fitting parametric curve models to images using local self-adapting separation criteria. *International Journal of Computer Vision*, 59(3):233–258, 2004.
- [12] C. Harris. Tracking with rigid models. In *Active vision*, pages 59–73. MIT Press, 1993.
- [13] R. Hartley and A. Zisserman. *Multiple view geometry in computer vision*. Cambridge Univ Pr, 2003.

Bibliography

- [14] M. Isard and A. Blake. ICONDENSATION: Unifying low-level and high-level tracking in a stochastic framework. *Computer Vision-ECCV'98*, page 893, 1998.
- [15] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *International journal of computer vision*, 1(4):321–331, 1988.
- [16] A. Kelemen, G. Székely, and G. Gerig. Elastic model-based segmentation of 3-D neuroradiological data sets. *Medical Imaging, IEEE Transactions on*, 18(10):828–839, 1999.
- [17] A. Kelemen, G. Székely, and G. Gerig. Three-dimensional model-based segmentation of brain MRI. In *Biomedical Image Analysis, 1998. Proceedings. Workshop on*, pages 4–13. IEEE, 1999.
- [18] L. Liu and S. Sclaroff. Deformable shape detection and description via model-based region grouping. In *Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on.*, volume 2. IEEE, 1999.
- [19] S. Lobregt and M.A. Viergever. A discrete dynamic contour model. *Medical Imaging, IEEE Transactions on*, 14(1):12–24, 1995.
- [20] D.G. Lowe. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2):91–110, 2004.
- [21] T. McInemey and D. Terzopoulos. Topology adaptive deformable surfaces for medical image volume segmentation. *Medical Imaging, IEEE Transactions on*, 18(10):840–850, 1999.
- [22] T. McInerney and D. Terzopoulos. Topologically adaptable snakes. In *iccv*, page 840. Published by the IEEE Computer Society, 1995.
- [23] G. Panin and A. Knoll. Fully automatic real-time 3d object tracking using active contour and appearance models. *Journal of Multimedia 2006*, 1(7):62–70, 2006.
- [24] G. Panin, A. Ladikos, and A. Knoll. An efficient and robust real-time contour tracking system. 2006.
- [25] J. Schick and E.D. Dickmanns. Simultaneous estimation of 3D shape and motion of objects by computer vision. In *Visual Motion, 1991., Proceedings of the IEEE Workshop on*, pages 256–261. IEEE, 1991.
- [26] S. Sclaroff and L. Liu. Deformable shape detection and description via model-based region grouping. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 23(5):475–489, 2001.
- [27] GL Scott. The alternative snake—and other animals. In *Proceedings 3rd Alvey Vision Conference (Cambridge)*, pages 341–347, 1987.
- [28] H.W. Sorenson. Parameter estimation: principles and problems. 1980.
- [29] G. Storvik. A Bayesian approach to dynamic contours through stochastic sampling and simulated annealing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 976–986, 1994.

- [30] D. Terzopoulos and D. Metaxas. Dynamic 3D models with local and global deformations: Deformable superquadrics. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 703–714, 1991.
- [31] D. Terzopoulos and R. Szeliski. Tracking with Kaiman snakes. *Active vision*, pages 3–20, 1992.
- [32] C. Xu and J.L. Prince. Snakes, shapes, and gradient vector flow. *Image Processing, IEEE Transactions on*, 7(3):359–369, 1998.
- [33] C. Xu and J.L. Prince. Gradient vector flow deformable models. *Handbook of Medical Imaging, Academic Press, pp*, pages 159–170, 2000.