



Computer Vision

A MODERN APPROACH

Second Edition

Forsyth • Ponce

ALWAYS LEARNING

PEARSON

COMPUTER VISION

A MODERN APPROACH

SECOND EDITION

DAVID A. FORSYTH

University of Illinois at Urbana-Champaign

JEAN PONCE

Ecole Normale Supérieure

International Edition contributions by

SOUMEN MUKHERJEE

RCC Institute of Information Technology

ARUP KUMAR BHATTACHARJEE

RCC Institute of Information Technology

PEARSON

Boston Columbus Indianapolis New York San Francisco Upper Saddle River
Amsterdam Cape Town Dubai London Madrid Milan Munich Paris Montreal Toronto
Delhi Mexico City Sao Paulo Sydney Hong Kong Seoul Singapore Taipei Tokyo

Computer Vision: A Modern Approach

Table of Contents

Cover

Title Page

Contents

I IMAGE FORMATION

1 Geometric Camera Models

1.1 IMAGE FORMATION

1.1.1 Pinhole Perspective

1.1.2 Weak Perspective

1.1.3 Cameras with Lenses

1.1.4 The Human Eye

1.2 INTRINSIC AND EXTRINSIC PARAMETERS

1.2.1 Rigid Transformations and Homogeneous Coordinates

1.2.2 Intrinsic Parameters

1.2.3 Extrinsic Parameters

1.2.4 Perspective Projection Matrices

1.2.5 Weak-Perspective Projection Matrices

1.3 GEOMETRIC CAMERA CALIBRATION

1.3.1 A Linear Approach to Camera Calibration

1.3.2 A Nonlinear Approach to Camera Calibration

1.4 NOTES

2 Light and Shading

2.1 MODELLING PIXEL BRIGHTNESS

2.1.1 Reflection at Surfaces

2.1.2 Sources and Their Effects

2.1.3 The Lambertian+Specular Model

2.1.4 Area Sources

2.2 INFERENCE FROM SHADING

2.2.1 Radiometric Calibration and High Dynamic Range Images

2.2.2 The Shape of Specularities

2.2.3 Inferring Lightness and Illumination

Table of Contents

2.2.4 Photometric Stereo: Shape from Multiple Shaded Images

2.3 MODELLING INTERREFLECTION

2.3.1 The Illumination at a Patch Due to an Area Source

2.3.2 Radiosity and Exitance

2.3.3 An Interreflection Model

2.3.4 Qualitative Properties of Interreflections

2.4 SHAPE FROM ONE SHADED IMAGE

2.5 NOTES

3 Color

3.1 HUMAN COLOR PERCEPTION

3.1.1 Color Matching

3.1.2 Color Receptors

3.2 THE PHYSICS OF COLOR

3.2.1 The Color of Light Sources

3.2.2 The Color of Surfaces

3.3 REPRESENTING COLOR

3.3.1 Linear Color Spaces

3.3.2 Non-linear Color Spaces

3.4 A MODEL OF IMAGE COLOR

3.4.1 The Diffuse Term

3.4.2 The Specular Term

3.5 INFERENCE FROM COLOR

3.5.1 Finding Specularities Using Color

3.5.2 Shadow Removal Using Color

3.5.3 Color Constancy: Surface Color from Image Color

3.6 NOTES

II EARLY VISION: JUST ONE IMAGE

4 Linear Filters

4.1 LINEAR FILTERS AND CONVOLUTION

4.1.1 Convolution

4.2 SHIFT INVARIANT LINEAR SYSTEMS

4.2.1 Discrete Convolution

4.2.2 Continuous Convolution

4.2.3 Edge Effects in Discrete Convolutions

4.3 SPATIAL FREQUENCY AND FOURIER TRANSFORMS

Table of Contents

4.3.1 Fourier Transforms

4.4 SAMPLING AND ALIASING

4.4.1 Sampling

4.4.2 Aliasing

4.4.3 Smoothing and Resampling

4.5 FILTERS AS TEMPLATES

4.5.1 Convolution as a Dot Product

4.5.2 Changing Basis

4.6 TECHNIQUE: NORMALIZED CORRELATION AND FINDING PATTERNS

4.6.1 Controlling the Television by Finding Hands by Normalized Correlation

4.7 TECHNIQUE: SCALE AND IMAGE PYRAMIDS

4.7.1 The Gaussian Pyramid

4.7.2 Applications of Scaled Representations

4.8 NOTES

5 Local Image Features

5.1 COMPUTING THE IMAGE GRADIENT

5.1.1 Derivative of Gaussian Filters

5.2 REPRESENTING THE IMAGE GRADIENT

5.2.1 Gradient-Based Edge Detectors

5.2.2 Orientations

5.3 FINDING CORNERS AND BUILDING NEIGHBORHOODS

5.3.1 Finding Corners

5.3.2 Using Scale and Orientation to Build a Neighborhood

5.4 DESCRIBING NEIGHBORHOODS WITH SIFT AND HOG FEATURES

5.4.1 SIFT Features

5.4.2 HOG Features

5.5 COMPUTING LOCAL FEATURES IN PRACTICE

5.6 NOTES

6 Texture

6.1 LOCAL TEXTURE REPRESENTATIONS USING FILTERS

6.1.1 Spots and Bars

6.1.2 From Filter Outputs to Texture Representation

6.1.3 Local Texture Representations in Practice

6.2 POOLED TEXTURE REPRESENTATIONS BY DISCOVERING TEXTONS

6.2.1 Vector Quantization and Textons

6.2.2 K-means Clustering for Vector Quantization

Table of Contents

6.3 SYNTHESIZING TEXTURES AND FILLING HOLES IN IMAGES

6.3.1 Synthesis by Sampling Local Models

6.3.2 Filling in Holes in Images

6.4 IMAGE DENOISING

6.4.1 Non-local Means

6.4.2 Block Matching 3D (BM3D)

6.4.3 Learned Sparse Coding

6.4.4 Results

6.5 SHAPE FROM TEXTURE

6.5.1 Shape from Texture for Planes

6.5.2 Shape from Texture for Curved Surfaces

6.6 NOTES

III EARLY VISION: MULTIPLEIMAGES

7 Stereopsis

7.1 BINOCULAR CAMERA GEOMETRY AND THE EPIPOLAR CONSTRAINT

7.1.1 Epipolar Geometry

7.1.2 The Essential Matrix

7.1.3 The Fundamental Matrix

7.2 BINOCULAR RECONSTRUCTION

7.2.1 Image Rectification

7.3 HUMAN STEREOPSIS

7.4 LOCAL METHODS FOR BINOCULAR FUSION

7.4.1 Correlation

7.4.2 Multi-Scale Edge Matching

7.5 GLOBAL METHODS FOR BINOCULAR FUSION

7.5.1 Ordering Constraints and Dynamic Programming

7.5.2 Smoothness Constraints and Combinatorial Optimization over Graphs

7.6 USING MORE CAMERAS

7.7 APPLICATION: ROBOT NAVIGATION

7.8 NOTES

8 Structure from Motion

8.1 INTERNALLY CALIBRATED PERSPECTIVE CAMERAS

8.1.1 Natural Ambiguity of the Problem

8.1.2 Euclidean Structure and Motion from Two Images

8.1.3 Euclidean Structure and Motion from Multiple Images

Table of Contents

8.2 UNCALIBRATED WEAK-PERSPECTIVE CAMERAS

- 8.2.1 Natural Ambiguity of the Problem
- 8.2.2 Affine Structure and Motion from Two Images
- 8.2.3 Affine Structure and Motion from Multiple Images
- 8.2.4 From Affine to Euclidean Shape

8.3 UNCALIBRATED PERSPECTIVE CAMERAS

- 8.3.1 Natural Ambiguity of the Problem
- 8.3.2 Projective Structure and Motion from Two Images
- 8.3.3 Projective Structure and Motion from Multiple Images
- 8.3.4 From Projective to Euclidean Shape

8.4 NOTES

IV MID-LEVEL VISION

9 Segmentation by Clustering

9.1 HUMAN VISION: GROUPING AND GESTALT

9.2 IMPORTANT APPLICATIONS

- 9.2.1 Background Subtraction
- 9.2.2 Shot Boundary Detection
- 9.2.3 Interactive Segmentation
- 9.2.4 Forming Image Regions

9.3 IMAGE SEGMENTATION BY CLUSTERING PIXELS

- 9.3.1 Basic Clustering Methods
- 9.3.2 The Watershed Algorithm
- 9.3.3 Segmentation Using K-means
- 9.3.4 Mean Shift: Finding Local Modes in Data
- 9.3.5 Clustering and Segmentation with Mean Shift

9.4 SEGMENTATION, CLUSTERING, AND GRAPHS

- 9.4.1 Terminology and Facts for Graphs
- 9.4.2 Agglomerative Clustering with a Graph
- 9.4.3 Divisive Clustering with a Graph
- 9.4.4 Normalized Cuts

9.5 IMAGE SEGMENTATION IN PRACTICE

- 9.5.1 Evaluating Segmenters

9.6 NOTES

10 Grouping and Model Fitting

10.1 THE HOUGH TRANSFORM

Table of Contents

10.1.1	Fitting Lines with the Hough Transform
10.1.2	Using the Hough Transform
10.2	FITTING LINES AND PLANES
10.2.1	Fitting a Single Line
10.2.2	Fitting Planes
10.2.3	Fitting Multiple Lines
10.3	FITTING CURVED STRUCTURES
10.4	Robustness
10.4.1	M-Estimators
10.4.2	RANSAC: Searching for Good Points
10.5	FITTING USING PROBABILISTIC MODELS
10.5.1	Missing Data Problems
10.5.2	Mixture Models and Hidden Variables
10.5.3	The EM Algorithm for Mixture Models
10.5.4	Difficulties with the EM Algorithm
10.6	MOTION SEGMENTATION BY PARAMETER ESTIMATION
10.6.1	Optical Flow and Motion
10.6.2	Flow Models
10.6.3	Motion Segmentation with Layers
10.7	MODEL SELECTION: WHICH MODEL IS THE BEST FIT?
10.7.1	Model Selection Using Cross-Validation
10.8	NOTES
11	Tracking
11.1	SIMPLE TRACKING STRATEGIES
11.1.1	Tracking by Detection
11.1.2	Tracking Translations by Matching
11.1.3	Using Affine Transformations to Confirm a Match
11.2	TRACKING USING MATCHING
11.2.1	Matching Summary Representations
11.2.2	Tracking Using Flow
11.3	TRACKING LINEAR DYNAMICAL MODELS WITH KALMAN FILTERS
11.3.1	Linear Measurements and Linear Dynamics
11.3.2	The Kalman Filter
11.3.3	Forward-backward Smoothing
11.4	DATA ASSOCIATION
11.4.1	Linking Kalman Filters with Detection Methods

Table of Contents

11.4.2 Key Methods of Data Association

11.5 PARTICLE FILTERING

11.5.1 Sampled Representations of Probability Distributions

11.5.2 The Simplest Particle Filter

11.5.3 The Tracking Algorithm

11.5.4 A Workable Particle Filter

11.5.5 Practical Issues in Building Particle Filters

11.6 NOTES

V HIGH-LEVEL VISION

12 Registration

12.1 REGISTERING RIGID OBJECTS

12.1.1 Iterated Closest Points

12.1.2 Searching for Transformations via Correspondences

12.1.3 Application: Building Image Mosaics

12.2 MODEL-BASED VISION: REGISTERING RIGID OBJECTS WITH PROJECTION

12.2.1 Verification: Comparing Transformed and Rendered Source to Target

12.3 REGISTERING DEFORMABLE OBJECTS

12.3.1 Deforming Texture with Active Appearance Models

12.3.2 Active Appearance Models in Practice

12.3.3 Application: Registration in Medical Imaging Systems

12.4 NOTES

13 Smooth Surfaces and Their Outlines

13.1 ELEMENTS OF DIFFERENTIAL GEOMETRY

13.1.1 Curves

13.1.2 Surfaces

13.2 CONTOUR GEOMETRY

13.2.1 The Occluding Contour and the Image Contour

13.2.2 The Cusps and Inflections of the Image Contour

13.2.3 Koenderinks Theorem

13.3 VISUAL EVENTS: MORE DIFFERENTIAL GEOMETRY

13.3.1 The Geometry of the Gauss Map

13.3.2 Asymptotic Curves

13.3.3 The Asymptotic Spherical Map

13.3.4 Local Visual Events

13.3.5 The Bitangent Ray Manifold

13.3.6 Multilocal Visual Events

Table of Contents

13.3.7 The Aspect Graph

13.4 NOTES

14 Range Data

14.1 ACTIVE RANGE SENSORS

14.2 RANGE DATA SEGMENTATION

14.2.1 Elements of Analytical Differential Geometry

14.2.2 Finding Step and Roof Edges in Range Images

14.2.3 Segmenting Range Images into Planar Regions

14.3 RANGE IMAGE REGISTRATION AND MODEL ACQUISITION

14.3.1 Quaternions

14.3.2 Registering Range Images

14.3.3 Fusing Multiple Range Images

14.4 OBJECT RECOGNITION

14.4.1 Matching Using Interpretation Trees

14.4.2 Matching Free-Form Surfaces Using Spin Images

14.5 KINECT

14.5.1 Features

14.5.2 Technique: Decision Trees and Random Forests

14.5.3 Labeling Pixels

14.5.4 Computing Joint Positions

14.6 NOTES

15 Learning to Classify

15.1 CLASSIFICATION, ERROR, AND LOSS

15.1.1 Using Loss to Determine Decisions

15.1.2 Training Error, Test Error, and Overfitting

15.1.3 Regularization

15.1.4 Error Rate and Cross-Validation

15.1.5 Receiver Operating Curves

15.2 MAJOR CLASSIFICATION STRATEGIES

15.2.1 Example: Mahalanobis Distance

15.2.2 Example: Class-Conditional Histograms and Naive Bayes

15.2.3 Example: Classification Using Nearest Neighbors

15.2.4 Example: The Linear Support Vector Machine

15.2.5 Example: Kernel Machines

15.2.6 Example: Boosting and Adaboost

15.3 PRACTICAL METHODS FOR BUILDING CLASSIFIERS

Table of Contents

- 15.3.1 Manipulating Training Data to Improve Performance
- 15.3.2 Building Multi-Class Classifiers Out of Binary Classifiers
- 15.3.3 Solving for SVMs and Kernel Machines

15.4 NOTES

16 classifying Images

16.1 BUILDING GOOD IMAGE FEATURES

- 16.1.1 Example Applications
- 16.1.2 Encoding Layout with GIST Features
- 16.1.3 Summarizing Images with Visual Words
- 16.1.4 The Spatial Pyramid Kernel
- 16.1.5 Dimension Reduction with Principal Components
- 16.1.6 Dimension Reduction with Canonical Variates
- 16.1.7 Example Application: Identifying Explicit Images
- 16.1.8 Example Application: Classifying Materials
- 16.1.9 Example Application: Classifying Scenes

16.2 CLASSIFYING IMAGES OF SINGLE OBJECTS

- 16.2.1 Image Classification Strategies
- 16.2.2 Evaluating Image Classification Systems
- 16.2.3 Fixed Sets of Classes
- 16.2.4 Large Numbers of Classes
- 16.2.5 Flowers, Leaves, and Birds: Some Specialized Problems

16.3 IMAGE CLASSIFICATION IN PRACTICE

- 16.3.1 Codes for Image Features
- 16.3.2 Image Classification Datasets
- 16.3.3 Dataset Bias
- 16.3.4 Crowdsourcing Dataset Collection

16.4 NOTES

17 Detecting Objects in Images

17.1 THE SLIDING WINDOW METHOD

- 17.1.1 Face Detection
- 17.1.2 Detecting Humans
- 17.1.3 Detecting Boundaries

17.2 DETECTING DEFORMABLE OBJECTS

17.3 THE STATE OF THE ART OF OBJECT DETECTION

- 17.3.1 Datasets and Resources

17.4 NOTES

Table of Contents

18 Topics in Object Recognition

18.1 WHAT SHOULD OBJECT RECOGNITION DO?

- 18.1.1 What Should an Object Recognition System Do?
- 18.1.2 Current Strategies for Object Recognition
- 18.1.3 What Is Categorization?
- 18.1.4 Selection: What Should Be Described?

18.2 FEATURE QUESTIONS

- 18.2.1 Improving Current Image Features
- 18.2.2 Other Kinds of Image Feature

18.3 Geometric Questions

18.4 SEMANTIC QUESTIONS

- 18.4.1 Attributes and the Unfamiliar
- 18.4.2 Parts, Poselets and Consistency
- 18.4.3 Chunks of Meaning

VI APPLICATIONS AND TOPICS

19 Image-Based Modeling and Rendering

19.1 VISUAL HULLS

- 19.1.1 Main Elements of the Visual Hull Model
- 19.1.2 Tracing Intersection Curves
- 19.1.3 Clipping Intersection Curves
- 19.1.4 Triangulating Cone Strips
- 19.1.5 Results
- 19.1.6 Going Further: Carved Visual Hulls

19.2 PATCH-BASED MULTI-VIEW STEREOPSIS

- 19.2.1 Main Elements of the PMVS Model
- 19.2.2 Initial Feature Matching
- 19.2.3 Expansion
- 19.2.4 Filtering
- 19.2.5 Results

19.3 THE LIGHT FIELD

19.4 NOTES

20 Looking at People

20.1 HMMS, DYNAMIC PROGRAMMING, AND TREE-STRUCTURED MODELS

- 20.1.1 Hidden Markov Models
- 20.1.2 Inference for an HMM

Table of Contents

20.1.3 Fitting an HMM with EM

20.1.4 Tree-Structured Energy Models

20.2 PARSING PEOPLE IN IMAGES

20.2.1 Parsing with Pictorial Structure Models

20.2.2 Estimating the Appearance of Clothing

20.3 TRACKING PEOPLE

20.3.1 Why Human Tracking Is Hard

20.3.2 Kinematic Tracking by Appearance

20.3.3 Kinematic Human Tracking Using Templates

20.4 3D FROM 2D: LIFTING

20.4.1 Reconstruction in an Orthographic View

20.4.2 Exploiting Appearance for Unambiguous Reconstructions

20.4.3 Exploiting Motion for Unambiguous Reconstructions

20.5 ACTIVITY RECOGNITION

20.5.1 Background: Human Motion Data

20.5.2 Body Configuration and Activity Recognition

20.5.3 Recognizing Human Activities with Appearance Features

20.5.4 Recognizing Human Activities with Compositional Models

20.6 RESOURCES

20.7 NOTES

21 Image Search and Retrieval

21.1 THE APPLICATION CONTEXT

21.1.1 Applications

21.1.2 User Needs

21.1.3 Types of Image Query

21.1.4 What Users Do with Image Collections

21.2 BASIC TECHNOLOGIES FROM INFORMATION RETRIEVAL

21.2.1 Word Counts

21.2.2 Smoothing Word Counts

21.2.3 Approximate Nearest Neighbors and Hashing

21.2.4 Ranking Documents

21.3 IMAGES AS DOCUMENTS

21.3.1 Matching Without Quantization

21.3.2 Ranking Image Search Results

21.3.4 Laying Out Images for Browsing

21.4 PREDICTING ANNOTATIONS FOR PICTURES

Table of Contents

- 21.4.1 Annotations from Nearby Words
- 21.4.2 Annotations from the Whole Image
- 21.4.3 Predicting Correlated Words with Classifiers
- 21.4.4 Names and Faces
- 21.4.5 Generating Tags with Segments

21.5 THE STATE OF THE ART OF WORD PREDICTION

- 21.5.1 Resources
- 21.5.2 Comparing Methods
- 21.5.3 Open Problems

21.6 NOTES

VII BACKGROUND MATERIAL

22 Optimization Techniques

22.1 LINEAR LEAST-SQUARES METHODS

- 22.1.1 Normal Equations and the Pseudoinverse
- 22.1.2 Homogeneous Systems and Eigenvalue Problems
- 22.1.3 Generalized Eigenvalues Problems
- 22.1.4 An Example: Fitting a Line to Points in a Plane
- 22.1.5 Singular Value Decomposition

22.2 NONLINEAR LEAST-SQUARES METHODS

- 22.2.1 Newtons Method: Square Systems of Nonlinear Equations.
- 22.2.2 Newtons Method for Overconstrained Systems
- 22.2.3 The GaussNewton and LevenbergMarquardt Algorithms

22.3 SPARSE CODING AND DICTIONARY LEARNING

- 22.3.1 Sparse Coding
- 22.3.2 Dictionary Learning
- 22.3.3 Supervised Dictionary Learning

22.4 MIN-CUT/MAX-FLOW PROBLEMS AND COMBINATORIAL OPTIMIZATION

- 22.4.1 Min-Cut Problems
- 22.4.2 Quadratic Pseudo-Boolean Functions
- 22.4.3 Generalization to Integer Variables

22.5 NOTES

Bibliography

Index

List of Algorithms