Perception basics for learning by experimentation

Matthias Schlemmer, Markus Vincze, Peter Gemeiner, Johann Prankl, Markus Bader
Overview

• Perception – what for? And how?
• State of the Art
  – Object detection / recognition
  – Gestalt based grouping
  – 3D
• Tools for XPERO – Learning by Experimentation
  – Colour blob recogniser
  – Stereo Gestalt based grouping tool
  – Ontology based grouping
  – [Self-localisation]
Overview

• **Perception – what for? And how?**
• **State of the Art**
  – Object detection / recognition
  – Gestalt based grouping
  – 3D
• **Tools for XPERO – Learning by Experimentation**
  – Colour blob recogniser
  – Stereo Gestalt based grouping tool
  – Ontology based grouping
  – [Self-localisation]
Tasks of Perception

• Detect objects of interest
  – Objects? Rather collection of primitives
  – Primitives or features
    • Interest point
    • Line, junction, parallel lines, rectangle, ...
    • Arc, ellipse
    • Surface patches
  – Result: object location, object identification

• Orientation in space, Localisation
  – Features just as above
  – Result: some coordinates or relationship
Perception, Options

- Odometry
- Distance sensors
  - Sonar
  - Infrared
  - Laser scanner
- Touch („Bumper“)
- Visual perception
  - Monocular vision, one camera
  - Stereo vision, two or three cameras
  - Direct range images: laser scans, time-of-flight
PART I: Basics of (artificial) visual perception
Overview

- Perception – what for? And how?
- **State of the Art**
  - Object detection / recognition
  - Gestalt-based grouping
  - 3D
- **Tools for XPERO – Learning by Experimentation**
  - Colour blob recogniser
  - Stereo Gestalt based grouping tool
  - Ontology based grouping
  - [Self-localisation]
Object Detection and Recognition

• Purely model and geometry driven
  – CAD model of object, whole environment

• Purely learning (model-free) and appearance based
  – Typically: modelling/learning of objects
  – Interest points or „whole“ object

• Mixture: structure in data – Gestalt principles
  – Model physics of world and imaging process (rather than objects)
  – Features, perceptual grouping
Object Detection and Recognition

• Purely model and geometry driven
  – CAD model of object, whole environment

• Purely learning (model-free) and appearance based
  – Typically: modelling/learning of objects
  – Interest points or „whole“ object

• Mixture: structure in data – Gestalt principles
  – Model physics of world and imaging process (rather than objects)
  – Features, perceptual grouping
Objects and Use of Interest Points

- Extraction of interest points (characteristic locations)
- Computation of local descriptors
- Determining correspondences
- Detect similar image parts (objects)
Extraction of Interest Points

• Corner detectors
  – Harris, Hessian

• Multi-scale corner detectors (with scale selection)
  – Scale invariant Harris and Hessian corners
  – Difference of Gaussian (DoG) (Lowe)

• Affine covariant regions
  – Harris-Affine (Mikolajczyk, Schmid ‘02, Schaffalitzky, Zisserman ’02)
  – Hessian-Affine (Mikolajczyk and Schmid ’02)
  – Maximally stable extremal regions (MSER) (Matas et al. ’02)
  – Intensity based regions (IBR) (Tuytelaars and Van Gool ’00)
  – Edge based regions (EBR) (Tuytelaars and Van Gool ’00)
  – Entropy-based regions (salient regions) (Kadir et al. ’04)
Scale invariant Harris points

- Multi-scale extraction of Harris interest points
- Selection of points at characteristic scale in scale space

Characteristic scale:
- Maximum in scale space
- Scale invariant

[Mikolajczyk 04]
Affine covariant regions - Motivation

[Mikolajczyk 04]
Computation of Local Descriptors

- Distinctive
- Robust
- Invariant to geometric & photometric transformation
- Descriptors
  - Sampled image patch
  - Gradient orientation histogram – SIFT (Lowe)
  - Shape context (Belongie et al. ’02)
  - PCA-SIFT (Ke and Sukthankar ’04)
  - Moment invariants (Van Gool ’96)
  - Gaussian derivative-based (Koenderink ’87, Freeman ’91)
  - Complex filters (Baumberg ’00, Schaffalitzky and Zisserman ’02)
Gradient orientation histogram (SIFT)

- Thresholded image gradients are sampled over 16x16 array of locations in scale space
- Create array of orientation histograms
- 8 orientations x 4 x 4 histogram array = 128 dimensions

[Lowe 04] Image gradients  Keypoint descriptor
Interest points can be used for ...

• Object recognition
• Object recognition and segmentation
• Robot Localization
• Tracking
Planar Recognition

- Planar surfaces can be reliably recognized at a rotation of 60° away from the camera
- Affine fit approximates perspective projection
- Only 3 points are needed for recognition
  → Cope with occlusion

[Lowe]
Robot Localization

[Se 05]
Interest Point Tracking and Occlusion Reasoning

- Grouping KLT features based on motion
- Detect occlusion based on appearance and disappearance of interest points
Object Detection and Recognition

- Purely model and geometry driven
  - CAD model of object, whole environment
- Purely learning (model-free) and appearance based
  - Typically: modelling/learning of objects
  - Interest points or "whole" object
- Mixture: structure in data – Gestalt principles
  - Model physics of world and imaging process (rather than objects)
  - Features, perceptual grouping
Appearance-based Object Recognition

- Training with segmented images
- Representation in high dimensional or reduced (Principal Component Analysis PCA) space
- Separate objects linear or non-linear (kernel methods, SVM)
- Challenges
  - Illumination, Scale, Occlusion

[Bischof, Summerschool 2005]
Object Recognition using SVM

- Approximate 200 trainings images / object (RGB, different views, different light)
- Background trainings images
- Hyperspace with 3072 dimensions
- Iterative calculation of separating surface between two classes of objects

[Zillich 01]
Overview

• Perception – what for? And how?

• State of the Art
  – Object detection / recognition
  – Gestalt based grouping
  – 3D

• Tools for XPERO – Learning by Experimentation
  – Colour blob recogniser
  – Stereo Gestalt based grouping tool
  – Ontology based grouping
  – [Self-localisation]
Comparison classical object recognition vs. vs2

• From Features to symbols
• Grouping of non-accidental edgels to Gestalts
Hierarchical Grouping

- Abstraction from Gestalts to Object shapes
- Learning is possible from features at every level
Perceptual Grouping

- No learning of specific objects!
- Bottom-up approach
- Uses
  - Gestalt principles
  - built-in knowledge
  - levels of abstraction
- Goal not to mimic human vision, exploit all the known structure in the data
Perceptual Grouping – State of the Art

- 3D Object Recognition from single 2D images [Lowe 1987]
- Integration of regions and contours for object recognition [Schlüter 2000]
- A computational structure for preattentive perceptual organization [Sarkar 1994]
3D Object Recognition using Perceptual Grouping

- Perceptual organization to form groupings and structures in image
  - Proximity
  - Parallelism
  - Collinearity
- Probabilistic ranking
- Projection of three dimensional objects

[Low 87]
Perceptual Grouping Steps

• Edge detection [Canny]
• Lines, arcs [Rosin, West 95]
• L- and T-junctions, parallel lines, ellipses
• Higher level shapes, objects
• Perceptual grouping + symbolic reasoning (Prolog)
Cylinder

cyl(E,A,B) :- tangent_l(E,A), tangent_r(E,B), parallel(A,B).
Cup Scene

Original image  Edges  First cylinder
Kitchen Scene

Original image

First 3 cylinders
Rectangle

rect(A, B, C, D) := \text{line}(A, B), \text{line}(B, C), \text{line}(C, D), \text{line}(D, A)

rect(A, B, C, D) := u(A, B, C), u(C, D, A)
Book Scene

Original image

First 2 rectangles

3847 edges

All rectangles
Office Scene
Perceptual Grouping – runtime considerations

- Simple edge detector fails to "see" square
- Square can be assembled with "growing" lines
- → Runtime decides which objects are found
- → More obvious shapes are found first, other ones later, which results in Ranking
Conclusion on Grouping

- Exploit local information (smoothness, ...)
- Avoid local decisions and early pruning of hypotheses (avoid thresholds)
  → Use ranking
- Perceptual grouping = well-defined pruning of hypotheses → reduce search space
- Image Space Indexing: $O(n^2) \rightarrow O(n)$
- Same method to detect different shapes

Overview

- Perception – what for? And how?
- **State of the Art**
  - Object detection / recognition
  - Gestalt based grouping
  - 3D
- Tools for XPERO – Learning by Experimentation
  - Colour blob recogniser
  - Stereo Gestalt based grouping tool
  - Ontology based grouping
  - [Self-localisation]
Range (3D) Vision Sensors

- Laser Range Scanner
- Stereo Vision System
- Time of Flight Sensor

→ Acquire point cloud of environment
3D Laser Range Scanner
Time of Flight Sensor

[CSEM]
Plane Detection

[Murray 04]
Model Fitting: Geometric Primitives
Model Fitting: Superquadrics

- Recover and Select Paradigm

[Leonardis 97]
Example: Table Scene

- Objects learned from one scan
- Detection in one view, 2 sec.
PART II: Tools used for robot experimentation in XPERO
Perception software used

- Recording data streams of
  - Robot’s motion and perception
    - Odometry - egomotion
    - Stereo cameras – object location and size ("vs2-stereo")
  - Ground truth ("ACIN_TrackColor")
    - Position and orientation of robot w.r.t world
    - Information for the human observer
## Movability experiment

<table>
<thead>
<tr>
<th>timestamp</th>
<th>odo_x</th>
<th>odo_y</th>
<th>odo_ori</th>
<th>obj_area</th>
<th>obj_dist</th>
<th>obj_ang</th>
<th>obj_rad</th>
</tr>
</thead>
<tbody>
<tr>
<td>116351578.4108</td>
<td>111.51</td>
<td>-0.826</td>
<td>6.25872</td>
<td>5650.9296</td>
<td>1395.7987</td>
<td>0.193378</td>
<td>53.249889</td>
</tr>
<tr>
<td>116351578.676759</td>
<td>111.51</td>
<td>-0.826</td>
<td>6.25872</td>
<td>5535.5274</td>
<td>1398.3284</td>
<td>0.193408</td>
<td>54.519947</td>
</tr>
<tr>
<td>116351578.936725</td>
<td>111.51</td>
<td>-0.826</td>
<td>6.25872</td>
<td>5544.1685</td>
<td>1400.9954</td>
<td>0.193448</td>
<td>53.748249</td>
</tr>
<tr>
<td>116351579.203685</td>
<td>113.162</td>
<td>-0.826</td>
<td>6.269458</td>
<td>5534.3393</td>
<td>1402.9049</td>
<td>0.191451</td>
<td>54.603863</td>
</tr>
<tr>
<td>116351579.457643</td>
<td>113.162</td>
<td>-0.826</td>
<td>6.269458</td>
<td>5547.6424</td>
<td>1405.5694</td>
<td>0.187391</td>
<td>54.575568</td>
</tr>
<tr>
<td>116351579.72161</td>
<td>113.162</td>
<td>-0.826</td>
<td>0.009204</td>
<td>5490.6025</td>
<td>1401.8033</td>
<td>0.180236</td>
<td>54.419564</td>
</tr>
<tr>
<td>116351579.980559</td>
<td>113.162</td>
<td>-0.826</td>
<td>0.009204</td>
<td>5491.5522</td>
<td>1415.194</td>
<td>0.162945</td>
<td>53.572767</td>
</tr>
<tr>
<td>116351580.237518</td>
<td>112.336</td>
<td>-0.826</td>
<td>0.052156</td>
<td>5412.4416</td>
<td>1422.0342</td>
<td>0.140015</td>
<td>53.446098</td>
</tr>
<tr>
<td>116351580.498476</td>
<td>111.51</td>
<td>-0.826</td>
<td>0.072098</td>
<td>5364.4796</td>
<td>1285.2201</td>
<td>0.119119</td>
<td>48.710912</td>
</tr>
<tr>
<td>116351580.755462</td>
<td>111.51</td>
<td>-0.826</td>
<td>0.072098</td>
<td>5281.8925</td>
<td>1247.5526</td>
<td>0.099671</td>
<td>46.188951</td>
</tr>
</tbody>
</table>
I. Bratko, et al., „Initial experiments in robot discovery in XPERO“, IEEE ICRA Workshop on Concept Learning for Embodied Agents“, April 2007
Overview

• Perception – what for? And how?
• State of the Art
  – Object detection / recognition
  – Gestalt based grouping
  – 3D
• Tools for XPERO – Learning by Experimentation
  – Colour blob recogniser
  – Stereo Gestalt based grouping tool
  – Ontology based grouping
  – [Self-localisation]
Colour Blob Recogniser

- Tool to detect colour blobs adapted to local illumination conditions with reference pattern
- Calibrated camera → computation of blob location
  → Localisation of robot (tracking at frame rate)
Camera Calibration CamCalb

- Calibrated camera $\rightarrow$ computation of blob location
Overview

- Perception – what for? And how?
- State of the Art
  - Object detection / recognition
  - Gestalt based grouping
  - 3D
- **Tools for XPERO – Learning by Experimentation**
  - Colour blob recogniser
  - **Stereo Gestalt based grouping tool**
  - Ontology based grouping
  - [Self-localisation]
XPERO Need: Extension of vs2 to Stereo

- Detect Gestalts in 2 images
- Compute epipolar geometry
- \(\rightarrow\) Gestalts in 3 dimensional space
- Allows, e.g., computation of distance to ball
  - Radius of ball, if several balls would be in scene
Overview

• Perception – what for? And how?
• State of the Art
  – Object detection / recognition
  – Gestalt based grouping
  – 3D
• **Tools for XPERO – Learning by Experimentation**
  – Colour blob recogniser
  – Stereo Gestalt based grouping tool
  – **Ontology based grouping**
  – [Self-localisation]
Ontology-based grouping

• From Features to Objects
  – Proto-objects = collection of connected features
  – Object = n proto-objects, \( n=1 \ldots \infty \)
  – Ellipse = ball or egg or …

• Simplifying learning
  – Easier to learn from higher level entities

• Relations between objects
  – Notion of objects, movable, pushable,…
Ontology-based grouping

- From “proto-objects” (vs2) to “objects” via ontology
- No fixed appearance or shape, only “concept” (=necessary conditions, i.e. structure,...) in ontology
- Actual appearance is tested for compliance
- Semantic content
- Up to now: only spatial relations („left_of“, „on_top_of“, etc.)
Ontology-based grouping

- Searching for object concepts in images
- Trying to fulfill a “task” (Where is an arch?)
- $\rightarrow$ gradual abstraction

$\rightarrow$ from qualitative relations to quantitative location
Conclusion

• Object recognition
  – Interest Points
  – „whole object“ appearance based (learning)
  – Gestalt-based (Perceptual Grouping)

• 3D Laser Scanning
  – Object recognition (by components/fitting)

• Tools for Robot‘s Experiments
  – Simple Colour tracking
  – Vs2-stereo
Thank you for your attention!