Exploiting Monge Properties in Optimum Subwindow Search

Senjian An, P. Peursum, W. Liu, S. Venkatesh and X. Chen

Curtin University of Technology, Australia
Exploiting Monge Structures in Optimum Subwindow Search

- **Optimum Subwindow Search**
  - Appears in object detection using sliding window methods
  - Can be formulated as finding the maximum entry of a 4-D array
  - Exhaustive search takes $O(n^4)$ time for $n$ by $n$ images

- We show that, if the quality function is submodular i.e., satisfying “diminishing return” property,
  - The 4-D array has some 2-D Monge subarrays.
  - By using the Monge property, the time complexity is reduced to be cubic ($O(n^3)$).
  - An approximate quadratic ($O(n^2)$) time algorithm is also developed.

Welcome to See Our Poster!
Unified Tracking and Recognition with Rotation-Invariant Fast Features

Gabriel Takacs, Vijay Chandrasekhar, Sam Tsai, David Chen, Radek Grzeszczuk, Bernd Girod
Unified Tracking and Recognition with Rotation-Invariant Fast Features

Real-time feature extraction on a smart-phone
Same features for both tracking and recognition
Fast Polygonal Integration
And Its Application in Extending Haar-like Features to Improve Object Detection

M.T. Pham, Y.Gao, V.D.D.Hoang, T.J. Cham
Fast Polygonal Integration

Motivation:
- To quickly integrate over a 2D polygonal region that is not necessarily \textit{rectilinear}.

Generalization of integral image: from 2 slopes (i.e. rectilinear) to many slopes
- Integration time: $O(T)$, where $T$ = the number of vertices
- Source code: \url{http://code.google.com/p/pyopencv/}

Extension of Haar-like features for Object Detection:
Poster Spotlights

Session: Object Recognition II, Tue 15 June 2010, 3:40-5:20 pm

Object Detection via Boundary Structure Segmentation

Alexander Toshev, Ben Taskar, Kostas Daniilidis
Problem: Shape representation for object detection/segmentation.

Major contributions:

1. Chordiogram: global / holistic shape descriptor

2. Coupling of chordiogram with segment and contour grouping

3. Simultaneous single-step chordiogram matching and object segmentation

Improvements over state-of-the-art on ETHZ Shape Dataset and INRIA Horses.
Implicit Hierarchical Boosting for Multi-view Object Detection

Xavier Perrotton, Marc Sturzel – EADS
Michel Roux – Telecom ParisTech
Implicit Hierarchical Boosting for Multi-view Object Detection

Multi-view object detection

- Boosting scheme
- Single AdaBoost cascade
- No pose annotation required

Benchmarks on Pedestrian INRIA, Pascal’06, ...

- Weak partition concept
  - Auto-adaptative to classification errors
  - Feature-based temporary segmentation of positive samples

- Implicit Hierarchical structure
  - Decreasing pattern
  - Adjusting complexity
Connecting Modalities: Semi-supervised Segmentation and Annotation of Images Using Unaligned Text Corpora

Richard Socher, Fei-Fei Li
Connecting Modalities

- Training with New York Times text and Flickr images
- Only 5 training images
- State of the art annotation performance
- Good segmentation performance

- Kernel Canonical Correlation Analysis
Support Vector Regression for Multi-View Gait Recognition based on Local Motion Feature Selection

Worapan Kusakunniran, Qiang Wu, Jian Zhang, Hongdong Li
Challenges and New Solutions to Multi-view Gait Recognition

- **Advantages:** an efficient biometric for recognising a person in a distance
- **Challenges:** performance drops when view angle changes.
- **3 general approaches:** for multi-view gait recognition
  - Approach based on 3D camera setting
  - Approach based on view invariant feature
  - Approach based on View Transformation Model (VTM) – *(Our new method is proposed in this paper in this category.)*

- **New method:**
  - Idea: VTM constructed based on local motion regression using **Support Vector Regression (SVR)**, where Gait Energy Image (GEI) is used as gait feature.
  - Experiments: extensive experiments are carried out on benchmark multi-view CASIA gait database B (total 124 subjects under 11 viewing angels)
  - Conclusions: Performance significantly improved when compared with other existing methods of the same category.

- **Reference:**
Integrated Pedestrian Classification and Orientation Estimation

Markus Enzweiler and Dariu M. Gavrila
- Pedestrian classification and orientation estimation from single 2D images
- We use the same model for both tasks
- Pose-specific Mixture-of-Experts model with
  - shape-based orientation priors
  - texture-based expert classifiers
- Approximation of continuous orientation density using GMM
- 50% less errors for both classification (reduction in false detections) and orientation estimation (difference between ground-truth and estimated orientation) vs. state-of-the-art
Poster Spotlights

Session: Object Recognition II, Tue 15 June 2010, 3:40-5:20 pm

Multi-Cue Pedestrian Classification with Partial Occlusion Handling

Markus Enzweiler, Angela Eigenstetter, Bernt Schiele and Dariu M. Gavrila
Multi-Cue Pedestrian Classification

- Most pedestrian classifiers assume full visibility of pedestrians
- They do not respond well to partially occluded pedestrians

Our approach
- Examines occlusion boundaries (discontinuities) in depth and motion space to derive visibility information of body components
- Visibility-related weights are used to fuse multi-cue (intensity, depth, motion) component-based classifiers in a Mixture-of-Experts framework
- Does not require special camera set-ups and is independent of the feature/classifier layout
  - Occlusion-handling reduces false positives at constant detection rates by factor of two
  - Multi-cue classification yields an additional factor of two in performance improvement
Poster Spotlights

Session: Object Recognition II, Tue 15 June 2010, 15:40-17:20 pm

Model Globally, Match Locally: Efficient and Robust 3D Object Recognition

Bertram Drost, Markus Ulrich, Nassir Navab, Slobodan Ilic
Model Globally, Match Locally: Efficient and Robust 3D Object Recognition

**Problems:**
- Free Form Objects
- Occlusion, Clutter, Noise
- Multiple Instances

**Our Approach:**
- Global model description using point pair features
- Local matching based on voting scheme

**Benefits:**
- Fast & Accurate
- Stable
Poster Spotlights

Session: Object Recognition II, Tue 15 June 2010, 3:40-5:20 pm

Visual Recognition and Detection under Bounded Computational Resources

Sudheendra Vijayanarasimhan and Ashish Kapoor
Visual Recognition and Detection under Bounded Computational Resources

Sudheendra Vijayanarasimhan and Ashish Kapoor

**Approach**

- Current approaches ignore computational constraints during recognition
- We propose a detection approach that actively determines both the **best location** and the **best feature** to extract given the **computational cost**

**Results**

- ETHZ/INRIA datasets
- **Anytime** behavior during detection
- Outperforms other Hough voting approaches

**Selected features**

- **Our Approach**
- **Passive approach**
Poster Spotlights

Session: Object Recognition II, Tue 15 June 2010, 3:40-5:20 pm

Talking Pictures:
Temporal Grouping and Dialog-Supervised Person Recognition

Timothee Cour, Benjamin Sapp, Akash Nagle, Ben Taskar
Talking Pictures: Temporal Grouping and Dialog-Supervised Person Recognition

Hey, Jack! What is it, Charlie?

- Temporal Grouping model
  - Tractable inference / learning
  - Incorporates high-order grouping relations

State of the art grouping results on 8 TV shows/movies

- Global weakly supervised learning
  - Incorporates: Grouping cues, Dialog cues, Gender cues
  - Hey, Jack!

Accuracy: 80% at 10% recall (resp. 60% at 50%) across 55 characters
An Efficient Divide-and-Conquer Cascade for Nonlinear Object Detection

Christoph H. Lampert (IST Austria)
Overall Topic: Efficient Object Detection in Natural Images

Sliding Window Paradigm: evaluate a classifier cascade for many windows

**Classical Cascade:** early stopping in stages, but exhaustive search in locations

**Proposed:** early stopping in stages, and divide-and-conquer in locations
New Features and Insights for Pedestrian Detection

Stefan Walk¹, Nikodem Majer¹, Konrad Schindler¹, Bernt Schiele¹,²

¹ TU Darmstadt  ² MPI Informatics, Saarbrücken
New Features and Insights for Pedestrian Detection

• **New Feature: Color Self-Similarity**

• Significant improvement (not only) on the Caltech Pedestrian Dataset in combination with motion features

• Which regions in the detection window are similar to each other?

**Insights:** Evaluation protocol matters! Training procedure matters as well!

---

![Graph showing miss rate vs. false positives per image for different methods and training rounds.](image)
Efficient Rotation Invariant Object Detection using Boosted Random Ferns

Michael Villamizar, Francesc Moreno-Noguer, Juan Andrade-Cetto, Alberto Sanfeliu
Efficient Rotation Invariant Object Detection using Boosted Random Ferns

Decoupled Approach: **Orientation Estimation** + **Object Classification**

**Boosted Random Ferns**
* Most discriminative Random Ferns computed in HOG space.
* They are common to all object orientations trained.

**Orientation Estimation**
* Boosted RFs used to compute initial class independent object pose hypotheses.
* Many false positives … with almost no false negatives.

**Object Classification**
* Orientation specific classifier is rotated and tested only over the initial hypotheses → very efficient!
* Competitive detection rates at reduced computational price.
Poster Spotlights

Session: Object Recognition II, Tue 15 June 2010, 15:40-17:20 pm

Fast and Robust Object Segmentation with the Integral Linear Classifier

D. Aldavert, A. Ramisa, R. Lopez de Mantaras and R. Toledo
Fast and Robust Object Segmentation with the Integral Linear Classifier

Densely Sample Descriptors

Bag-Of-Features Linear Classifier Weights

Where is the bike?

Final Segmentation

Multiple Segmentations at Different Scales

Source code at: http://www.cvc.uab.cat/~aldavert/plor

Up to 10 FPS
Poster Spotlights

Session: Object Recognition II, Tue 15 June 2010, 15:40-17:20 pm

Segmenting Video Into Classes of Algorithm-Suitability

O. Mac Aodha, G. J. Brostow, M. Pollefeys
Problem Statement
Given a scene and a set of algorithms, which one(s) should you apply to compute optical flow, or perform feature matching?

Method
Our hypothesis is that the most suitable algorithm can be chosen for each video automatically, through supervised training of a classifier.

Results
- Can you trust this optical flow?
- Can you trust this SIFT match?
Latent Hierarchical Structural Learning for Object Detection

L. Zhu, Y. Chen, A. Yuille, W. Freeman
What are good structures?

1. Deep is better than shallow
2. Simple grid structure is sufficient
3. Concave-Convex procedure (CCCP)
A Steiner tree approach to efficient object detection

Olga Russakovsky and Andrew Y. Ng
Stanford University
Goal: efficiently detect multiple object classes in a scene

(1) Classifier-agnostic object detection pipeline

(2) Steiner tree optimization framework

- shares pipeline parameters amongst object classes
- amortizes computational cost of proposing regions

10-15x faster detection

Olga Russakovsky and Andrew Y. Ng
Poster Spotlights

Session: Object Recognition II, Tue 15 June 2010, 3:40-5:20 pm

Cascaded Pose Regression
Piotr Dollár, Peter Welinder, Pietro Perona
Cascaded Pose Regression

training data

pose indexed features

training procedure

**Input**: Data \((I_i, \theta_i)\) for \(i = 1 \ldots N\)

1. \(\theta^0 = \arg \min_\theta \sum_i d(\theta, \theta_i)\)
2. \(\theta^0_i = \theta^0\) for \(i = 1 \ldots N\)
3. **for** \(t = 1\) **to** \(T\) **do**
4. \(x_i = h^t(\theta^{t-1}_i, I_i)\)
5. \(\tilde{\theta}_i = \theta^{t-1}_i \circ \theta_i\)
6. \(R^t = \arg \min_R \sum_i d(R(x_i), \tilde{\theta}_i)\)
7. \(\theta^t_i = \theta^{t-1}_i \circ R^t(x_i)\)
8. \(\epsilon_t = \sum_i d(\theta^t_i, \theta_i)/\sum_i d(\theta^{t-1}_i, \theta_i)\)
9. **if** \(\epsilon_t \geq 1\) **stop**
10. **end for**
11. **Output** \(R = (R^1, \ldots, R^T)\)

**typical (sorted) results**
Poster Spotlights

Session: Object Recognition II, Tue 15 June 2010, 15:40-17:20 pm

Free-Shape Subwindow Search for Object Localization

Zhiqi Zhang, Yu Cao, Dhaval Salvi, Kenton Oliver, Jarrell Waggoner, Song Wang
Motivation:

Localization using rectangle subwindow search

Localization using free-shape subwindow search

Localization rates for the VOC 2007 database:

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Proposed</td>
<td>ESS</td>
</tr>
<tr>
<td>dog</td>
<td>0.419</td>
<td>0.389</td>
</tr>
<tr>
<td>cat</td>
<td>0.433</td>
<td>0.422</td>
</tr>
<tr>
<td>sheep</td>
<td>0.132</td>
<td>0.095</td>
</tr>
<tr>
<td>cow</td>
<td>0.217</td>
<td>0.176</td>
</tr>
<tr>
<td>horse</td>
<td>0.398</td>
<td>0.388</td>
</tr>
</tbody>
</table>

Sample results:

Proposed Method

ESS
Improving web image search using query-relative classifiers

Josip Krapac, Moray Allan, Jakob Verbeek, Frédéric Jurie
Goal: re-rank images retrieved using a textual query (e.g. using Google Images) using visual content and textual meta-data

Most existing work
- Learns classifiers from noisily labeled examples (top search engine results)
- Models do not generalize to new queries; requires classifier training for each query

Our contributions
1. Definition of query-relative visual features, where feature meaning depends on the query
2. Learn single model (using image and meta-data)
   - Generalizes to new unseen queries
   - Even when image set is over 50% irrelevant
3. New annotated dataset of images and meta-data (for 353 search queries)