Incremental Learning of Object Detectors Using a Visual Shape Alphabet

A. Opelt, A. Pinz & A. Zisserman

CVPR '06

Presented by Medha Bhargava*

* Several slides adapted from authors' presentation, CVPR '06
OUTLINE

• Motivation, Goals & Overview of the approach

• Learning the model
  Stage 1: the visual alphabet of shape
  Stage 2: jointly/incrementally learned detectors

• Detection
  • Invariances (scale, rotation, viewpoint)
  • Experiments and Results
  • Summary
MOTIVATION

Class: Bicycle

Class: Person
MOTIVATION

Classification + Localization + rough Segmentation

Proposed approach uses

Alphabet of Shape

+ Geometry
GOALS

• Object detection

• Localization and crude segmentation

• Learning new models from previously trained detectors
  – Incremental learning
  – Sharing of model features ....underlying theme

• Sublinear learning complexity

NEW CONCEPTS

• Boundary fragment based shape alphabet

• Incremental joint-AdaBoost algorithm
**BOUNDARY-FRAGMENT-MODEL 1/2**

• **Learning** the BFM
  
  – *Training* set
    – Object delineated by bounding box
      – 20 images/class
  
  – *Validation* set
    – Labeled as +ve/-ve image
    – Object centroid marked
      – 50 images/class -- 25 +ve, 25 –ve

  – A candidate boundary fragment **MUST**
    – match edge chains in +ve set
    – have good localization of the centroid in +ve set
• **Scoring** a Boundary Fragment

\[ C(\gamma_i) = c_{\text{match}}(\gamma_i) c_{\text{loc}}(\gamma_i) \]

- \( c_{\text{match}}(\gamma_i) \): Ratio of cumulative Chamfer matching costs of fragment to edge chains in validation images

\[
c_{\text{match}}(\gamma_i) = \frac{\sum_{i=1}^{L} \text{distance}(\gamma_i, V_i^+)/L^+}{\sum_{i=1}^{L} \text{distance}(\gamma_i, V_i^-)/L^-}
\]

- \( c_{\text{loc}}(\gamma_i) \): pixel distance between true centroid and predicted centroid, averaged over +ve validation set
The Boundary-Fragment-Model

first proposed in Opelt, Pinz and Zisserman ECCV 2006

Geometric model related to Leibe, Leonardis and Schiele (Workshop at ECCV 2004)

Similar model proposed by Shotton et al. (ICCV 2005)

OVERVIEW OF THE APPROACH 1/2

More categories →

Two possibilities: Learning JOINTLY or INCREMENTALLY
STAGE 1: THE VISUAL ALPHABET OF SHAPE 1/3
STAGE 1: THE VISUAL ALPHABET OF SHAPE 1/3

For each class $C_i$

For $i=1:N$ trials

1. Grow candidate fragment in training images around random starting point $i$

2. Evaluate the fragment at each step on the validation set of the category
   - calculate costs

3. If the fragments costs are above a certain threshold discard this fragment, otherwise go on with step 4.
STAGE 1: THE VISUAL ALPHABET OF SHAPE 2/3

For each class $C_i$

For $i=1:N$ trials

1. ...

2. ...

3. ...

4. Evaluate the boundary fragment on the validation sets of the other categories.

5. Add this fragment with costs on all categories and the geometric information to the alphabet.

Horses

Faces

Update centroid vectors

COSTS
STAGE 1: THE VISUAL ALPHABET OF SHAPE 3/3

Clustering shape

Visual Shape Alphabet
INCREMENTAL LEARNING

• **Enlarging the alphabet codebook**
  1. Add more boundary fragments
  2. Allow a single fragment to vote for additional object centroids

• **Sharing to build**
  1. If fragments from different categories match, update centroid info
  2. Evaluation of fragment on –ve validation set
  3. Granting additional voting privileges
EXAMPLE OF SHARING

Over Classes

Over Aspects

Benefit: One class/aspect can build on what has been learnt from another
STAGE 2: WEAK DETECTOR CANDIDATES

Combinations of 2 boundary fragments as pool for learning

Matching $\gamma_a$ on the edge image

Overlap of centroid predictions

Matching $\gamma_b$ on the edge image

voting for same centroid

Calculated for ALL combinations on ALL validation sets
STAGE 2: JOINTLY LEARNED DETECTORS

Visual Shape Alphabet

Combinations of 2 alphabet entries form POOL of CANDIDATES

Validation Set

Class 1

Class 2

Class C

Background

Based on Torralba et al. CVPR 2004

JOINTBOOST

Each candidate that is valid (boosting) for at least one category is added to Collection of weak detectors

h(Horse, CarFront)
STAGE 2: INCREMENTALLY LEARNED DETECTORS

Knowledge:
Collection of weak detectors
e.g. CarsSide, Horses, Bicycles

1. Update existing knowledge (share)

2. Add new weak detectors (discriminative)
DETECTION FOR THE MULTICLASS CASE

Collection of votes in Hough voting space

Mode above threshold → Detection Class 3
INVARIANCES

- **Translation** → Mode search in the Hough voting space
- **In-plane Rotation** → Hough voting with oriented model
- **Scale invariance** → 3D-Balloon-Meanshift-Mode-Est.
- **Viewpoint** →

![Graph showing rotation influence on detection confidence for different objects.](image)
EXPERIMENTS
MULTICLASS DATASET

Collection of 17 categories, From Caltech, Graz02, Magee, ImageGoogle (available at: http://emt.tugraz.at/~pinz/data)

Different numbers of training images per category (10-100)

Different aspects and similar categories
RESULTS 1/6

Similarities at the alphabet level
RESULTS 2/6

Incremental vs. Joint-Boosting
RESULTS 3/6

Sharing of weak detectors
Examples of detection results
Examples of detection results
**RESULTS**

Detection results: Independent learning, Joint learning, one-class, multi-class

<table>
<thead>
<tr>
<th>Class</th>
<th>Plane</th>
<th>CarR</th>
<th>MBx</th>
<th>Face</th>
<th>B-S</th>
<th>B-E</th>
<th>B-F</th>
<th>Car23</th>
<th>CarF</th>
<th>Bottle</th>
<th>CowS</th>
<th>H-S</th>
<th>H-F</th>
<th>CowF</th>
<th>Pers</th>
<th>Mug</th>
<th>Cup</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ref.</td>
<td>6.3</td>
<td>6.1</td>
<td>3.6</td>
<td>6.0</td>
<td>6.0</td>
<td>6.0</td>
<td>6.0</td>
<td>6.0</td>
<td>6.0</td>
<td>6.0</td>
<td>6.0</td>
<td>6.0</td>
<td>6.0</td>
<td>6.0</td>
<td>6.0</td>
<td>6.0</td>
<td></td>
</tr>
<tr>
<td>L</td>
<td>7.4</td>
<td>2.3</td>
<td>4.4</td>
<td>3.6</td>
<td>25.0</td>
<td>15.7</td>
<td>12.5</td>
<td>10.0</td>
<td>9.0</td>
<td>0.0</td>
<td>8.2</td>
<td>13.8</td>
<td>18.0</td>
<td>47.4</td>
<td>6.9</td>
<td>18.8</td>
<td></td>
</tr>
<tr>
<td>J</td>
<td>7.4</td>
<td>3.3</td>
<td>3.9</td>
<td>3.7</td>
<td>22.5</td>
<td>21.3</td>
<td>12.5</td>
<td>7.6</td>
<td>10.7</td>
<td>0.0</td>
<td>7.8</td>
<td>11.5</td>
<td>12.0</td>
<td>42.0</td>
<td>6.7</td>
<td>12.5</td>
<td></td>
</tr>
<tr>
<td>L,M</td>
<td>1.1</td>
<td>7.0</td>
<td>6.2</td>
<td>1.4</td>
<td>10.0</td>
<td>7.7</td>
<td>5.2</td>
<td>7.6</td>
<td>7.1</td>
<td>1.6</td>
<td>10.0</td>
<td>8.2</td>
<td>9.5</td>
<td>29.1</td>
<td>5.7</td>
<td>8.0</td>
<td></td>
</tr>
<tr>
<td>J,M</td>
<td>1.5</td>
<td>4.3</td>
<td>4.3</td>
<td>1.6</td>
<td>8.9</td>
<td>7.7</td>
<td>3.8</td>
<td>8.5</td>
<td>6.1</td>
<td>1.3</td>
<td>11.0</td>
<td>4.7</td>
<td>6.8</td>
<td>27.7</td>
<td>5.8</td>
<td>5.8</td>
<td></td>
</tr>
</tbody>
</table>

Motorbikes: Shotton et al. 2005 → 7.6%  
Ours: 4.4 % (indep.), 3.9 % (joint)

Bicycle (Rear): Ours: 25.0 % (indep.), 20.8 % (joint)

Cups: Ours: 18.8 % (indep.), 10.0 % (joint)

See paper for details!
SUMMARY

• Shape and geometry for categorization and detection

• Shared over categories (and aspects)

• Required number of weak detectors grows sublinearly with the number of categories

• Alphabet and the detector can be updated incrementally

• Joint learning gives better results with the same amount of training data
THANK YOU!