Spatio-temporal Reasoning for Traffic Scene Understanding

Raluca Brehar*, Carolina Fortuna+, Silviu Bota*, Dunja Mladenic+, Sergiu Nedevschi*

*Technical University of Cluj-Napoca, Romania
+Josef Stefan Institute, Ljubljana, Slovenia
Agenda

• Context and Objective
• Method description
• Results
• Conclusions, future work
Context

• Object detection in traffic scenes

• Ontology for performing general knowledge base and commonsense reasoning engine (OpenCyc)

• Create semantic concepts corresponding to classified objects and map them to the concepts in the ontology.

• Perform reasoning based on those concepts.
Objective

Generate descriptions of the traffic scene in cvasi-natural language.

Perform the semantic understanding of the traffic scene and summarization.
Overview of the proposed method

1. Stereo based object detection system
2. 3D Reconstructed Scene with object hypotheses and attributes
3. Object Recognition
4. Labels / Annotations for the traffic scenes
5. Ontological Modelling of the World
   - Knowledge Base
     - OpenCyc Ontology + rules
     - OpenCyc Inference Engine
   - Improved model of the traffic scene
   - Textual annotation of the image by means of transliterated knowledge
Steps of the method

- Object detection and object attributes generation
- Object recognition
- Ontological modeling of the world
Object Detection and Object Attributes Generation

• Use stereovision for constructing object hypotheses

• Objects are modeled as cuboids, right prisms with rectangular bases. The bases of the prism are always parallel to the xOz plane.
Object Detection and Object Attributes Generation

Each object can be fully described by:

– The \((x; z)\) coordinates of its base corners
– The \(y\) coordinates for the top and bottom bases.
– Speed vector components, along the \(x\) and \(z\) axes.
– Mass center .
– Height
– Width and length
– \(z\) distance from the ego vehicle
Object Recognition

• Machine learning for object recognition:
  – Construct a database of models for each class
  – Extract relevant features for the models in each class
  – Train level 1 classifiers. These learners are trained on histogram of oriented gradients features for the Pole and Pedestrian class. Their recognition score will be used as input feature (among other features) for a meta-classifier.
  – Train meta-classifier on all features and on the responses of level 1 classifiers (AdaBoost and RandomForest).
Object Recognition - Relevant feature extraction

- **Object dimension** - width and height
- **Motion features** are represented by: Speed (on Ox and Oz axis) and motion signature – the horizontal variance of the 2D motion field.
- The **HoG** Pedestrian score and the HoG Pole Score are given by previously trained boosted classifiers on Histogram of Oriented Gradient features for the class pole and pedestrian.
- The **pattern matching** score is given by the degree of match between the contour of an obstacle and templates from a previously generated hierarchy of pedestrian contours.
- The **vertical texture dissimilarity** measures the maximum vertical dissimilarity that can be found in the object’s area.
Ontological Modeling of the World

• Use OpenCyc for modeling the world (acquired images, objects and information about the objects)

• OpenCyc knowledge base is composed of:
  – Ontology
  – Instances of concepts
  – Relations
  – Rules

• For each image, the object recognition module provides a list of objects described by a set of attributes. The attributes are mapped to existing OpenCyc concepts or expressed using corresponding predicates.

• For attributes with no equivalence in the OpenCyc ontology new corresponding concepts and predicates have been defined and linked.
### Ontological Modeling of the World

<table>
<thead>
<tr>
<th>Descriptor + ObjectNo</th>
<th>Value</th>
<th>Constant</th>
</tr>
</thead>
<tbody>
<tr>
<td>ObjectClass</td>
<td>N</td>
<td>Automobile, UtilityPole, Person, Obstacle</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Car5A000287 isa : Automobile, Pole4A000282 isa : UtilityPole, Pedestrian2A000282 isa : Person, Unclassified1000282 isa : Obstacle</td>
</tr>
<tr>
<td>NearLeftX, NearLeftZ, NearRightX, NearRightZ</td>
<td>X, Z Point</td>
<td>Point00A000282 isa : Point</td>
</tr>
<tr>
<td></td>
<td></td>
<td>memberOfList : ListOfVerticesFP0A000282</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(positionAlongXAxisInCoordinateSystem Point00A000282)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(SomeExampleFn CartesianCoordinateSystem)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>((Milli Meter) 3))</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(positionAlongZAxisInCoordinateSystem Point00A000282)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(SomeExampleFn CartesianCoordinateSystem)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>((Milli Meter) 2.4))</td>
</tr>
<tr>
<td>CenterX</td>
<td>X</td>
<td>positionAlongXAxisInCoordinateSystem Pedestrian1A000285</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(SomeExampleFn CartesianCoordinateSystem)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>((Milli Meter) 2))</td>
</tr>
<tr>
<td>CenterY</td>
<td>Y</td>
<td>positionAlongYAxisInCoordinateSystem Pedestrian1A000285</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(SomeExampleFn CartesianCoordinateSystem)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>((Milli Meter) 2))</td>
</tr>
<tr>
<td>CenterZ</td>
<td>Z</td>
<td>positionAlongZAxisInCoordinateSystem Pedestrian1A000285</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(SomeExampleFn CartesianCoordinateSystem)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>((Milli Meter) 2.4))</td>
</tr>
<tr>
<td>Height, Width, Length</td>
<td>H, W, L heightOfObject</td>
<td>heightOfObject : ((Milli Meter) 2)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>widthOfObject : ((Milli Meter) 9)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>lengthOfObject : ((Milli Meter) 10)</td>
</tr>
<tr>
<td>SpeedX, SpeedZ</td>
<td>X, Z speedAlongXAxisInCoordinateSystem Pedestrian1A000285</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(KilometersPerHour -11)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(SomeExampleFn CartesianCoordinateSystem)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(speedAlongYAxisInCoordinateSystem Pedestrian1A000285)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(KilometersPerHour 15)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(SomeExampleFn CartesianCoordinateSystem)</td>
</tr>
</tbody>
</table>

### Descriptors and Constants

<table>
<thead>
<tr>
<th>Descriptor</th>
<th>Constant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Direction of motion</td>
<td>movesInDirection-Average, North-Generally, Northeast-Generally, West-Generally, Northwest-Generally</td>
</tr>
<tr>
<td>Speed of object</td>
<td>speedOfObject-Translation, velocityOfObject</td>
</tr>
<tr>
<td>Relative position between objects</td>
<td>near, farFrom, northOf, northeastOf, eastOf, southeastOf, spatiallyIntersects, overlappingSpaceRegions</td>
</tr>
<tr>
<td>Action execution</td>
<td>performedBy, doneBy</td>
</tr>
<tr>
<td>Events</td>
<td>Walking, Jogging, Running</td>
</tr>
</tbody>
</table>
Examples of conceptual representation of aspects in images

**Representation of a time ordered sequence of images**

Individual: VideoClip0  
isa: Series  
seriesLength: 15  
seriesMembers:  
ImageA000288 ImageA000287 ImageA000286  
ImageA000285 ImageA000284 ImageA000283  
ImageA000282 ImageA000281 ImageA000280  
ImageA000279 ImageA000278 ImageA000277  
ImageA000276 ImageA000275 ImageA000274  
seriesOrderedBy: after

**Representation of an image in a sequence**

Individual: ImageA000282  
isa: Image TemporalThing  
imageDepicts:  
Unclassified10A000282 Unclassified9A000282  
Unclassified8A000282 Unclassified7A000282  
Unclassified6A000282 Unclassified5A000282  
Pole4A000282 Unclassified3A000282  
Pedestrian2A000282 Unclassified1A000282  
Unclassified0A000282 PhySeriesA000282  
(precedesInSeries ImageA000282 ImageA000283 VideoClip0)  
(precedesInSeries ImageA000281 ImageA000282 VideoClip0)  
(seriesMembers VideoClip0 ImageA000282)
Ontological Modeling of the World

• The purpose of the paper is to define rules to improve classification, object recognition and scene understanding.

• Once the relevant concepts have been modeled and introduced in the knowledge base, generic or application specific rules can be defined to achieve the desired functionality:
  – Scene Understanding (i.e. pedestrian on the left side of a pole)
  – Action Recognition (i.e. approaching, moving from left to right)
English transliteration

• The transliteration of a traffic scene can be seen as a textual summary of the scene.

• The proposed method is able to:
  – Infer relationships between objects
  – Understand situations in a scene
  – Output the results in a quasi-natural language manner using transliteration
### Experimental Results: Object Recognition

<table>
<thead>
<tr>
<th>Classifier: Random Forest evaluated using stratified cross-validation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correctly Classified Instances</td>
</tr>
<tr>
<td>Incorrectly Classified Instances</td>
</tr>
<tr>
<td>Kappa statistic</td>
</tr>
<tr>
<td>Mean absolute error</td>
</tr>
<tr>
<td>Root mean squared error</td>
</tr>
<tr>
<td>Relative absolute error</td>
</tr>
<tr>
<td>Root relative squared error</td>
</tr>
<tr>
<td>Total Number of Instances</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Classifier: AdaBoost with J48 decision stumps evaluated using stratified cross-validation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correctly Classified Instances</td>
</tr>
<tr>
<td>Incorrectly Classified Instances</td>
</tr>
<tr>
<td>Kappa statistic</td>
</tr>
<tr>
<td>Mean absolute error</td>
</tr>
<tr>
<td>Root mean squared error</td>
</tr>
<tr>
<td>Relative absolute error</td>
</tr>
<tr>
<td>Root relative squared error</td>
</tr>
<tr>
<td>Total Number of Instances</td>
</tr>
</tbody>
</table>
Experimental Results:

Scene Understanding Rule

There are two consequent images in a video: Img1 and Img2

- Img1 contains Obj1 that is an Obstacle
- Img2 contains Obj2 that is a Person

The two objects have similar dimensions and relative positions to the origin of the coordinate system

Obstacle is a Person!
Experimental Results: Scene Understanding

List of things:
- Image depicts physeriesa000282
- Image depicts unclassified0a000282
- Image depicts unclassified1a000282
- Image depicts pedestrian2a000282
- Image depicts unclassified3a000282
- Image depicts pole4a000282
- Image depicts unclassified5a000282
- Image depicts unclassified6a000282
- Image depicts unclassified7a000282
- Image depicts unclassified8a000282
- Image depicts unclassified9a000282
- Image depicts unclassified10a000282

Classification of object instances to semantic concepts:
- pedestrian2a000282 is a person.
- pole4a000282 is a utility pole.
- unclassified6a000282 is a utility pole.
- unclassified7a000282 is a utility pole.
- unclassified8a000282 is a utility pole.
- unclassified9a000282 is a utility pole.
- unclassified10a000282 is a utility pole.
- unclassified3a000282 is a car.

Inference of belonging to related semantic concepts:
- unclassified3a000282 is a roadvehicle.
- unclassified3a000282 is a mechanicaldevice.
- unclassified3a000282 is a powereddevice.

Relative positioning:
- unclassified0a000282 near unclassified1a000282.
- unclassified0a000282 northOf unclassified1a000282.
- unclassified0a000282 eastOf pedestrian2a000282.
- unclassified1a000282 northeastOf pedestrian2a000282.
- unclassified8a000282 farFrom pole4a000282.
- unclassified8a000282 near unclassified9a000282.
- unclassified0a000282 northOf unclassified9a000282.
- unclassified1a000282 northOf unclassified9a000282.
- pedestrian2a000282 farFrom unclassified9a000282.
Experimental Results: Action Recognition

Motion modeling by translation and direction:
- pedestrian2a000282 moves in direction northwest—generally, speed of object kilometers per hour 17.205.
- unclassified3a000282 moves in direction northwest—generally, speed of object kilometers per hour 17.205.

Action recognition:
- actionpedestrian2a000282 performed by pedestrian2a000282
- actionpedestrian2a000282 isa running.

- actionpedestrian2a000282 performed by pedestrian2a000282
- actionpedestrian2a000282 isa ambulation.

- actionpedestrian2a000282 performed by pedestrian2a000282
- actionpedestrian2a000282 isa movingalllegs.

- movement02a000282 done by pedestrian2a000282
- movement02a000282 isa approaching.

- movement02a000282 done by pedestrian2a000282
- movement02a000282 isa translation-locationchange.

- movement02a000282 done by pedestrian2a000282
- movement02a000282 isa movement-translationevent.
Experimental Results:
English transliteration

Image ImageA000282 depicts SpatialThing?
  unclassified0a000282 is a utility pole,
  every utility pole is an open-air,
  every open-air is a localized spatial
  thing,
  every localized spatial thing is a spatial
  thing.
unclassified1a000282 is an obstacle,
  every obstacle is a tangible thing,
  every tangible thing is three dimensional
  thing,
all three dimensional thing is a thing
  with two or more
  dimensions,
every thing with two or more dimensions
  is a spatial thing with one or more
  dimensions,
every spatial thing with one or more
  dimensions is a spatial thing.

Image ImageA000282 depicts ObjectWithUse?
  unclassified0a000282 is a utility pole,
  every utility pole is a post,
  every post is a shaft,
  every shaft is a rod,
  every rod is an implement,
  every implement is a device,
  every device is an object with uses.
unclassified3a000282 is a car,
  every car is a device that is
  not a weapon,
every device that is not a weapon
  is a device,
every device is an object with uses.
Conclusions and Future work
Thank you!

Questions?