ROBIN Challenge
Competitions

Emmanuel D'Angelo[1] Stéphane Herbin[2] Matthieu Ratiéville [1]

[1] DGA/CEP/GIP, 16 bis av. Prieur de la Côte d'Or, F-94114 Arcueil cedex
[2] ONERA, BP72 - 29 avenue de la Division Leclerc, F-92322 CHATILLON cedex

http://robin.inrialpes.fr

November 24, 2006- DRAFT

1 Introduction

This document is produced by the ROBIN Project Committee. Its objective is to describe the six data sets used in the ROBIN challenge provided by the following industrial companies or agencies:

- CNES for SPOT5 spatial images;
- MBDA, EADS and Thales for aerial infrared and visible images;
- Bertin Technologies, ECA and Sagem for infrared ground images.

Each data set is divided into several competitions committed to the evaluation of various types or levels of difficulties. Participants to a competition will receive:

- initially a number of images dedicated to the *Training* of the algorithms,
- later on some images for the *Validation* of the evaluation process (blank test),
- eventually the images for the *Test*.

A geometrical acceptance criterion is needed for the Detection tasks evaluated on 5 of the data sets (Bertin/ECA, EADS, MBDA, Sagem and Thales). Its definition is presented in details in the document describing the performance evaluation metrics ([1]).

2 CNES data set

2.1 Data

The data set contains patches of size 128×128, and complete scenes of size 24000×24000.

Patches contain centered objects and background. Patches containing parts of objects are considered background.

2.2 Ground truth annotation

Patches are assigned a single category.

Scenes are assigned a list of categories and centers. Object size or extension is not annotated.

A global scene can not be wholly annotated with all potential categories. Only a limited number of objects are identified.

2.3 List of categories

1. Background
2. Roundabouts
3. Crossroads
4. Highways and trunk roads
5. Secondary roads
6. Minor roads
7. Tracks and ways
8. Insulated building
9. Suburban Area, house gathering
10. Bridges (PT: Ponts)
11. Railways (VF : Voies Ferrées)

2.4 Protocole

The evaluation is divided into two groups of competitions:

Detection task

The goal is to detect the presence of an object on a patch. The detection is evaluated as a binary decision between two groups of categories, including background.

- INPUT : list of target object categories, test image;
- OUTPUT : binary decision with optional confidence coefficient.

The detection task will be evaluated on three sub-challenges:

Detection sub-challenge	*Positive categories*	*Negative categories*	*Train*	*Validation*	*Test*
D1 : Linear structures	Tracks, Minor roads, Highways, Secondary roads, Railways	Insulated building, Crossroads, Bridges, Roundabouts, Suburban area, Background	1286+1441	1286+1440	2573+2882
D2 : Compact structures	Insulated building, Crossroads, Bridges, Roundabouts	Tracks, Minor roads, Highways, Secondary roads, Railways, Suburban area, Background	961+1766	961+1765	1923+3532
D3 : Structures	Tracks, Minor roads, Highways, Secondary roads, Railways, Insulated building, Crossroads, Bridges, Roundabouts, Suburban area	Background	2502+225	2501+225	5005+450

Categorization task

The goal is to assign a category to a patch containing a centered object.

- INPUT : list of target object categories, test image;
- OUTPUT : object category with optional confidence coefficient.

The Categorization task will be evaluated on three sub-challenges:

Categorization sub-challenge	*Target categories*	*Train*	*Validation*	*Test*
C1 : Linear structures	Tracks, Minor roads, Highways, Secondary roads, Railways	1286	1286	2573
C2 : Compact structures	Insulated building, Crossroads, Bridges, Roundabouts	961	961	1923
C3 : Structures	Tracks, Minor roads, Highways, Secondary roads, Railways, Insulated building, Crossroads, Bridges, Roundabouts, Suburban area	2502	2501	5005

2.5 Acceptance criterion

The definition of a geometrical acceptance criterion is only necessary for the object localization competition. Good localization is defined according to the type of output:

Point : a location defined by a position $\boldsymbol{x}$ compared to a ground truth bounding box $[\boldsymbol{x}^*_{min}, \boldsymbol{x}^*_{max}]$ is correctly assigned if

$$\|\boldsymbol{x} - \frac{1}{2}(\boldsymbol{x}^*_{min} + \boldsymbol{x}^*_{max})\| \leq \delta$$

Bounding box : a location defined by a bounding box $[\boldsymbol{x}_{min}, \boldsymbol{x}_{max}]$ compared to a ground truth bounding box $[\boldsymbol{x}^*_{min}, \boldsymbol{x}^*_{max}]$ is correctly assigned if

$$\|(\boldsymbol{x}_{min} + \boldsymbol{x}_{max}) - (\boldsymbol{x}^*_{min} + \boldsymbol{x}^*_{max})\| \leq 2\delta$$

The location tolerance value δ is typically 3 pixels.

3 THALES

3.1 Data

The data set contains scenes observed from two airborne infrared sensors, sampled along a sequence of views under various viewing conditions (altitude, date ...) and environments (urban, semi-urban, rural).

Sensors used are:

- Micro Bolometer, generating images of size 320×240;
- InSb sensor, generating images of size 320×256;
- QWIP sensor, generating images of size 640×512.

Objects to be detected are simple in order to be observable with all types of sensor and viewing conditions.

3.2 Ground truth annotation

All images will be annotated by rectangular bounding boxes along image axes, and object category.

It is assumed that all objects of interest in a scene have a corresponding bounding box.

3.3 List of categories

Scenes contain various objects and buildings. Objects of interest are:

1. Car (three sub-categories)
2. Bus
3. Truck
4. Boat

The three car sub-categories are designed to assess the ability of classification algorithms to discriminate between various behaviors, (parked, moving and halted vehicles) using infrared phenomena.

3.4 Protocole

Evaluation is divided into two competitions, collecting various viewing conditions.

Detection Task

The goal is to detect objects in a global scene by providing a list of their localization.

- INPUT : list of target object categories, test image;

- OUTPUT : list of detections (bounding boxes) with optional confidence coefficients.

The detection task will be evaluated on three sub-challenges. The following table shows the number of images and objects (in parenthesis) for each sub-challenge.[1]

Detection sub-challenge	*Target categories*	*Train*	*Validation*	*Test*
D1 : Boats	Boat	16 (180)	9 (76)	*30 (300)*
D2 : Cars	Car	292 (1826)	155 (390)	*500 (2000)*
D3 : Vehicles	Car, Truck, Bus	309 (1957)	162 (427)	500 (2500)

Categorization Task

The goal is to assign a category to a designated object. Object location is a point in image coordinates. No bounding box is provided.

- INPUT : list of object categories, test image with list of object positions;
- OUTPUT : object category with optional confidence coefficient for each location.

The detection task will be evaluated on two sub-challenges. The following table shows the number of objects to categorize.[2]

Categorization sub-challenge	*Target categories*	*Train*	*Validation*	*Test*
C1 : Vehicles	Car, Truck, Bus	1957	512	*3000*
C2 : Cars	Moving car, Parked car, Halted car	1826	475	*2500*

Rejection Task

No rejection ability is evaluated with this data set.

4 EADS

4.1 Data

The data set contains patches, and complete scenes. It is built from 9 high resolution images on which object appearances have been inserted (25 to 30 objects per scene).

The scenes have various resolution, noise level, viewing and illumination conditions reflecting typical remote sensing conditions.

The patches contain centered objects with shadow. Other objects may be included in thoses patches, but only the centered one is considered for the object categorization competition.

[1] The exact database size for the final test will be confirmed later.

[2] The exact database size for the final test will be confirmed later.

4.2 Ground truth annotation

Scenes are assigned a list of possible categories and viewing conditions (resolution, viewing angle, illumination angle, date, noise level).

Patches are assigned a single ground truth category and a reference to the original scene (bounding box in scene coordinates).

It is assumed that all objects of interest in a scene have a corresponding patch.

4.3 List of categories

1. Trucks (5 subcategories)
2. Buses
3. Cars
4. Airplanes (7 subcategories)

4.4 Protocole

Evaluation is divided in several scenarii based on nine different scenes, sampling several types of difficulties. On each scenario, object detection and categorization will be tested.

Evaluation is divided into two competitions:

Detection Task

The goal is to detect objects in a global scene by providing a list of their localization.

- INPUT : list of object categories, test image;
- OUTPUT : list of detections (bounding boxes) with optional confidence coefficients.

The detection task will be evaluated on six sub-challenges:

Detection sub-challenge	*Target categories*	*Train*	*Validation*	*Test*
D1V : Small vehicles	Car	140	23	47
D2V : Big vehicles	Long truck, Long truck with cabin, Middle truck, Semitrailer, Bus	559	23	47
D3V : All vehicles	Car, Van, Long truck, Long truck with cabin, Middle truck, Semi-trailer, Bus	793	23	47
D1P : Small planes	Business jet, Tourism plane, Small size transport plane	257	18	35
D2P : Big planes	Medium size transport plane, Medium size transport plane with propeller, Medium size transport plane with booster, Big size transport plane.	343	18	35
D3P : All planes	Business jet, Tourism plane, Small size transport plane, Medium size transport plane, Medium size transport plane with propeller, Medium size transport plane with booster, Big size transport plane.	583	18	35

Categorization Task

The goal is to assign a category to a patch containing a centered object.

- INPUT : list of object categories, test image with centered object;
- OUTPUT : object category with optional confidence coefficient.

The detection task will be evaluated on four sub-challenges:

Categorization sub-challenge	*Target categories*	*Train*	*Validation*	*Test*
C1V : Small vehicles	Car, Van	1143	1143	2285
C2V : All vehicles	Car, Van, Long truck, Long truck with cabin, Middle truck, Semi-trailer, Bus	1588	1585	3170
C1P : Small planes	Business jet, Tourism plane, Small size transport plane	283	281	563
C2P : All planes	Business jet, Tourism plane, Small size transport plane, Medium size transport plane, Medium size transport plane with propeller, Medium size transport plane with booster, Big size transport plane.	544	540	1080

Rejection Task

The goal is to state that an object is in a given list of categories or not. Test data sets will contain object categories that are not in the training set.

- INPUT : list of object categories, test image with centered object;
- OUTPUT : binary rejection flag.

The rejection task will be evaluated on two sub-challenges:

Rejection sub-challenge	*Target categories*	*Train*	*Validation*	*Test*
R1V : Trucks	Long truck, Long truck with cabin, Middle truck, Semitrailer	371	442	885
R1P : Big planes	Medium size transport plane, Medium size transport plane with booster, Big size transport plane.	190	259	517

5 MBDA

5.1 Data

Mainly synthetic images are provided,with an additional small number of real images. The size and acquisition parameters of the images are consistent with aerial or ground sensors. Three spectral bands are equally represented:

- IR 3-5 μ
- IR 8-12 μ
- panchromatic

The real images have been acquired by the three following sensors, depending on the spectral band:

- an ORION camera of size 256^2 for IR 3-5 μ
- a microbolometric camera of size 320×240 for IR 8-12 μ
- a camera of size 458×341 for the panchromatic band

The synthetic images have been produced using a physical model of a real matricial sensor of size 512^2. Various environmental factors are rendered: visibility and contrast linked to the season, the time of the day, and the weather; perturbatory elements such as shadows, smoke, masking, clutter. Some variability is also introduced in the objects: engine on/off, open/closed door... However the sky, if it appears in an image, is not rendered realistically.

The elevation angle of the sensor ranges from 0 to 70 degrees, the distance to the main object of interest from 50 to 2000 m, which means the sizes of the targets range from 10 to 120 pixels roughly.

Several sceneries are modelized:

- countryside
- half-urban
- airport
- seaside with industrial activity

For the purpose all learning, only synthetic images will be available: their size is 128x128 and each one features a single object of interest, in central position on a uniform background.

5.2 Categories

The objects of interest are civilian ground vehicles and landed aircrafts, plus a couple of infrastructure elements:

- cars (4 models : 3 for training and validation + 1 for evaluation only)
- trucks (3 models: 2 for training and validation + 1 for evaluation only)
- airplanes (3 models: 2 for training and validation + 1 for evaluation only)
- helicopters (3 models: 2 for training and validation + 1 for evaluation only)
- Telecom towers (2 models: 1 for training and validation + 1 for evaluation only)

5.3 Ground Truth annotation

For each image and each object of interest present in the image, the following information will be available:

- the category of the object
- a bounding box excluding the possibly masked parts of the object (the axes of the bounding box parallel to the image borders)

5.4 Protocole

For each band (IR2,IR3,VIS) the following three competitions are proposed, which makes a total of nine competitions:

Vehicle Detection Task

The goal is to detect the vehicles (i.e. any object belonging to the categories listed in section 5.2 except Telecom towers) in an image by providing a list of their localizations.

- INPUT : test images
- OUTPUT : list of detections (a detection is preferentially a bounding box, otherwise the coordinates of a point) for each image, with optional confidence coefficient.

The number of images used for this task is as follows:

Training images	*Validation images*	*Test images*
10368	124	1250

Telecom Tower Detection Task

The goal is to detect the Telecom towers in an image by providing a list of their localizations.

- INPUT : test images
- OUTPUT : list of detections (a detection is preferentially a bounding box, otherwise the coordinates of a point) for each image, with optional confidence coefficient.

The number of images used for this task is as follows:

Training images	*Validation images*	*Test images*
1152	33	273

Vehicle Categorization Task

The goal is to assign a category to the vehicles indicated in the image.

- INPUT : test image with a list of bounding boxes.
- OUTPUT : category of the object for each bounding box, among: *Car*, *Truck*, *Helicopter*, *Plane*, *Ambiguous*; or a confidence coefficient for each of these categories except *Ambiguous*.

All the provided bounding boxes do contain one of the four categories of vehicles: no rejection ability will be evaluated with the MBDA data set.

The number of images used for this task is as follows:

Training images	*Validation images*	*Test images*
10368	107	1115

Note that the Training images are the same as for the Vehicle Detection Task.

6 SAGEM

6.1 Data

The provided images are infrared real images in the range 3-5 μ, acquired by the MATIS matricial camera of size 384×256 and extracted from short (about 10 seconds) video sequences at 50 frames per second. The images mainly reflect ground acquisition conditions, except for most of the training set which has been acquired from a helicopter.

6.2 Categories

There are three kinds of objects of interest: vehicles, people and environmental elements.

As far as the vehicles are concerned, the following categories are present:

- road cars
  - Peugeot 307
  - Volkswagen Golf
  - ND
- city cars
  - Peugeot 106, 205, 206, 306
  - Citroën Twingo, AX
  - Renault Clio
  - ND
- saloon
  - Renault Megane, Laguna
  - Ford Escort
  - Citroën Xsara break
- monospace
  - Renault Scenic, Espace
  - ND
- MPVs
  - Citroën C25
  - ND
- 4 Wheel Drive (Land Rover Range Rover)
- utility vehicles
  - Renault Express
  - Citroën
  - Peugeot Expert

The people may be standing up, sitting in a car, riding a bicycle.

Many environmental elements have also been annotated: street lights, traffic lights, pylons, bus stops...

6.3 Ground Truth annotation

For each image and each object of interest present in the image, the following information will be available:

- the category of the object
- a bounding box excluding the masked parts of the objects

6.4 Protocole

A detection and a categorization competitions are proposed. They are defined by the types of expected Input and Output of the evaluated algorithms.

Detection Task

The aim is to localize objects of interest in an image. This task is divided into three sub-challenges:

- 4 wheel vehicle detection
- people detection
- street light detection.

The training, validation and test images are the same for the three challenges, but participants can choose to answer to one or several challenges, and the challenges will be evaluated independently.

- INPUT : test images
- OUTPUT : list of detections (a detection is preferentially a bounding box, otherwise the coordinates of a point) for each image, with optional confidence coefficient.

The number of images used for this task is as follows:

Training images	*Validation images*	*Test images*
317	41	367

Car Categorization Task

The aim is to recognize the model of the car whose bounding box is provided.

- INPUT : test image with a list of bounding boxes,
- OUTPUT : model of the vehicle for each bounding box, among: *Xsara break*, *Peugeot 106*, *Renault Express*, *Ambiguous*; or a confidence coefficient for each of these categories except *Ambiguous*.

This is a discrimination test: the given bounding boxes do contain one of the above three car models. No assessment of rejection ability will be performed with SAGEM data set.

The number of images used for this task is as follows:

Training images	*Validation images*	*Test images*
317	42	375

7 Bertin - Cybernetix

7.1 Data

The proposed datasets are made of colour and infrared images of both vehicles and pedestrians (see the exact list below). The imaging devices used for these datasets are:

- a color Sunkwang Electronics SK-2172X camera with a Sony CCD;
- a FLIR A20M thermal imager delivering pictures with a resolution of 320×240 pixels coded over 16 bits.

The field of view of the visible camera has been adapted in order to fit the one of the IR camera.

7.2 Categories

The objects of interest will be of two kinds: vehicles and persons. As far as the vehicles are concerned, three categories are present:

- cars (sedans), with the following models in the complete database :
  - Renault Laguna
  - Renault Twingo
  - Renault Scenic
  - Peugeot 407 SW
  - Peugoet 309
  - Ford Ka
  - Ford Focus
- utility vehicles :
  - Citroën Berlingo
  - Citroën Jumper
- motorbikes :
  - Peugeot Satelis (scooter)
  - Yamaha TDR

Six different persons are also present, and will be identified by a first name (Clement, Serge, Bachir, Murielle, Tony and Laurent).

The images were shot in 8 different places in order to provide different backgrounds, which are mostly rural and semi-rural backrounds.

Some sequences have been acquired with moving objects of interest, although movement detection task will not be considered in the ROBIN framework. These sequences will however differ from the static shots, because some characteristics of the sensors (for example, integration time or shutter speed) will create some blurring over the objects, and occlusions are more likely to happen. These images are hence expected to be more difficult to deal with for the algorithms.

7.3 Ground Truth annotation

The images will be provided with the following information:

- the category of the object
- a bounding box including the masked parts of the objects

All the objects contained in the images, and not only the ones specially used as objects of interest, will be annotated in the ground truth, but with less accuracy in their description: for example, a car will be described by the class ”sedan” and not by the exact model.

The object categories included in the ground truth will be the precise type (for a car) or name (for a person). However, in the challenges, the lists of classes to detect will not be so accurate : for instance, one will have to detect cars or utility vehicles. The evaluation dataset will indeed include new models of vehicles and new people, in order to assess the generality of the the diferent systems.

For a small subset of images, a detailed pixelic mask of the present objects of interest may be provided.

Remarks : background objects and occlusions The image acquisitions have been made using a small number of vehicles (see section 7.2), called ”Test Vehicles”. These vehicles will be used for training and testing the algorithms on known data, and will be most frequently present in the foreground of the images since they are the main element of interest in these images.

However, the different sequences were shot in public or open spaces. Hence, some vehicles may be present in the background, *e.g*. vehicles parked along a road. All these background objects cannot be assigned to a precise category as defined in section 7.2, but need to be present in the ground-truth. Otherwise, an algorithm which would be able to detect all the vehicles, including the ones in the background, would be penalized. In consequence, the following rule has been applied for the annotation of these background objects :

- if more than 25% of the vehicle is visible, then the bounding box surrounding it (and not two or more boxes for each visible part) is provided in the ground truth
- if less than 25% of the vehicle is hidden, then it is not added to the ground truth

- the label assigned to background objects is their general class (*e.g*. sedan, utility vehicle), and will not be the exact type of the vehicle
- in the case of several overlapping background objects, only one bounding box included the complete group of vehicles is provided with a specific comment.

Furthermore, if a test vehicle is occluded by some element of the image (which is the case in some sequences), the occluded part of the vehicle is also annotated and a comment with the evaluation of the percentage of occlusion is added in the metadata file associated to the image.

7.4 Protocole

Three competitions are proposed : one detection task and two classification tasks. They are defined by the types of expected Input and Output of the evaluated algorithms.

Detection Task

The Detection Task aims at finding the instances of the general classes defined in section 7.2, *i.e*. car, utility vehicle, motorbike, person. All the models used during the acquisition of the images will not be included in the training database. Hence, there will be some unknown vehicles and persons in order to assess the generality and the flexibility of an algorithm and of its embedded data representation.

- INPUT : list of classes to research = **Car, Utility Vehicle, Motorbike, Person**
- INPUT : test images
- OUTPUT : list of detections (bounding boxes and assigned class) for each image.

Classification Task 1

In this first classification task, the classes to research correspond to vehicles and people. The goal is to assign the correct label to a patch which may contain an element of a class or some background.

- INPUT : list of classes to research = **Car, Utility Vehicle, Motorbike, Person, Background, Other**
- INPUT : test images with list of bounding boxes
- OUTPUT : class of detected object for each bounding box

Classification Task 2

In this second classification task, the classs to research do not correspond to vehicles or man-made objects. The test patches will contain a person, and the goal is here to perform a posture recognition task. Hence, the classes to research will be in this case:

- standing
- crouching

- lying

Except for the definition of the different classes, the task remains the same as above:

- INPUT : list of postures to research = **Standing, Crouching, Lying**
- INPUT : test images with list of bounding boxes
- OUTPUT : posture of the person inside each bounding box

References

[1] “ROBIN Challenge: Evaluation principles and metrics”, Emmanuel D’Angelo, Stéphane Herbin and Matthieu Ratiéville