

## ESTIMATING VEHICLE ACTIVITY USING THERMAL IMAGE SEQUENCES AND MAPS

U. Stilla, E. Michaelsen

FGAN-FOM Research Institute for Optronics and Pattern Recognition, Gutleuthausstr. 1, 76275 Ettlingen, Germany  
stilla@fom.fgan.de, mich@fom.fgan.de

**KEY WORDS:** Thermal images, Large scale vector maps, Spot detection, , Grouping

### ABSTRACT:

Monitoring traffic in dense build-up areas is a difficult task. In this contribution we describe a method to assess the activity of vehicles based on airborne image sequences from an infrared camera. At the resolution of approximately one meter vehicles appear as elongated spots. In urban areas many additional other objects have the same property. To discriminate vehicles from these objects we fuse information from IR-images and vector-maps. Besides evidence from overlapping areas of frames an IR-sequence is accumulated in the scene. We also perform grouping of vehicles into rows along the margins of the roads. These tasks can be automatically performed by a production net. The generic productions for grouping become feasible through the use of map knowledge as context.

### 1. INTRODUCTION

Monitoring traffic in dense urban areas is a difficult task. Among the many objects participating in traffic, like pedestrians, bicyclists, trams, subways, busses, trucks, etc. cars are a special problem in cities. Monitoring moving cars is being studied for a long time, e.g. [Dreschler & Nagel, 1982],[Haag & nagel, 199],[Partinevelos et al., 2000]. But the activity of cars cannot only be described by their movement. We distinguish moving cars from stationary ones. These may still be active, i.e. waiting at a traffic light or in a traffic jam. We also like to distinguish cars that halt only for a short time - e.g. for loading, in a drive through service or just quickly picking things - from those who are parked for hours or days. Thus the temperature of a car is an important feature for its activity. Temperature can be captured by IR-sensors.

There is mutual evidence between cars and certain GIS-segments. Cars will be present only on roads and parking lots. While some authors suggest to use car detectors as hints for parking lots [Quint, 1996] or roads [Hinz & Baumgartner, 2001], we use the map as support for searching cars. For oblique views and building recognition a 3-d GIS database is used [Stilla et al., 2000], but in dense urban areas the cars will be visible only in nearly vertical views, and roads and parking lots are flat. While in [Stilla, 1995],[Stilla et al., 2000] we focussed on the geometric aspects, we here emphasize the importance of the radiometric information. So conventional 2-d large-scale vector maps are sufficient.

Stationary passive cars appear as dark stationary spots, because they are colder than the surrounding. They will usually occur along the margin of roads or in parking lots, and they will be grouped into rows. Fig. 1 shows such a situation. Active vehicles will appear as moving bright spots on the roads. A stationary bright spot within a row of stationary dark spots can be interpreted as a car that is still warm (has been moved short time before) or as warm spot on the bare concrete giving the hint, that there has been a vehicle short time before, that moved away. Rows of bright spots in the roads are probably cars, that are waiting at a traffic light or in a traffic jam. Thus the

percentage of bright versus dark spots gives a good estimation of the activity of cars in the scene.

Fig.1 shows a typical example of a thermal image of an urban site with rather low car activity. While some buildings and cars have good contrast to the background, roads do not necessarily have clear margins in this kind of imagery

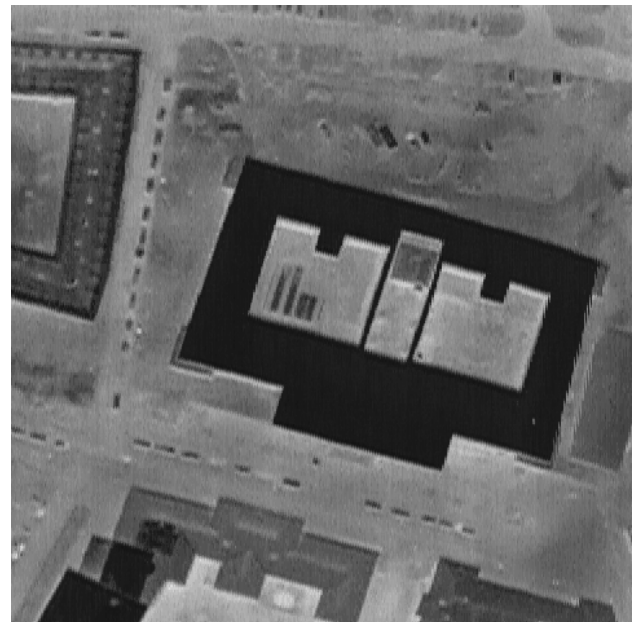


Fig. 1. Example of a thermal IR image of an urban area. Note the parking cars appearing as dark spots grouped in approximately equally spaced rows along the margin of the roads.

Several authors have proposed to use perceptual grouping in urban areas with the contextual aid of maps. E.g. in [Quint, 1996] parking lots known from the map are used to analyze

monocular aerial images and to group cars into rows. The knowledge is represented in a semantic net, and the instantiation of corresponding structures in the data is done by A\* search. Here we use a production net to represent the knowledge and a pruned bottom-up search.

## 2. IMAGE PROCESSING

In this section we present the proposed method and show example results: In Sec. A the Operator is explained, that we use for spot cue detection. Sec. B presents the cluster formation, that assembles pixels into spots with considerable mass and second moments. The fusion with map information, that filters out most of the clutter spots is described in Sec. C. We then focus on the grouping process in Sec. D. In Sec. E we recommend to also fuse the information from different images in the overlapping regions. While this section focuses on the application, the following Sec. III gives a brief methodological overview.

### 2.1 Iconic spot detection

For detection of cues for cars a spot detector [Kohnle et al., 1993] is used. For each pixel  $p=(x,y)$  it looks for the maximum gray value  $m_d$  along the margin of a square of radius  $d$  (Pixels  $(x+j,y+i)$ , where  $\text{Max}(|i|, |j|)=d$ ). See Fig. 2a. The radius  $d$  varies from 1 to a maximum, which is a parameter of the operator. Of all maxima  $m_d$  the minimal value is compared to the gray value in the center. If the central value is significantly higher, then there is evidence for a hot spot at this position, and the operator sets this difference as value at position  $(x,y)$  of its output. Otherwise it sets the output to zero. Dark spots are found in an analogue way.

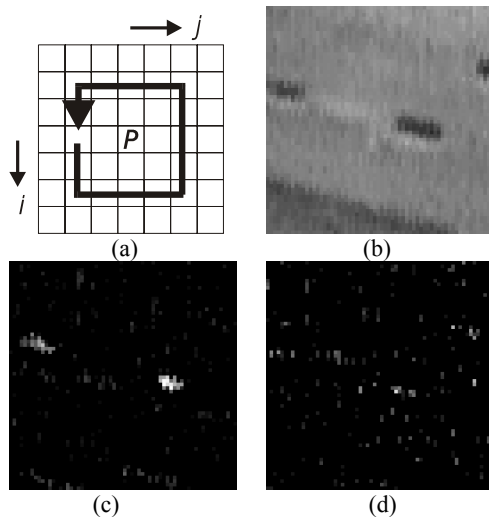


Fig. 2. Spot detector and its application to an IR-image: a) Pixel locations considered in the inequalities of the operator; b) section of an IR-Image, c) result of the cold spot detector; d) result of the hot spot detector.

Fig.2 shows the action of this operator on a section of the image displayed in Fig. 1b shows the section from the center, Fig. 2c shows the cold spot image and Fig. 2d the hot spot image. For clarity Fig. 2c and Fig. 2d have been adjusted in brightness.

### 2.2 Spot Formation by Clustering

Non-zero pixels are entries to the symbolic reasoning system presented in this contribution. The first step in this system is the formation of spots from sets of adjacent pixels of consistent temperature (hot or cold). Apart from this temperature a spot gains a location, a mass and second moments as attributes.

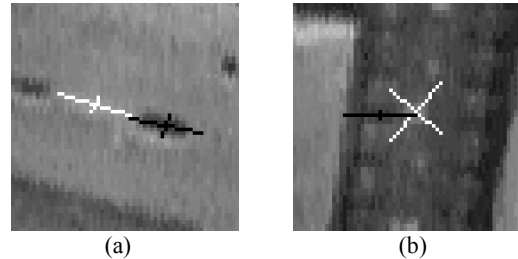


Fig. 3. Examples of spots found by the cluster formation: a) A strong cold spot (a car) on the right, and a weak warm spot on the concrete on the left; b) a cold and a warm spot on a roof.

Fig. 3 shows examples of such spots. Fig. 3a displays one example of a cold spot (black) caused by a vehicle and one example of a hot spot (white) caused by still warm road surface in a place, where a car has moved away short time ago. The crosses indicate the position, orientation and eccentricity of the spots. The mass of the cold spot is one order of magnitude higher than the mass of the hot spot. While these are examples for the kind of spots we are interested in, Fig. 3b shows examples of spots with similar or more prominent attribute values, that do not correspond to vehicles or phenomena caused by vehicles. These are due to structures on a roof in the upper left region of the example image. The cold one is very eccentric but has only little mass (comparable to the hot one in Fig. 3a, the hot one has high mass, but its eccentricity is so low, that the orientation has barely any meaning. Among all spots in the image, the ones corresponding to vehicles like in Fig. 3a are a minority, while clutter like in Fig. 3b forms the majority. This motivates the use of map knowledge to concentrate the search on the road margins and parking lots.

### 2.3 Additional Map Information

Large scale vector maps of roads provide a proper means to exclude most of the non-vehicle spots. Fig. 4 shows such a map overlaid as white lines on the image. The regions of interest inferred from the map are displayed in dark gray-tone, while the non-interest regions are displayed much brighter.

For the registration of map and image attention must be paid to the interior distortions of the camera (in this case a quadratic distortion with one parameter was used) and to the exterior distortions of the view (here we assumed the scene to be flat and thus used a projective homography with eight parameters).

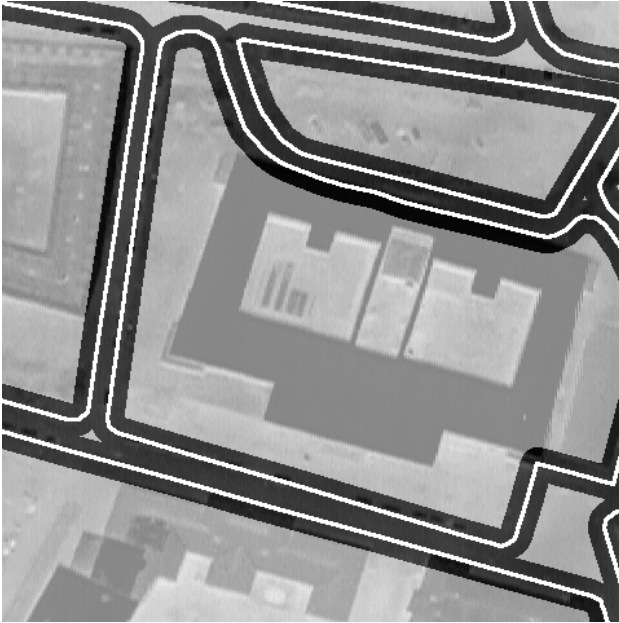


Fig. 4. Interest regions imposed by a map on the IR-image. The road margins of the map are displayed as white lines. The resulting interest regions are shaded much darker than the regions that are excluded from the search for cars.

#### 2.4 Grouping Spot-Rows along the Road Margins

Cars tend to be placed in equidistantly spaced rows along the margin of roads. This criterion allows to discriminate them from other spot-shaped objects. Grouping of such spots into rows of arbitrary length is a generic operation. Fig. 5a shows a section of an IR image containing a row of cars.

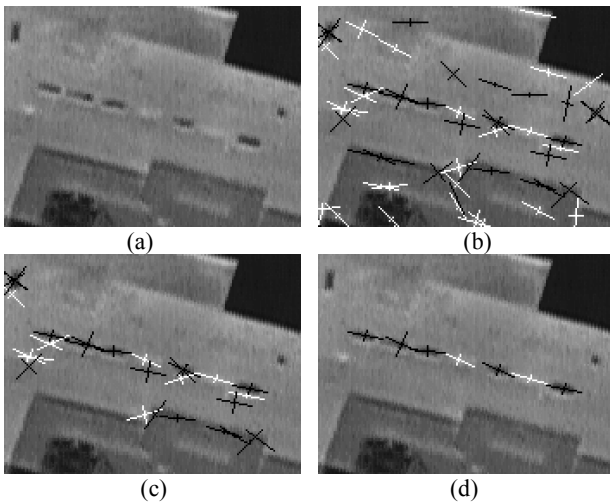


Fig. 5. The benefit of grouping: a) A section of an IR-image; b) all spots constructed in that region; c) spots in the interest region given by fusion with the map; d) car-spots remaining consistent with the row model after grouping.

All detected warm and cold spots in this section are displayed in Fig. 5b. Spots caused by cars constitute only a subset of these. Fig. 5c shows those spots, that are sufficiently close to a road

margin. The grouping starts only from spots, which exceed a minimal mass. The grouping direction is constrained by the road margin. All spots are taken into account, that exceed another lower threshold. Only those spots fitting into the straight and equidistantly spaced row model are grouped.

Still there may be several alternatives of grouping, e.g. if two spots are close to another in a location consistent with the model (see Fig. 5c, most right member of the row). Among the alternatives one group is selected based on an assessment, which is calculated from the number of spots, total mass, regularity in spacing and straightness and consistency in orientation of the spots. Fig. 5d shows the best group containing seven spots.

#### 2.5 Processing image sequences

Compared to standard aerial imagery the image size of the IR images used here is small (standard video). An appropriate resolution on the ground implies that only a small section of the scene can be captured by one image (frame). To monitor a broader scene a sequence of images has to be recorded. The analysis should not be performed on the images independently.

Rather we suggest to project image spots of sufficient mass into a fixed Cartesian coordinate system in the scene and perform a clustering of the spots in the scene. Utilizing map information and grouping is then performed in the scene instead of the image. Large structures, e.g. very long rows of vehicles, that don't fit into a single image can be analyzed.

Parameters and tolerances can be defined in real scene measures (e.g. meter), which is more appropriate than in the image (e.g. pixels). For cameras with non-projective distortions straight lines in the scene are not necessarily straight in the image. This poses additional problems to non-local grouping techniques like the row formation, if the pixel coordinate system is used. Image sequences allow a multiple measurement of the attributes of a spot. Averaging mass, orientation and eccentricity reduces the noise. Image sequences in time allow to determine velocity as additional attribute of a spot.

### 3. THE PRODUCTION NET

The automatic extraction of vehicle rows described in the previous section is implemented in a blackboard based production system [Stilla, 1995],[Stilla & Michaelsen, 1997]. A production captures a structural relation of the object model. The production defines how a given configuration of objects is transformed into a single more complex object. In the condition part of a production, geometrical topological and other relations of objects are examined. If the condition part of production holds, an object specific generation function is executed to generate a new object.

The hierarchical organization of object concepts and productions can be depicted by a production net which displays the part-of hierarchies of object concepts. Concepts represent object types and define a frame for concrete objects (instances) which are described by their attribute values .

The production net for this contribution is shown in Fig. 6. Production p1 generates objects  $i\_SPOT$  from sets of objects  $i\_PIXEL$  which are close together and consistent in temperature. Production p2 generates objects  $s\_SPOT$  in an analogue way from sets of objects  $i\_SPOT$ . While p1 captures a pure part-of relation staying in one domain (the image), production p2 has concrete-of semantic and changes domains from image to scene.

Production p3 initiates the grouping process. Here the link from the object m\_LINE of the map domain has a context semantic, while the link to the object i\_SPOT is a part-of link. Production p5 performs the generic perceptual grouping of objects i\_SPOT into objects i\_ROW and uses objects m\_LINE as context. Productions p4 and p6 are analogue to productions p3 and p5 working in the scene domain.

The recursive productions p5 and p6 need to use the object concept m\_LINE as context. This prevents them from grouping all possible rows of objects i\_SPOT or s\_SPOT. Otherwise there would be very many possibilities, making the process infeasible.

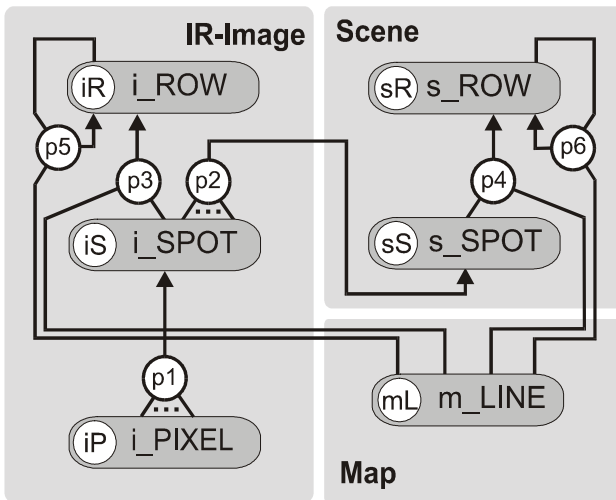


Fig. 6. Production net for automatic vehicle row extraction for activity estimation. Such production nets give a good overview over the workflow. Cycles in the net give a hint, where there is danger of potentially high search effort.

#### 4. RESULTS

The image in Fig 7 shows a scene section obtained further to the left side of the section displayed in the image in Fig. 1. The same groups of cars are seen in the right margin of Fig. 7 and the left margin of Fig.1 In Fig. 8 this group appears in the upper central part. This figure displays all spots which are accumulated in the scene from approx. 40 images and the objects m\_LINE of the map. Cold spots are drawn in black and warm spots are drawn in white. Many of the hot spots result from structures on the roofs. The percentage of hot spots in relation to all spots is 34%.

Using the constrains described above only the spots displayed in Fig. 9 remain as cars. All Spots from roofs and some spots from roads disappear. The percentage of active cars in relation to all cars is 3%.

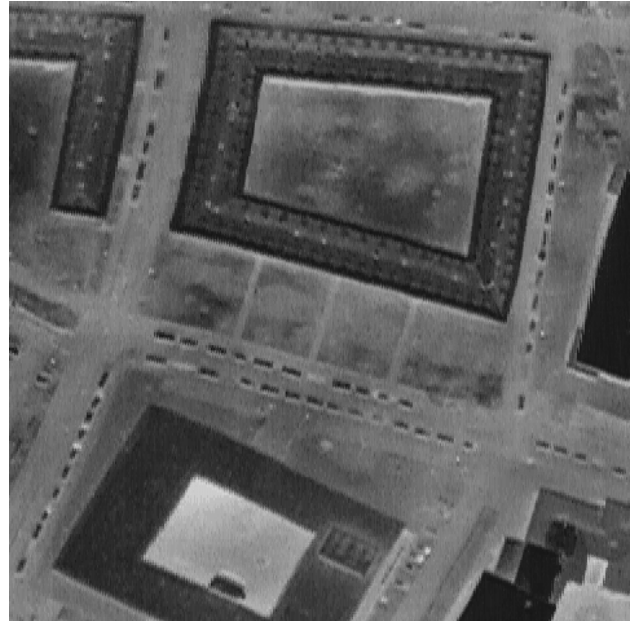


Fig. 7. Other IR image of the same sequence. The images shows a scene section left of scene section displayed in Fig. 1.



Fig. 8. Section of the scene with cold and hot spots (comp. Fig. 5b). The cars are contained, but also many warm spots on the roofs.

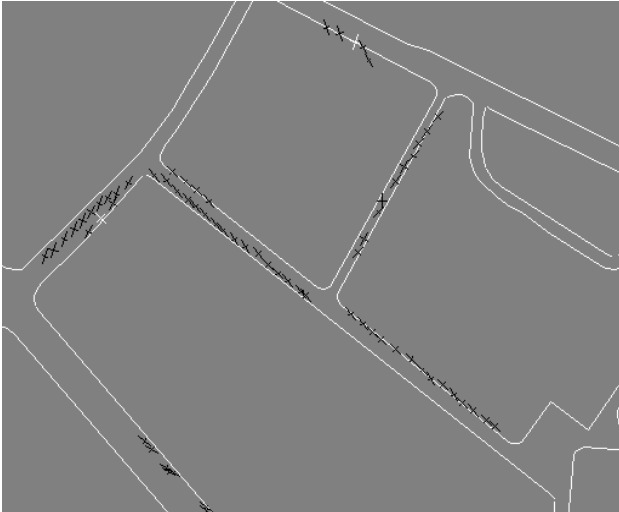


Fig. 9. Result of grouping along the margin of the roads in the scene (comp. Fig 5d). All spots correspond to cars. Only two spots are warm. The activity is estimated low.

## 5. DISCUSSION

Crucial for the success of such a system is an adequate choice of the parameters [Michaelsen & Stilla, 2000]. One example of such a parameter is the width of the interest regions around the road margins. This should be chosen with respect to the maximal errors of the registration and additionally with respect to the distance from the road margins in which cars typically appear. If it is chosen too small with respect to these criteria, a large portion of cars will not be detected. If this parameter is chosen too liberal, spots on roofs will be mistaken for cars. First results show that not all cars are found. The parameter setting was tuned to avoid false alarms. E.g., we only accept groups with at least four cars. Estimating the number of cars would require a different parameter setting. The estimation of the ratio of active versus all cars only needs a sample of the cars present in the scene.

The group of cars on the right margin of the road in the lower left region of Fig. 7 results in a distributed cluster of spots in the scene (Fig. 8, left). It leads to two groups of cars on either side of the road (Fig. 9, left). We blame this fault to inexactness in the determination of the external camera orientation. Because GPS-INS data were not provided with this IR image sequence, we estimated the orientation data on every tenth image manually and interpolated in-between. Provided that orientation data are recorded with sufficient accuracy such faults can be avoided.

Perceptual grouping and the utilization of additional data sources like maps can help in the automatic interpretation of airborne high-resolution thermal image sequences. Even in highly structured urban areas it becomes possible to discriminate vehicles from other objects and determine their activity. Working in the scene instead of the image has advantages, e.g. evidence from multiple images can be accumulated.

The test data set shows a scene with rather low activity. To make assessments on how the method works on scenes with high activity further data have to be processed, e.g. scenes containing major roads with considerable traffic and scenes containing parking lots of shopping centers in the opening hours.

In scenes with moving cars tracking the cars and the determination of their velocity will be another important discriminative attribute. Active vehicles will move preferably on roads at speeds, for which a reasonable bound is given. So hot spots of considerable mass and with their velocity fulfilling these criteria can be counted as active vehicles, even if they are not members of a row.

## 6. REFERENCES

Dreschler L, Nagel HH (1982) Volumetric model and trajectory of a moving car derived from monocular TV frame sequence of a street scene. *CGIP*, 20, 199-228.

Haag M, Nagel HH (1999) Combination of edge element and optical flow estimates for 3D-model based vehicle tracking in traffic image sequences. *IJCV*, 35:3, 295-319.

Hinz S, Baumgartner A (2001) Vehicle detection in aerial images using generic features, grouping and context. In: Radig, B., Florczyk S., Pattern recognition, DAGM 2001, LNCS 2191, Springer, Berlin, 45-52.

Kohnle A, Neuwirth W, Schuberth W, Stein K, Höhn DH, Gabler R, Hofmann L, Euing W (1993) Evaluation of essential design criteria for IRST systems. 76/SPIE Vol.2020 Infrared technology XIX.

Michaelsen E, Stilla U (2000) problems in geometric modeling and perceptual grouping of man-made objects in aerial images. *International archives of photogrammetry and remote sensing*, vol. 33, part B3, 577-583.

Partsinelovos P, Agouris P, Stefanidis A (2000) Modelling movement relations in dynamic urban scenes. *International archives of photogrammetry and remote sensing*, vol. 33, part B4, 818-825.

Quint F (1996) Recognition of structured objects in monocular aerial images using context information. In: Leberl F, Kallianay R, Gruber M (eds) Mapping buildings, roads and other man-made structures from images. IAPR TC-7 workshop in Graz, Oldenburg, Wien, 213-228.

Stilla U (1995) Map-aided structural analysis of aerial images. *ISPRS Journal of photogrammetry and remote sensing*, 50(4):3-10

Stilla U, Michaelsen E (1997) Semantic modeling of man-made objects by production nets. In: Gruen A, Baltasavias EP, Henriksen O (eds) Automatic extraction of man-made objects from aerial and space images (II). Basel: Birkhäuser, 43-52

Stilla U, Soergel U, Jaeger K (2000) Generation of 3D-city models and their utilization in image sequences. *International archives of photogrammetry and remote sensing*, vol. 33, part B2, 518-524.