Dynamic Scene Understanding for Mobile Robot Navigation

Ondrej Miksik

Laboratory of Telepresence and Robotics & Computer Vision Group, Faculty of Electrical Engineering and Communication, Brno University of Technology, Brno, Czech Republic

Self-supervised Learning for General Road Extraction

- Suitable for primarily teleoperated mobile robots, e.g. Orpheus-AC (military reconnaissance mobile robot)
  - Automatic return from teleoperated mission in case of signal loss
- System demands
  - Diverse light conditions (direct sunlight, strong shadows, . . .)
  - Structured and unstructured roads (gravel, tarmac, . . .)

1. Texture flow estimation
   - A bank of self-similar Gabor wavelets decomposed into linear combinations of Haar-like box functions
   - Efficient computation – integral images; over-complete dictionary: NP-hard → OOMP
2. Vanishing point voting & Smoothing
   - Coarse-to-fine voting scheme reduces computational complexity
   - Smoothing filter CONDENSATION reduces the influence of noise and the jumpy characteristics of output

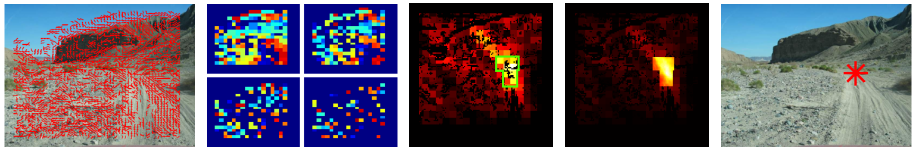


Figure 1 : Vanishing Point – texture flow (a), superpixels (b), coarse-to-fine voting (c) and (d), output (f)

3. Road Extraction – Gaussian Mixture Model (GMM)
   - Vanishing point determines the non-static training area (cf. Fig. 3)
   - Color models are constructed from sample pixels by self-supervised learning algorithm and adaptively updated
   - A few simple rules define properties of the color segmentation system (adaptivity speed, selectivity, robustness or behavior in shady and/or overexposed highlighted road segments)

Results - Adaptivity & Robustness

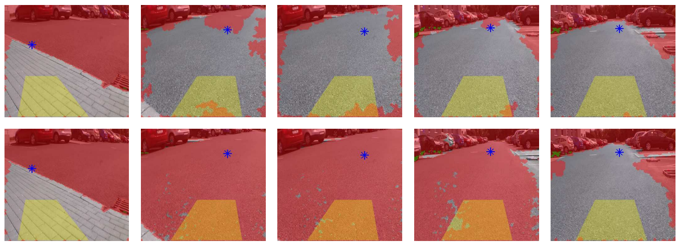


Figure 2 : Anti wind-up and decay (top) and without (bottom)

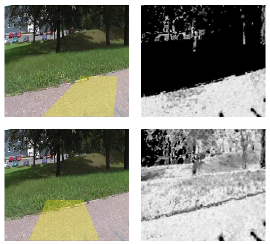
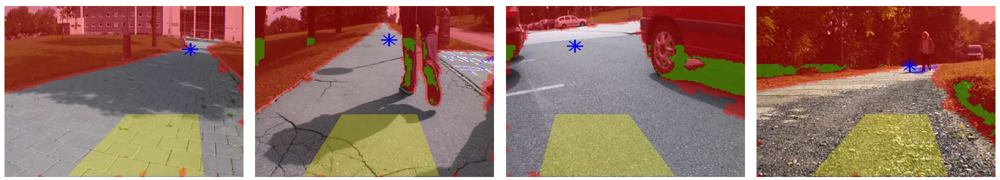


Figure 3 : Training area

Results (road/non-road regions)



Spatio-temporal Consistency for Total Scene Understanding

- The vision systems for advanced applications should provide
  - More reliable predictions
  - Predictions should be consistent in both, space and time
  - Information about the semantic classes present in the scene: objects (cars, pedestrians, etc.) and stuff (sky, grass, etc.)
- The dynamic scene understanding can be formalized as
  - A set of uncalibrated monocular images $\mathcal{I} = \{\mathbf{i}^{(1)}, \mathbf{i}^{(2)}, \dots, \mathbf{i}^{(n)}\}$
  - Random variables over data $\mathbf{i}^{(t)} = \{\mathbf{x}_1^{(t)}, \mathbf{x}_2^{(t)}, \dots, \mathbf{x}_m^{(t)}\}$
  - Assign a unique label l_i from $\mathcal{L} = \{l_1, l_2, \dots, l_k\}$
  - Label l_i represents a sample that corresponds to the highest probability of the random variable over the labels $\mathbf{Y}_i^{(t)} = \{y_{i,1}^{(t)}, y_{i,2}^{(t)}, \dots, y_{i,k}^{(t)}\}$ consistent in both, space and time

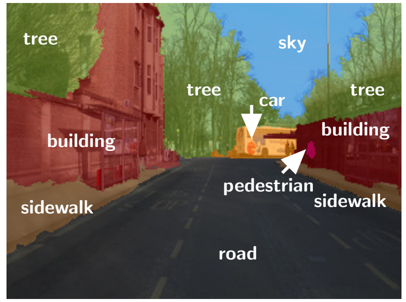


Figure 4 : Output of our system

1. Labels are propagated from frame $t-1$ to frame t (large displacement optical flow)
   - Matching pixels – not all pixels in a neighborhood are matches

2. Learning similarity metric
   - The standard radial basis function kernel is usually used to express the similarity between the features $\mathbf{f}_i$ and $\mathbf{f}_j$
   - Color features and Euclidean distance are not sufficient to distinguish small objects from the background

   - We need a better feature representation (LBPs, textons, . . .) and a novel similarity metric based on a Mahalanobis distance parametrized by matrix $\mathbf{M}$ obtained with off-line subgradient optimization
   - We aim to obtain an $\mathbf{M}$ that results in small distances between the features that belong to the same semantic class and a large distance between the others
   - Sparse bundle adjustment → positive $\mathcal{E}_p$ (matching points) and negative $\mathcal{E}_p$ (distractors) examples
3. Temporal smoothing
   - Measurement: Mahalanobis distance parametrized by matrix $\mathbf{M}$ with radial basis function
   - Update:

$$\hat{\mathbf{Y}}_i^{(t)} = (1-\lambda)\frac{1}{Z_1}\left[\sum_{j\in N_i} w_{ij}\hat{\mathbf{Y}}_j^{(t-1)} + c\mathbf{Y}_i^{(t)}\right] + \lambda\frac{1}{Z_2}\left[\sum_{j\in<t-s_1,t+s_2>} w_{ij}\mathbf{Y}_i^{(j)} + c\mathbf{Y}_i^{(t)}\right] \quad (1)$$

Appeared in *Proceedings of the IEEE ICRA 2011, ICMLA 2011, ICRA 2012, ICPR 2012*

ondra.miksik@gmail.com

http://www.miksik.co.uk