Driven to Distraction: Self-Supervised Distractor Learning for Robust Monocular Visual Odometry in Urban Environments

Dan Barnes, Will Maddern, Geoffrey Pascoe and Ingmar Posner dbarnes@robots.ox.ac.uk

Objective

Robust visual odometry in urban environments with only a monocular camera.

Challenges

Visual-only approaches to motion estimation often **fail** with large moving distractor (***ephemeral***) objects.

In this example the bus causes our visual odometry (VO) to fail.

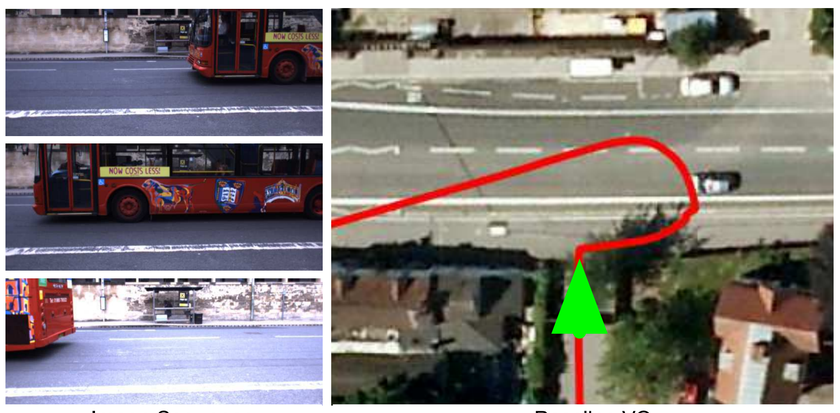


Image Sequence

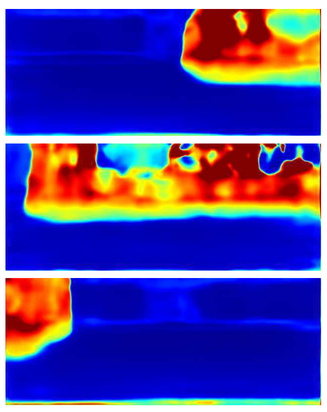
Baseline VO

Solution

Train a CNN to predict:

- pixel-wise ***ephemerality masks*** for each image to ignore distractor objects.

- ***disparity*** to give scale when using only a monocular camera.

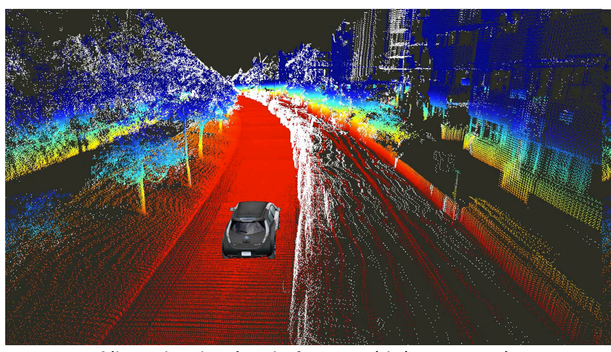


Ephemerality Mask

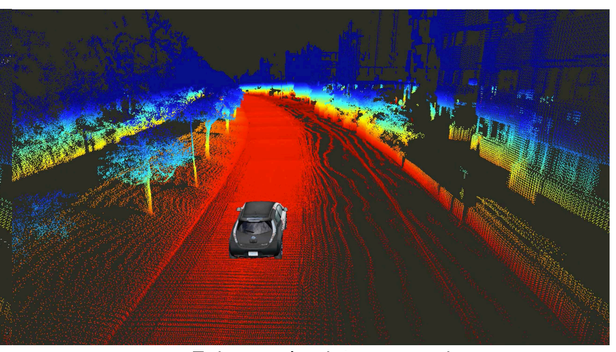
Robust VO

Learning Ephemerality Masks

1) Prior 3D Mapping - We align multiple traversals of our environment with an offline multi-session mapping system and use an entropy-based approach to determine what constitutes the static (non-ephemeral) structure of the scene.

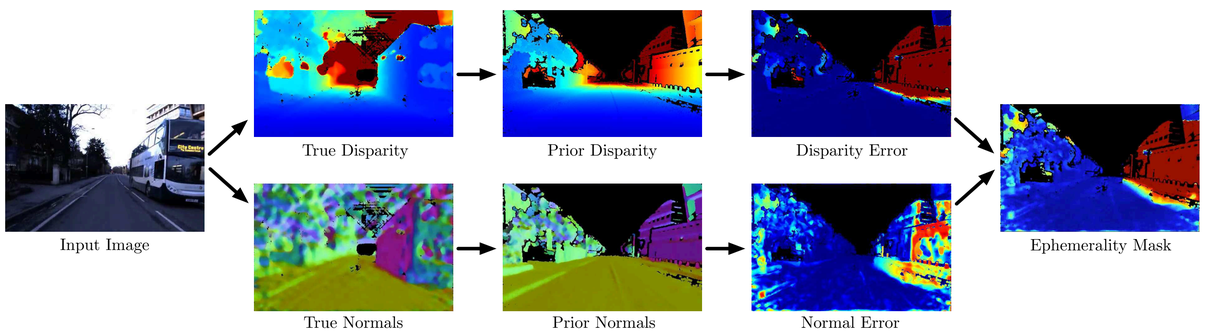


Aligned pointclouds from multiple traversals

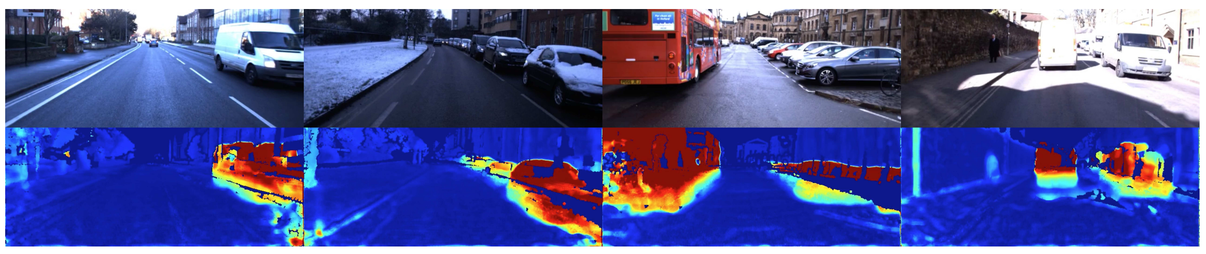


Ephemeral points removed

2) Ephemerality Labelling - We project the static structure into collected stereo camera images. In the presence of traffic or dynamic objects these differ considerably and we compute ephemerality as a weighted sum of disparity and normal differences.



3) Network Training - We train a deep convolutional network to predict pixel-wise disparity and ephemerality masks with only monocular input images. As a self-supervised approach, we can generate vast quantities of training data covering **lighting, weather and traffic** conditions **without manual labelling**.



Ephemerality mask training data

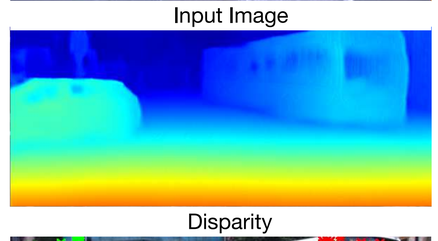
Deployment

We leverage the live ***disparity*** and ***ephemerality mask*** produced by the network to produce reliable VO estimates accurate to metric scale with ***only a monocular camera.*** We integrate our approach into two different odometry systems:

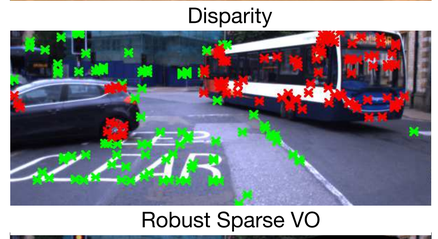
1) Sparse VO - Feature based VO extracts feature points from images and tracks them across an image sequence. We use the predicted ephemerality mask to disable features which are unlikely to belong to the underlying static structure (red crosses) using only the remaining stable features (green crosses).

2) Dense VO - Direct based VO uses the pixel intensity values directly rather than extracting explicit features. We use the predicted ephemerality mask to directly to weight the photometric residual; no thresholding is required. The darker regions illustrate the high ephemerality predictions.

Input Image



Disparity

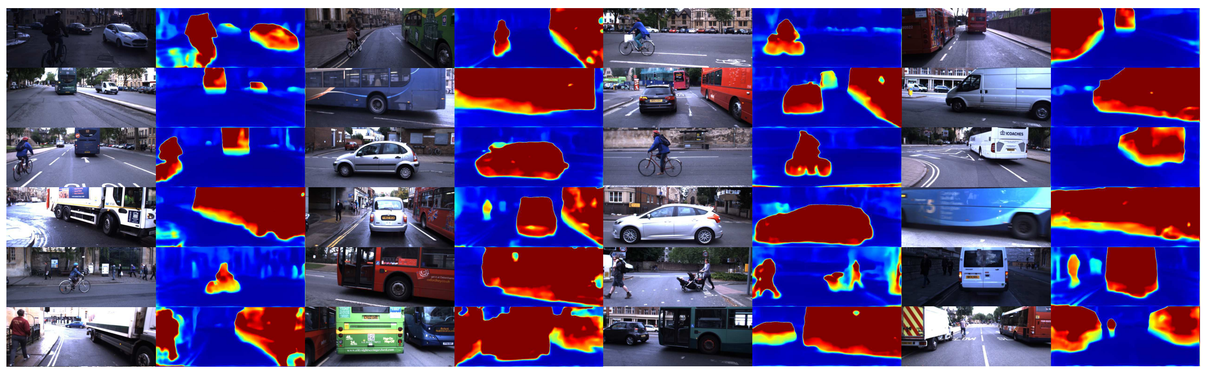


Robust Sparse VO

Robust Dense VO

Results

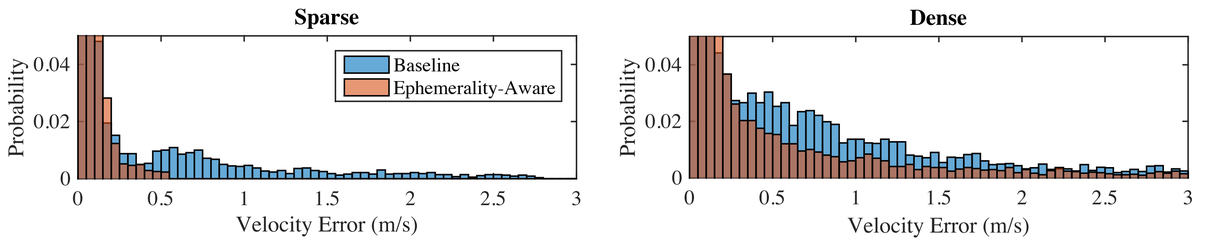
Qualitative ephemerality mask predictions in challenging urban environments. The masks reliably highlight a diverse range of dynamic objects (cars, buses, trucks, cyclists, pedestrians, strollers) with highly varied distances and orientations.



Qualitative ephemerality mask predictions

Evaluated over 400km of driving from the Oxford RobotCar Dataset, we demonstrate **reduced odometry drift** and **significantly improved egomotion estimation** in the presence of large moving vehicles in urban traffic.

Of particular note, our robust sparse VO approach is almost unaffected by distractors, whereas the baseline method reports errors over four times greater.



Velocity estimation errors in the presence of distractors. The vertical axis is scaled to highlight the outliers.

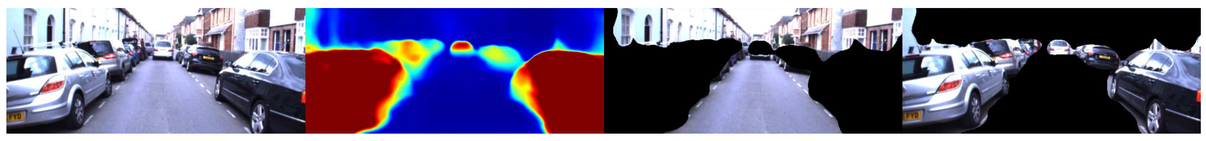
Conclusion

We introduce the concept of ***ephemerality masks***, which estimate the likelihood that any pixel in an input image corresponds to either reliable or static structure.

We use an entirely ***automatic self-supervised approach*** to train our system and do not require any manual labelling.

At run-time we only require a ***single monocular camera*** to produce ***reliable ephemerality-aware visual odometry to metric scale***.

We suggest that ***ephemerality masks*** could be utilised in other applications such as localisation, object detection and scene understanding.



Ephemerality masks could be used for foreground / background segmentation in other applications

UNIVERSITY OF OXFORD

A²I APPLIED ARTIFICIAL INTELLIGENCE LAB
OXFORD ROBOTICS INSTITUTE