ESTIMATION OF TERRAIN GRADIENT CONDITIONS & OBSTACLE DETECTION USING A MONOCULAR VISION-BASED SYSTEM

Final Demonstration

Connor Luke Goddard (clg11)

Department of Computer Science, Aberystwyth University

Background

The ability for an autonomous robot to navigate from one location to another in a manner that is both **safe**, yet **objective** is vital to the **survivability** of the machine, and the **success of the mission** it is undertaking.

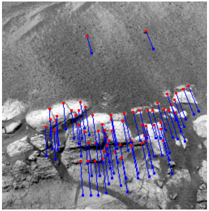


Figure 1: Tracked features from MER [2].

Vision based obstacle detection has enjoyed a high level of research focus over recent years. Less work has concentrated on providing a means of reliably detecting changes in terrain slope through the exploitation of observed changes in motion.

RELATED WORK

[1] *A Robust Visual Odometry and Precipice Detection System Using Consumer-grade Monocular Vision*, J. Campbell *et al.*, IEEE, 2005

[2] *Obstacle Detection using Optical Flow*, T. Low and G. Wyeth, School of Information Technology and Electrical Engineering, University of Queensland, 2011

[3] *Appearance-based Obstacle Detection with Monocular Color Vision*, I. Ulrich and I. Nourbakhsh, AAAI, 2000

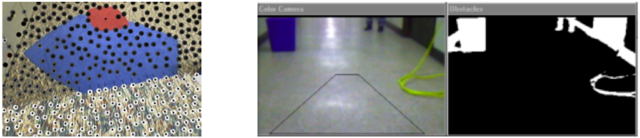


Figure 2: **Left:** Precipice detection [1], **Right:** Obstacle colour detection [3]

Focus

Investigation into a system capable of utilising a single, forward-facing colour camera to provide an estimation into current terrain gradient conditions, obstacle location & characteristics and robot ego-motion.

Potential Metrics For Observation

- Changes in terrain gradient (i.e. slopes).
- Location & characteristics of positive and negative obstacles (i.e. rock or pit respectively).
- Current robot speed of travel.
- Changes in robot orientation.

Working Hypothesis

WORKING HYPOTHESIS

"From images captured using a ***non-calibrated*** *monocular camera system, analysis of optical flow vectors extracted using* ***localised appearance-based*** *dense optical flow techniques can provide certain estimates into the current condition of terrain gradient, the location and characteristics of obstacles, and robot ego-motion."*

Approach focussed on the effects of **motion parallax**.

i.e. Objects/features that are at a greater distance from the camera appear to move less from frame-to-frame than those that are closer.

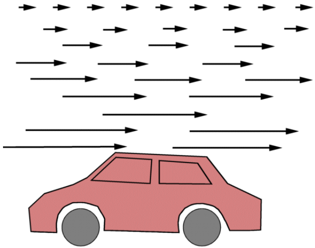


Figure 3: Typical example of motion parallax [4].

Key Aspects

1. The exclusive use of appearance-based template matching techniques to provide a localised variant of dense optical flow analysis over the use of sparse optical flow techniques relying upon a high number of available image features.
2. The use of a formalised model to represent detected changes in vertical displacement as part of efforts to estimate changes in terrain gradient in addition to the location characteristics of potential obstacles.

Provides a mechanism by which to evaluate the prediction:

*"Features captured towards the **bottom** of an image will show **greater displacement** between subsequent frames, than those features captured towards the **top** of the image."*

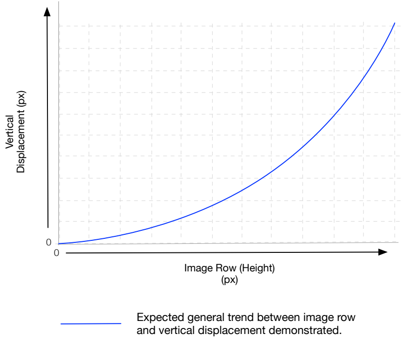


Figure 4: Example of "perfect" vertical displacement model.

From observing discrepancies between the current model and "baseline" model, it should be possible to infer the presence of potential obstacles:

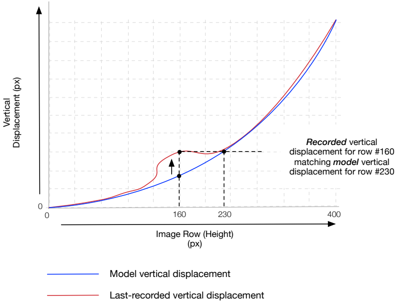
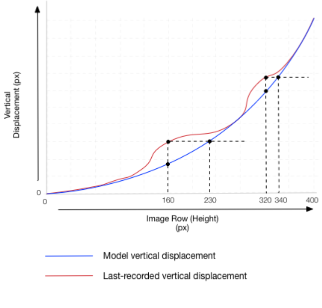


Figure 5: Vertical displacement model indicating the potential presence of a positive obstacle.

Differences in model discrepancy behaviour should provide enough information to establish between a potential obstacle and a change in terrain slope:



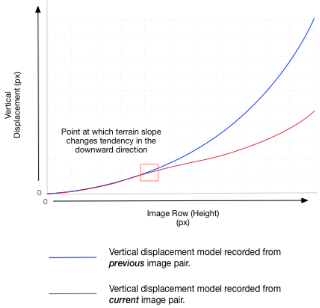


Figure 6: Example of observed differences between detection of obstacles and change in terrain slope.

Two types of observed motion:

- **Translation** (*Vertical* component of optical flow vectors)
- **Rotation** (*Horizontal* component of optical flow vectors)

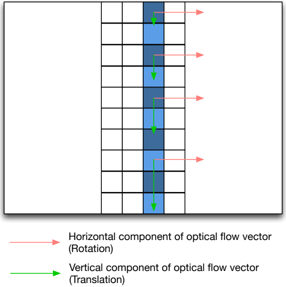


Figure 7: Diagram demonstrating the expected difference in behaviour between the separated vertical and horizontal components of observed optical flow vectors calculated from *forward-turning motion*.

By recording multiple “calibration” models at different speeds, it should be possible to later infer the current speed using the standard *speed-distance-time* equation.

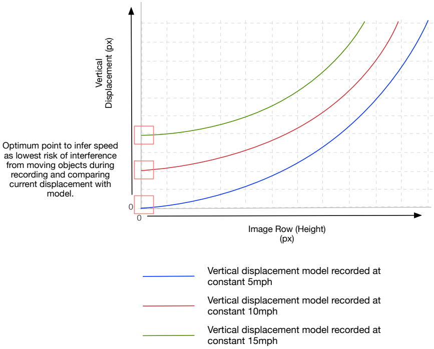


Figure 8: Diagram indicating examples of alternative vertical displacement models calculated while travelling at different constant speeds.

Primary Aims

1. Establish which out the following appearance-based template matching similarity measures:
    - Euclidean Distance
    - Correlation Coefficient
    - Histogram Correlation
    - Histogram Chi-Square

    best supports the predicted positive correlation between the vertical position of features within an image, and the level of vertical displacement demonstrated.
2. Use the generated model of vertical displacement to identify potential obstacles and changes in terrain gradient based upon the predicted ‘comparative trend’ behaviour.

Secondary Aims

1. Further extend the capabilities of the vertical displacement model to provide estimates into the current speed of travel demonstrated by the robot.
2. Add support for providing an estimation into the change in orientation.

Experiment Methods

GATHERING DATASETS

No existing imagery datasets were found to be suitable, therefore new sets had to be collected.

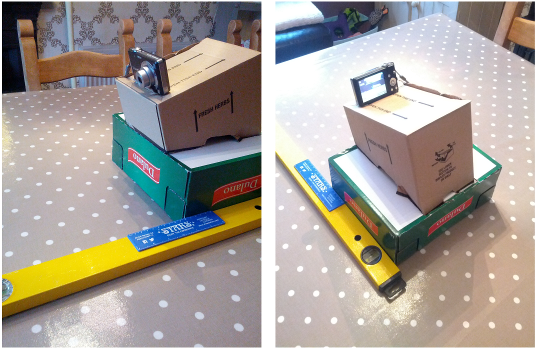


Figure 9: Camera-rig setup used to capture the experiment datasets from front and back profiles.

Images taken at a number of locations, but **four** were eventually selected based upon their variety in terms of **terrain texture** and **lighting conditions**.

Site 1: Living Room Rug | Site 2: Brick Paved Road | Site 3: Asian Rug | Site 4: Slate Tile Footpath

Figure 10: Examples of the first frame captured from each of the four location datasets.

Over time, investigation focus changed from exploring a series of reasonably "generalised" research aims, to instead concentrating almost exclusively on one small, but very important aspect;

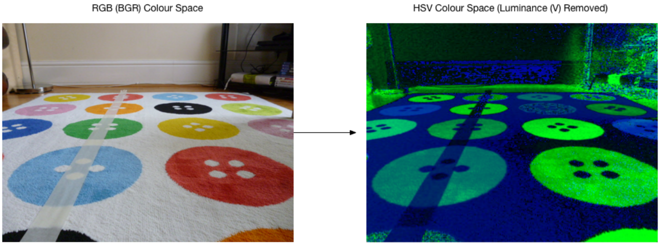
Ensuring that accurate results for appearance-based motion displacement could continue to be achieved, across terrain lacking any significant features and/or demonstrating highly repetitive patterns.

Three Experiments

1. Template Matching - Multiple Small Patches
2. Template Matching - Full-width Patches (Non-Scaled)
3. Template Matching - Full-width Patches (Geometrically Scaled)

Stage One

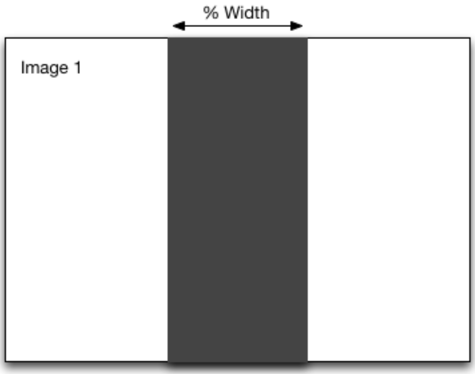
Import two consecutive images, and convert from RGB (BGR in OpenCV) colour space to HSV.



The 'V' channel is then removed in order to improve robustness to lighting changes between frames.

Stage Two

Extract a percentage-width region of interest (ROI) from centre of first image.



Why do we need to extract a ROI?

Focus-of-expansion: Objects in images do not actually move in 1-dimension (i.e. straight down the image).

This effect is minimised towards the centre of the image.

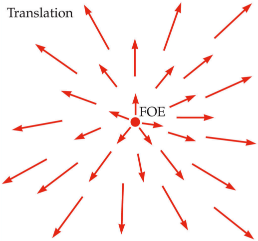


Figure 11: Diagram indicating the perceived effects caused by Focus of Expansion. Courtesy: [5]

Stage Three

Extract patches of a fixed size around each pixel within the extracted ROI.

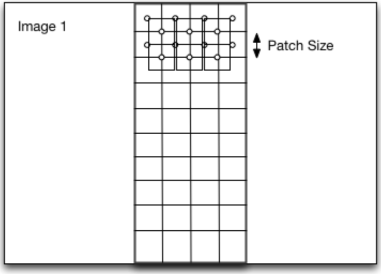


Figure 12: Simplified example of patch extraction within ROI.

Stage Four

For each patch extracted from *image one*, move down through a localised search window (column) in *image two* searching for the best match against the template patch.

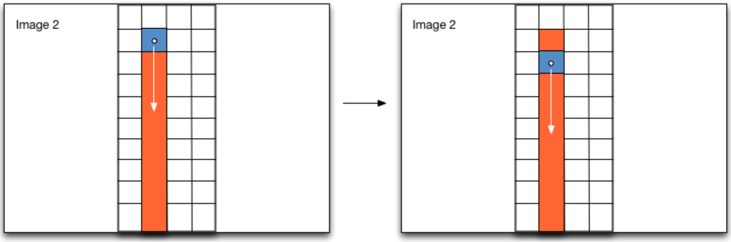
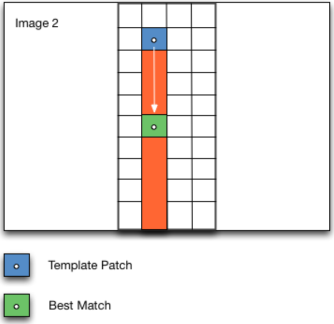


Figure 13: Example of “best match” search within local column.

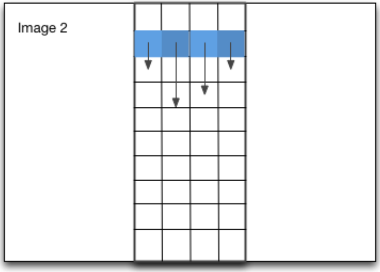
Stage Five

Identify the patch within the localised search window that provides the “best match” via correlation-based matching (e.g. Euclidean Distance, SSD or Correlation coefficient).



Stage Six

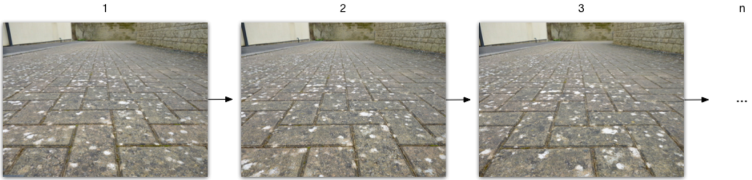
Average all measured displacements for each pixel along a given row.



Outliers are removed by ignoring any displacements that lie outside of (2 x Standard Deviation) of the mean.

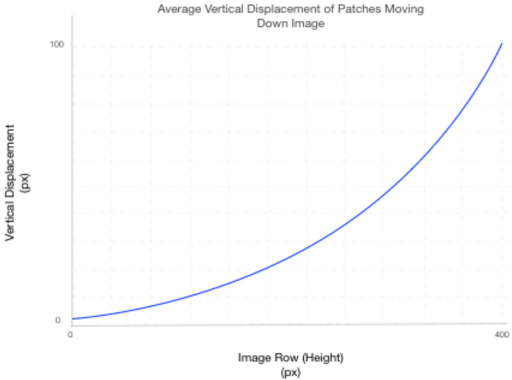
Repeat Stages 1-6

Repeat stages 1-6 for an entire collection of “calibration” images taken of *flat, unobstructed* terrain.



Stage Seven

Plot the *average displacement* for each ROI row, calculated from the displacements recorded over all calibration images.



Focus

Establishing if adopting ***larger*** patches in fewer numbers provided better appearance-based matching accuracy than using many more, but critically much ***smaller*** overlapping patches.

While the underlying method remained the same as in the first experiment, some key changes were made:

- Extracting a central region of interest relative to the **ground plane**, as opposed to extracting a fixed-width column from the image plane.
- Moving away from multiple overlapping small patches, in favour of adopting a single, **full-width** patch to represent a single row in the image.

Issue discovered regarding the use of **non-exhaustive** searching causing erroneous results. All tests in experiments two and three were conducted using both exhaustive and non-exhaustive approaches.

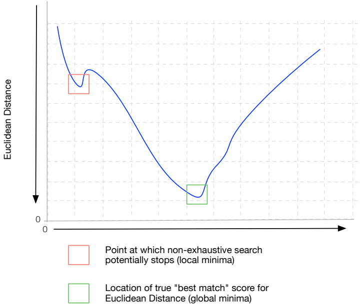


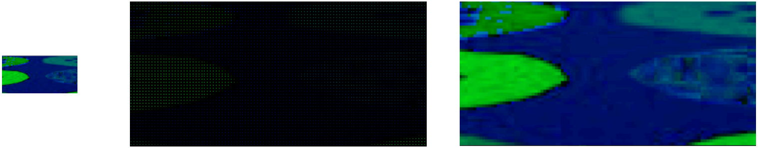
Figure 14: Example of potential early termination of non-exhaustive search due to entering local minima caused by noise, rather than global minima which represents to true best match.

To extract a region-of-interest along the *ground plane*, the system would need to take account of **perspective distortion**. Therefore, a simple **calibration tool** was implemented, enabling users to define the required region of interest within an image.

Figure 15: Results of the verification test for the approach towards geometric scaling of template patch pixel coordinates.

Focus

Add **geometric scaling** of template patch in order to account for objects appearing larger as they approach the camera.



Original Template Patch   Re-scaled Template Patch (Geometric Transformation of Pixel Coordinates)   Re-scaled Template Patch (Linear Interpolation)

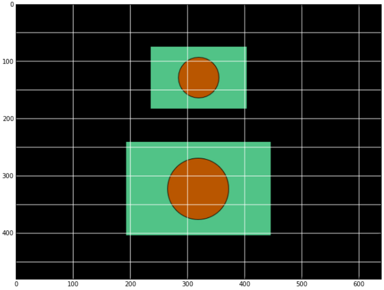
Figure 16: Comparison between various approaches of performing scaling of the original template patch image. In the case of linear interpolation, additional information has been added to image as a result of “estimating” the colour that lies between two original pixels.

6 Stage Process

1. Obtain the width and height of the current template patch.
2. Obtain the calibrated width of the search window with respect to the current image row.
3. Calculate the independent scale factor for the width and height between the template patch and the search window.
4. Treating each pixel as a 2D coordinate in geometric space, scale the position of each pixel coordinate within the template patch relative to the position of centre coordinate.
5. For each scaled pixel coordinate, extract the value of the pixel at the corresponding position with the search window and add to a temporary "image" data structure.
6. Perform template matching between the original template patch, and the new "extracted" temporary image.

Testing of Scaling Approach

A simple test was devised to confirm geometric scaling worked as expected under “perfect” conditions.



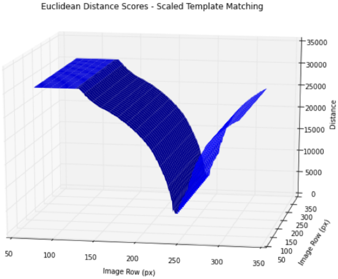


Figure 17: Results of the verification test for the approach towards geometric scaling of template patch pixel coordinates.

ADDITIONAL EXPERIMENTAL WORK

Some additional experimental work was conducted as part of the larger project investigation, but these did not actively contribute to the final investigation results.

- Investigation of Campbell *et al.* [1] approach to feature tracking to provide optical flow field (based heavily upon C# implementation provided by Dr. Rainer Hessmer (Source))
- Investigation of edge-based template matching discussed by Perveen *et al.* [3]

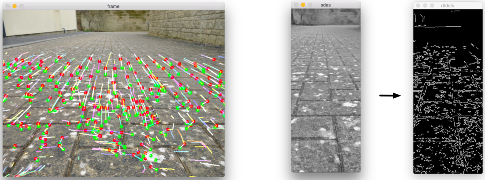


Figure 18: Left: Example of feature tracking to establish optical flow field. **Right:** Edge-based template matching using Canny edge detector.

Investigation Results

View Results (Courtesy: nbviewer)

Results can be described as generally inconclusive.

However from the results obtained, we have learned:

1. Under certain terrain conditions, it is possible to potentially establish a relationship between row height in an image, and average downwards pixel displacement.

2. The performance of each appearance-based similarity measure can vary widely, based upon factors including:
   - Terrain texture (e.g. highly patterned, isotropic or homogenous)
   - Patch size (Balance between reducing “noise” and representing true displacement behaviour)
   - Exhaustive vs. Non-exhaustive search
   - Scaling vs. Non-scaling

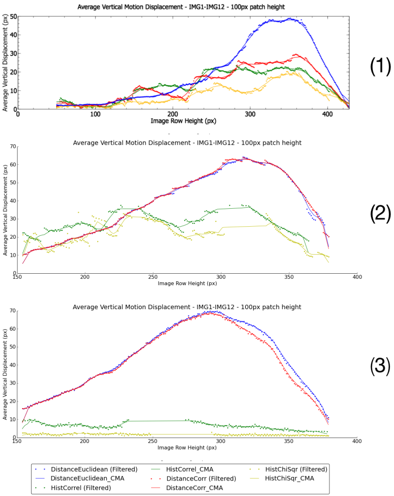


Figure 19: Comparison of results for the ‘Living room rug’ dataset using an exhaustive search and 100px patch size. Small patches (Graph 1), full-width patches (Graph 2) and geometrically scaled full-width patches (Graph 3).

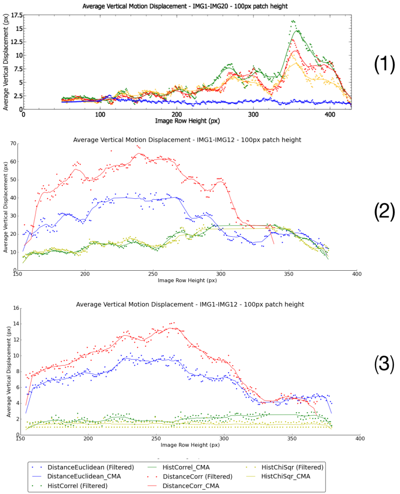
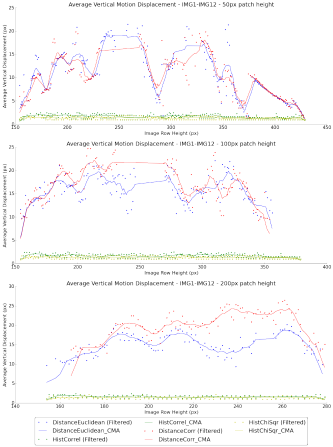


Figure 20: Comparison of results for the ‘Brick-paved road’ dataset using an exhaustive search and 100px patch size. Small patches (Graph 1), full-width patches (Graph 2) and geometrically scaled full-width patches (Graph 3).

Non-Exhaustive



Exhaustive

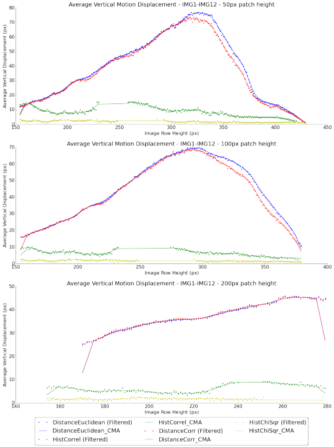


Figure 21: Comparison between non-exhaustive and exhaustive search performance using geometrically scaled template patches and the ‘Living room rug’ dataset.

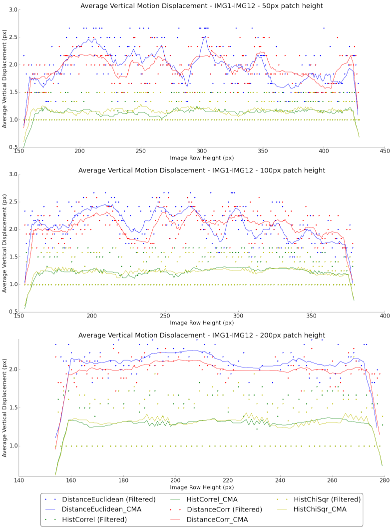


Figure 22: Comparison of results for the ‘Living room rug’ dataset using geometrically-scaled patches of 50px, 100px and 200px heights under an exhaustive search approach.

Conclusion

FUTURE WORK

Great potential for future work, most likely focussing on areas including:

- Further experimentation to establish most appropriate combination of similarity-measure, search approach and patch size for a **given terrain type**.
- Development of a system to identify “significant” changes in vertical displacement model that are potentially indicative of a potential obstacle or change in terrain gradient.
- Development of system to estimate change in robot heading.
- Development of system to estimate current robot speed.

Project has laid the groundwork upon which further development can subsequently be conducted.

While not all of the original aims were accomplished, research efforts had to be focussed on confirming the underlying hypothesis, which as a general concept, has been proven.

Not one single approach or solution will suffice for all terrain conditions.

Questions?

Project Blog

Slide Design: Matthias Vogelgesang - (Github)

[1] Jason Campbell, Rahul Sukthankar, Illah Nourbakhsh, and Aroon Pahwa. A robust visual odometry and precipice detection system using consumer-grade monocular vision. In *Robotics and Automation, 2005. ICRA 2005. Proceedings of the 2005 IEEE International Conference on*, pages 3421–3427. IEEE, 2005.

[2] Mark Maimone, Yang Cheng, and Larry Matthies. Two years of visual odometry on the mars exploration rovers. *Journal of Field Robotics*, 24(3):169–186, 2007.

[3] Nazil Perveen, Darshan Kumar, and Ishan Bhardwaj. An Overview on Template Matching Methodologies and its Applications. *IJRCCT*, 2(10):988–995, 2013.

[4] Jonathan Pillow. Motion Perception 2. `http://homepage.psy.utexas.edu/homepage/faculty/pillow/courses/perception09/slides/Lec13A_Motion_part2.pdf`, October 2009.

[5] Akshay Roongta. Depth Cue Theory. `http://akshayroongta.in/notes/depth-cue-theory/`, October 2013.