# Quasi-thematic Feature Detection And Tracking For Future Rover Long-Distance Autonomous Navigation

**Authors:** Affan Shaukat, Conrad Spiteri, Yang Gao, Said Al-Milli, and Abhinav Bajpai

**Surrey Space Centre, University of Surrey**

May 10, 2013

# Contents

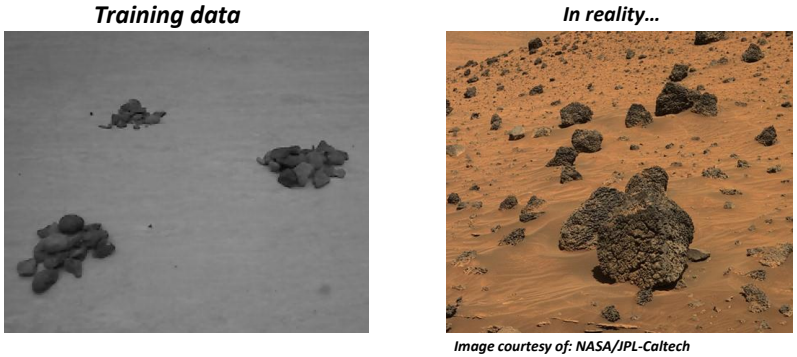
1. Introduction
   - Problem
   - Research overview
   - Performance Evaluation
2. Visual Saliency-based Rock Detection & Tracking
   - Bottom-up visual saliency
   - Binary Maps (Rock Detection)
   - Object Tracking
3. Blob (Shape-based) Detection and Tracking
   - Object Detection (Method of Moments)
   - Object Tracking
4. Quantitative Evaluation & Results
   - Object Detection Accuracy (N-MODA)
   - Object Tracking Accuracy (MOTA)
   - Datasets
   - Results & Conclusion

# Detect & Track Rocks on Planetary Surfaces

- Autonomous navigation systems for planetary rovers need to detect, track and avoid obstacles (e.g., rocks)
- Vision-based perceptual inputs deemed very useful for (rover) auto-navigation (e.g., MER & MSL missions)
- More importantly: requirement of robust detection and tracking of objects in visual scenes for (not limited to):
  - Obstacle detection and avoidance (e.g., *rocks*)
  - Visual odometry
  - Path planning
  - Path following and autonomous navigation
  - SLAM (loop closures etc)

# Problems...

- Use standard supervised learning techniques (e.g., SVMs, Tree learning, Gaussian classifiers etc), however...



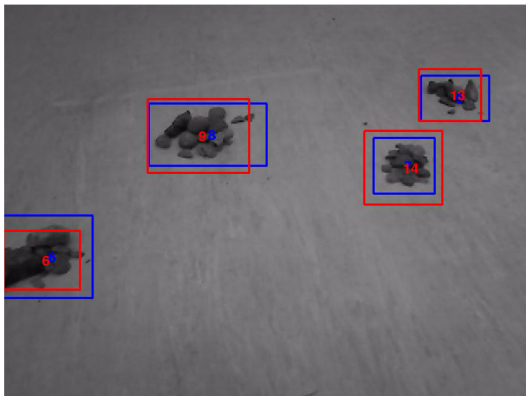
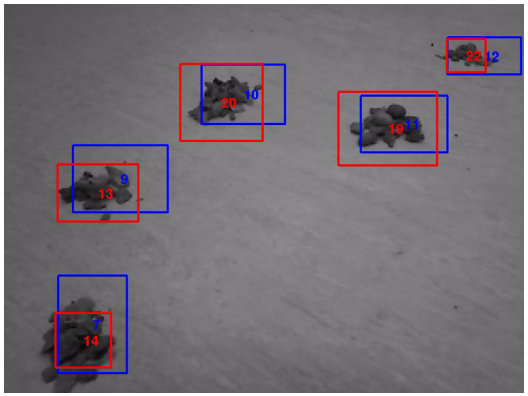
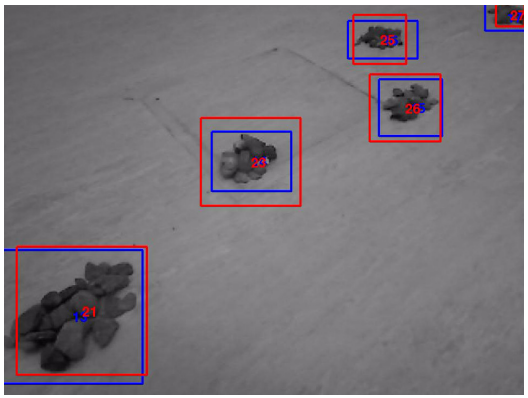

- Use over saturated contextual features (e.g., SIFT), however...
  - *Complex features detected that describe objects (e.g., rocks)*
  - *Require computationally expensive tracking algorithms*
  - *Higher memory requirements*
- Possible solutions? (Use symbolic descriptors for objects)
  - *Cognitive techniques, saliency maps to describe objects*
  - *Shape-based features to describe objects*

# Rock Detection & Tracking Using Thematic Features

1. **Visual Saliency Based Detection and Tracking:**
   - Bottom-up visual saliency model for object detection (*Itti-Koch-Niebur 98*)
   - Histogram shape-based image thresholding for binary saliency maps (*Otsu's method*)
   - Binary saliency blobs (describing rocks) are used as semantic feature descriptors
   - Use heuristic instance-based search algorithm (*k-NN search*) for tracking these blobs over subsequent frames
2. **Blob (Shape-based) Detection and Tracking:**
   - Image segmentation via binerisation using a threshold selection criterion
   - Contours of individual patches (i.e., blobs) extracted using a border following method
   - **Hu** set of invariant moments computed for each contour
   - **Hu** moments between two subsequent frames collated to achieve tracking

# Quantitative Evaluation Using Ground-truth Data

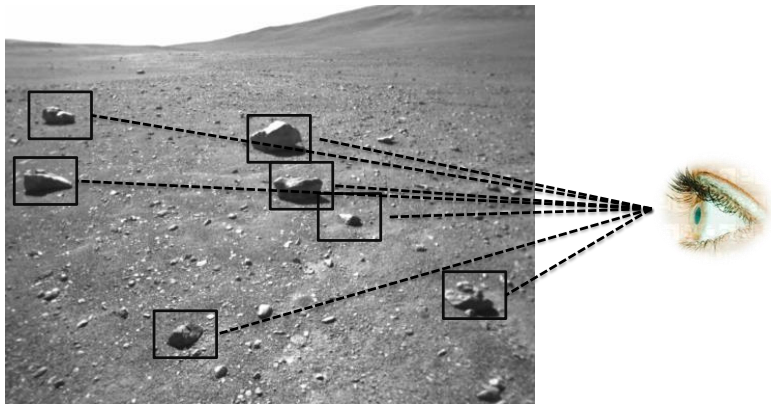
- Quantitative analysis using standard detection/tracking evaluation metrics and protocols set out in [1]
- Datasets used: Lab-based, simulated (PANGU) and real-world (SEEKER) replicating planetary surfaces
- Datasets are hand-labelled using a planetary rock annotation tool purposely built at the Surrey Space Centre

[1] R. Kastsuri *et al.* *"Framework for performance evaluation of face, text, and vehicle detection and tracking in video: Data, metrics, and protocols,"* Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 31, no. 2, pp. 319-336, 2009.

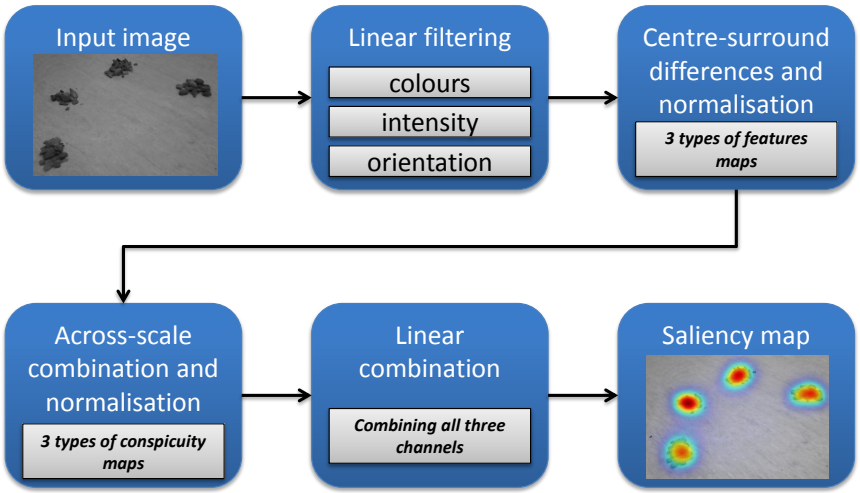
# Visual Saliency-based Paradigms

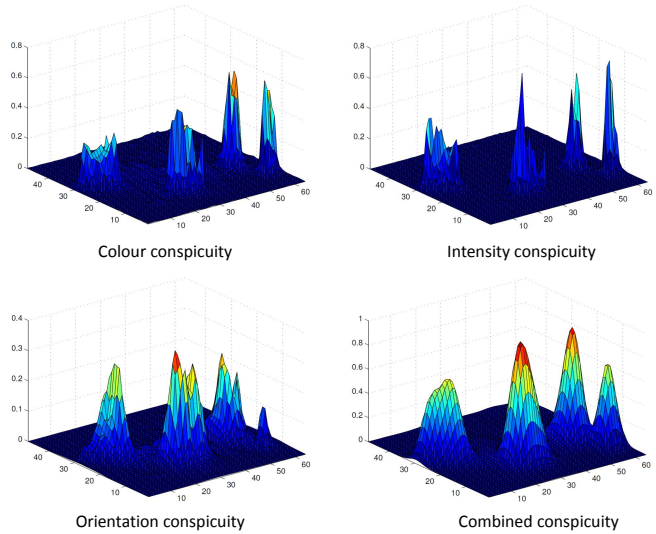
- Inspired by the information selection property of biological visual systems
- Models can either based on computational or cognitive research findings
- A saliency map shows the conspicuity of each pixel in probabilistic terms
- A number of characteristics can act as a stimulus towards conspicuity: *texture, colour, size, shape, orientation etc.*



# Itti-Koch-Niebur model

- We use the Itti-Koch-Niebur model (*Itti-98* to describe rocks in terms of saliency maps [2])
- Uses centre-surround differences across multi-scale image features (*colour, intensity, and orientation*)

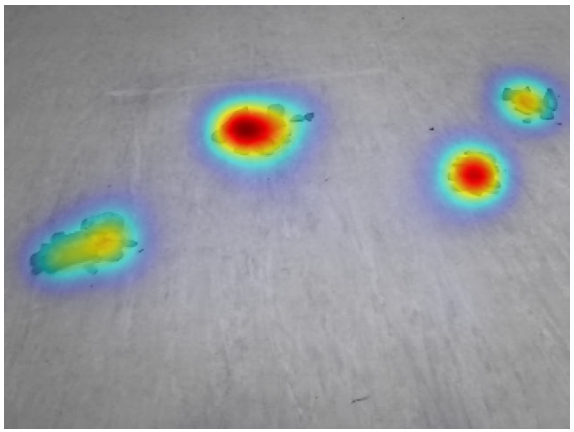


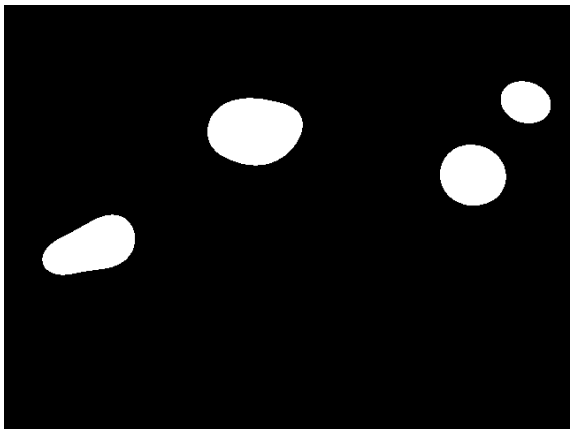




[2] L. Itti *et al.* *"A model of saliency-based visual attention for rapid scene analysis,"* Pattern Analysis and Machine Intelligence, IEEE Transactions on, pp. 1254-1259, 1998.

# Otsu's method

- Histogram shape-based thresholding to convert saliency map into a binary image [3]
- Assumption: Saliency map images have a bimodal distribution (i.e., two classes of pixels: salient objects (rocks) and the background)
- Exhaustive search strategy to compute the optimum threhold that minimises the intra-class variance



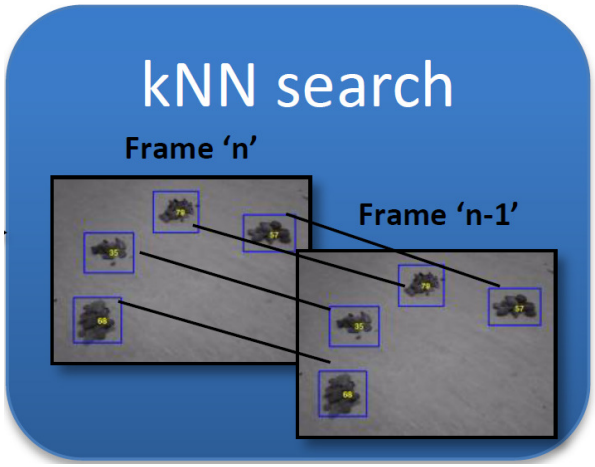


[3] N. Otsu. *"A threshold selection method from gray-level histograms,"* Systems, Man and Cybernetics, IEEE Transactions on, vol. 9, no. 1, pp. 62-66, 1979.

# k-NN Search-based Tracking

- k-NN search algorithm to track binary maps without the requirement of explicity model training or a priori knowledge of the dataset
- Collate ROI blobs throughout subsequent frames by applying Euclidean norm ($l^2$ norm)

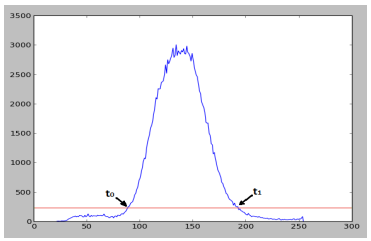
$$\forall I \forall r,\ kNN(R_I^t) \rightarrow \mathcal{L} : \mathcal{L} = \underset{r}{\operatorname{argmin}} \| R_I^t - Q_r^{t-1} \| \qquad (1)$$



# Binarised Image Segmentation via Thresholding

- Image segmentation using the MAT algorithm [4]
- Utilises local image statistics of *'mean'* and *'variance'* within a cluster and two thresholds obtained from intensity distribution histogram
- The algorithm uses a simple percentile (of the brightness) measurement procedure

$$dst_{(r,c)} = \begin{cases} 0 \ if \ \ dst_{(r,c)} > t_0 \\ 0 \ if \ \ dst_{(r,c)} < t_1 \\ 1 \ otherwise. \end{cases} \quad (2)$$



[4] F. Yan *et al.* *"A multistage adaptive thresholding method,"* Pattern Recognition Letters, 26, pp. 1183-1191, 2005.

# Continued...

- Binary blobs representing ROIs are selected based on zeroth moment (i.e., area) to eliminate outliers using an *a priori* defined threhold
- A border following method (based on [5]) is performed to extract blob contours
- Hu set of (translation and rotation) invariant moments [6] are computed for these contours
- This yields a vector of 7 values describing each blob (rock)

[5] S. Suzuki. *"Topological Structural Analysis of Digitized Binary Images by Border Following,"* Computer Vision, Graphics and Image Processing, 32-46.

[6] H. K. Hu. *"Visual Pattern Recognition by Moment Invariants,"* PROC. IRE vol. 49, p. 1428.

# Hu Moments Matching to Achieve Tracking

- Exhaustive search strategy is applied in order to carry out a comparison of Hu moments between two subsequent frames
- Results in matched pairs that are uniquely labelled throughout all the frames to achieve tracking
- More formally,

$$I(A,B) = \sum_{i=1}^{7} \left| \frac{1}{F_j^A} - \frac{1}{F_j^B} \right| \tag{3}$$

Where, $F_j^A$ and $F_j^B$ are defined (for Hu moment $h_j^A$ and $h_j^B$ of objects A & B) as follows,

$$F_j^A = sign(h_j^A) \cdot \log |h_j^A| \tag{4}$$

$$F_j^B = sign(h_j^B) \cdot \log |h_j^B| \tag{5}$$

# Normalised Multiple Object Detection Accuracy

For any given frame $t$, the number of '*false positives*' ($fp_t$), '*misses*' ($ms_t$) and '*true positives*' ($tp_t$) is calculated by measuring the spatial overlap between the ground-truth objects and the detector/tracker outputs.
Given, $G_i^t$ is the $i^{th}$ ground-truth object and $D_i^t$ is the $i^{th}$ detected object then the spatial overlap ratio ($OR_i^t$) is calculated as,

$$OR_i^t = \frac{|G_i^t \cap D_i^t|}{|G_i^t \cup D_i^t|} \tag{6}$$

where,

$$\forall i \forall t, D = \begin{cases} D_i^t \rightarrow \textit{true positive} : OR_i^t \geq 0.2 \\ D_i^t \rightarrow \textit{false positive} : OR_i^t < 0.2 \\ D_i^t \rightarrow \textit{miss} : \textit{unmatched.} \end{cases} \tag{7}$$

We calculate the Normalised Multiple Object Detection Accuracy (N-MODA) for the entire image sequence of each test data as follows,

$$N\text{-}MODA = 1 - \frac{\sum_{t=1}^{N_{frames}} \left(c_{ms}(ms_t) + c_f(fp_t)\right)}{\sum_{t=1}^{N_{frames}} N^t} \tag{8}$$

where,

$$\forall t, N^t = \begin{cases} N_G^t & \text{if } N_G^t \geq N_D^t \\ N_G^t & \text{if } N_G^t < N_D^t \end{cases}$$

for $\sum_{t=1}^{N_{frames}} N^t = 0$ we force $N\text{-}MODA = 0$. The parameters, $c_{ms}$ and $c_f$ are weighting parameters, set to $c_{ms} = c_f = 1$). $N_G^t$ is the number of ground-truth objects and $N_D^t$ is the number of detected objects.

# Multiple Object Tracking Accuracy

In order to evaluate the performance of the tracking system, we use the Multiple Object Tracking Accuracy (MOTA) as follows,

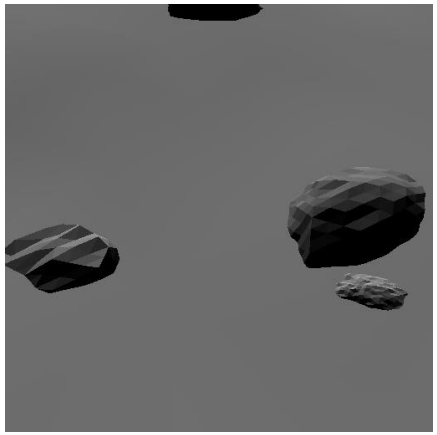
$$MOTA = 1 - \frac{\sum_{t=1}^{N_{frames}} \left(c_{ms}(ms_t) + c_f(fp_t) + c_s(ID_{SW_t})\right)}{\sum_{t=1}^{N_{frames}} N^t} \tag{9}$$

where, $ID_{SW_t}$ is the number of object label mismatches in the current frame '$t$' relative to the previous frame '$t-1$', and $c_s = \log_{10}$ (count for the number of mismatches always start from 1).

We compute these evaluation metrics along with important ROC measures (i.e., true positives per image ($Tp/img$), false positives per image ($Fp/img$), false negatives per image ($Fn/img$), *miss rate* and true positive rate ($TPR$)).

# Test Datasets

**Lab-based dataset**



**PANGU (simulated) dataset**

**RAL Space (SEEKER) dataset**

# Evaluation Results

Table: Visual Saliency-based Rock Detection & Tracking

| | Dataset | | | |
|---|---|---|---|---|
| **Measure** | **Lab-based** | **PANGU** | **SEEKER** | **Mean** |
| **N-MODA** | 0.80 | 0.84 | 0.61 | 0.75 |
| **MOTA** | 0.80 | 0.83 | 0.61 | 0.75 |
| **Tp/img** | 3.83 | 2.37 | 2.19 | 2.80 |
| **Fp/img** | 0.22 | 0.05 | 0.26 | 0.18 |
| **Fn/img** | 0.74 | 0.41 | 1.11 | 0.75 |
| **Objs/img** | 4.79 | 2.83 | 3.56 | 3.73 |
| **Miss rate** | 0.15 | 0.11 | 0.29 | 0.18 |
| **TPR** | 0.85 | 0.88 | 0.71 | 0.81 |

# Continued...

Table: Blob (Shape-based) Detection and Tracking

| | Dataset | | | |
|---|---|---|---|---|
| **Measure** | **Lab-based** | **PANGU** | **SEEKER** | **Mean** |
| **N-MODA** | 0.75 | 0.63 | 0.65 | 0.68 |
| **MOTA** | 0.71 | 0.61 | 0.62 | 0.65 |
| **Tp/img** | 4.36 | 2.14 | 3.11 | 3.20 |
| **Fp/img** | 1.36 | 1.07 | 1.54 | 1.32 |
| **Fn/img** | 0.08 | 0.18 | 0.13 | 0.13 |
| **Objs/img** | 5.80 | 3.39 | 4.78 | 4.66 |
| **Miss rate** | 0.01 | 0.07 | 0.03 | 0.04 |
| **TPR** | 0.99 | 0.93 | 0.97 | 0.96 |

# Conclusion

- Proposed to use two distinct approaches towards object detection and tracking using semantic feature descriptors
- The use of sparse semantic features reduces computational load, enables the use of heuristic tracking techniques
- Performed quantitative evaluations to check for performance using lab-based, simulated and real-world datasets
- We believe such techniques could potentially form a very effective basis for object detection/tracking for application in future long-distance autonomous rover navigation
- We anticipate to experiment with more challenging noisy datasets to achieve a solid foundation for the proposed concept

# END!

**THANKYOU!!**