Anomaly detection and target prioritization in planetary imagery via the automated global feature analyzer (AGFA): Progress towards a driver for autonomous C4ISR missions

Wolfgang Fink, Alexander J.-W. Brooks, Mark A. Tarbell

Event: SPIE Defense + Security, 2018, Orlando, Florida, United States
Invited Paper

Anomaly Detection and Target Prioritization in Planetary Imagery via the Automated Global Feature Analyzer™ (AGFA™): Progress Towards a Driver for Autonomous C⁴ISR Missions

Wolfgang Fink*, Alexander J.-W. Brooks†, Mark A. Tarbella†

†Visual and Autonomous Exploration Systems Research Laboratory, College of Engineering, University of Arizona, 1230 E Speedway Blvd, Tucson, AZ 85721, USA

ABSTRACT

The Automated Global Feature Analyzer™ (AGFA™) is a generically applicable automated sensor-data-fusion, feature extraction, feature vector clustering, anomaly detection, and target prioritization framework. AGFA™ operates in the respective feature space delivered by the sensor(s). In this paper we provide an overview of the inner workings of AGFA™ and apply AGFA™ to planetary imagery, representative of past, current, and future planetary missions, to demonstrate its automated and objective (i.e., unbiased) anomaly detection and target prioritization (i.e., region-of-interest delineation) capabilities. Imaged operational areas are locally processed via a cascade of image segmentation, visual and geometric feature extraction, agglomerative clustering, and principal components analysis. Resulting clusters are labeled based on relative size and location in feature space. Anomalous regions may be considered immediate targets for follow-up in-situ investigation by local robotic agents, which can be directed via autonomous telecommanding, e.g., as part of a Tier-Scalable Reconnaissance mission architecture. These capabilities will be essential for driving fully autonomous C⁴ISR missions of the future, since the speed of light prohibits “real time” Earth-controlled conduct of planetary exploration beyond the Moon.

Keywords: Autonomous decision making, sensor-data-fusion framework, objective anomaly detection, target prioritization, agglomerative clustering, principal components analysis, autonomous C⁴ISR systems, multi-tiered robotic exploration architectures

1. INTRODUCTION AND MOTIVATION

In (planetary) geology, biology, and medical sciences the reasoning process begins with observation. From observations and extracted features, these disciplines move on to the formulation of working hypotheses of overarching processes that may have lead to the observable evidence. These working hypotheses trigger the type and direction of follow-on investigations, observations, and experiments, the outcome of which may or may not lead to corroboration. This type of inference follows an abductive-inductive reasoning framework, in contrast to the deductive-inductive approach typical of physics and mathematics.

“Geology has discovered many unique and amazing phenomena. Usually the processes causing or having produced these phenomena are not observed, nor were they determined from fundamental laws. Most advancements of understanding the natural world have come from deduction. This deduction, especially in (planetary) geology, is never made from a single piece of evidence, but rather a suite of different and unique evidence.” [paraphrased statement, credit: Prof. Dr. James Dohm] As such, in particular planetary exploration has to be conducted by a multi-modal, multi-sensor, sensor-data-fusion, and unbiased anomaly detection based approach, especially when executed by robotic spacecraft operating in unknown environments. One of the main reasons for this is: unlike automated target recognition (ATR) for Earth-bound applications, where the target is usually known at least to some extent, the challenge of robotic planetary exploration lies in “expecting the unexpected!”

As NASA and other space agencies venture out to explore planetary bodies of high interest (Fig. 1), especially from an astrobiological point of view, i.e., the quest for extant/extinct life beyond Earth, planetary field geologists will have to be replaced/emulated, at least initially, by robotic spacecraft. As such, these robotic explorers will have to be equipped with observation, analysis, and reasoning capabilities of a field geologist.

* e-mail: wfink@email.arizona.edu; phone: +1-520-621-8734; website: http://autonomy.arizona.edu
Operational Autonomy Requirements for C4ISR Missions of the Future:

- Automatic mapping of an operational area from different vantage points
- Automatic feature extraction and target-of-interest identification within mapped operational area
- Automatic target prioritization for follow-up or close-up examination
- Automatic and unbiased anomaly detection
- Automatic science data interpretation
- Automatic (targeted) deployment and (concurrent) navigation/relocation of agents

Figure 1. Example planetary bodies of high interest for past, current, and future robotic space exploration missions (left to right, top to bottom): Titan, Venus, Moon, asteroid Itokawa, Mars, Earth, and Enceladus [Image credit: NASA, JAXA].

Figure 2. Operational requirements (left) for autonomous C4ISR missions of the future, such as Tier-Scalable Reconnaissance (TSR)1-8 mission architectures, e.g., for the comprehensive exploration of Mars (right).
Moreover, to mimic the geologic approach of local to regional to global reconnaissance in an integrated, mutually informing fashion, these robotic explorers will have to operate as part of multi-tiered robotic mission architectures, such as Tier-Scalable Reconnaissance (TSR),\textsuperscript{8-9} a foundational mission paradigm for autonomous C\textsuperscript{4}ISR systems of the future (Fig. 2, right). Devised for autonomous robotic space exploration of planetary bodies, the tier-scalable reconnaissance paradigm integrates multi-tier (space, atmosphere, surface, subsurface) and multi-agent hierarchical mission architectures (Fig. 2, right), introducing not only mission redundancy and safety, i.e., robustness, but also enabling intelligent, (science) objective-driven, and distributed reconnaissance of planetary bodies such as Mars, Titan, Venus, Europa etc. in near real-time.

As reported before\textsuperscript{8}, such systems and missions require operational autonomy (Fig. 2, left), comprising at least the following fundamental elements: (1) automatic characterization of operational areas from different vantages (i.e., space, air, ground, subsurface), (2) automatic sensor deployment and data gathering, (3) automatic feature extraction, anomaly detection, and region-of-interest or target identification, (4) automatic region-of-interest or target prioritization, and (5) subsequent automatic redeployment and navigation of robotic agents to regions or targets of interest.

2. AUTOMATED GLOBAL FEATURE ANALYZER\textsuperscript{TM} (AGFA\textsuperscript{TM}): OPERATIONAL STEPS

To address the above stated requirements for operational autonomy (in particular the ones inside the red box, Fig. 2, left) in a non-specific, generally applicable way, the Visual and Autonomous Exploration Systems Research Laboratory has been developing the Caltech-patented Automated Global Feature Analyzer\textsuperscript{TM} (AGFA\textsuperscript{TM}),\textsuperscript{8-13} which performs automated target identification using superpixel-based image segmentation,\textsuperscript{14} feature extraction, feature vector clustering, unbiased anomaly detection (i.e., based on feature space only), and target prioritization\textsuperscript{15} within mapped or imaged areas. AGFA\textsuperscript{TM} enables automated science analysis aboard robotic spacecraft, and, embedded in tier-scalable reconnaissance mission architectures, is a primary driver of future intelligent, distributed, and autonomous robotic exploration.

AGFA\textsuperscript{TM} objectively identifies and characterizes operational areas, both from an aerial/space perspective (i.e., overhead reconnaissance) and from the surface level. Moreover, it objectively flags anomalous targets via image analysis, i.e., based on the sensor-provided feature space alone as opposed to user bias. The software’s target identification and prioritization capabilities set the groundwork for autonomous surface investigation via robotic behavior modules/motifs in addition to robust path planning and adequate land/sea rover distribution in the presence of multiple targets and anomalies.\textsuperscript{8,16}

2.1 AGFA\textsuperscript{TM} Anomaly Detection Pipeline

The AGFA\textsuperscript{TM} procedure for anomaly detection in images can be decomposed into several functional blocks. Figure 3a gives an abstract, top-level overview of the major functional blocks of AGFA\textsuperscript{TM}. Figure 3b provides a more detailed diagram of how these functional blocks are subdivided and how they are engaged (i.e., AGFA\textsuperscript{TM} process flow). Each of these functional blocks is further described in the following subsections.

![Figure 3a. Abstraction of AGFA\textsuperscript{TM} anomaly detection pipeline.](image-url)
2.2 Image Segmentation

An image is partitioned into superpixels using the SLIC algorithm, which consists of localized clustering in a 5D space composed of X- and Y-coordinates and three LAB color space values\(^\text{14}\). The resulting superpixels are iteratively fused based on the correlation of the 2D histogram describing their respective A and B components within the LAB color space in order to create perceptually uniform regions (Fig. 4).

![Figure 4. Left: Raw image to be processed for anomaly detection. Center: Initial SLIC superpixel segmentation. Right: Resulting regions after histogram fusion.](image-url)
2.3 Contour Shape Fitting

Geometric descriptors are computed for each region extracted from the image segmentation algorithm. These descriptors consist of the approximate enclosing ellipse\(^{17}\) the best-fit ellipse\(^{18}\) and the minimum-area bounding rectangle\(^ {19}\) (Fig. 5).

![Figure 5. Left: Approximate enclosing ellipses for fused superpixel regions. Center: Approximate best-fit ellipses for fused superpixel regions. Right: Approximate minimum area bounding boxes of fused superpixel regions.](image)

2.4 Feature Calculations

Feature values are calculated for each of the segmented regions. These features can be broken down into visual features, i.e., features that can be calculated directly on the image or as a convolutional response, and geometric features, i.e., features that are calculated using the region shape and associated geometric descriptors. The following features are currently supported:

**Visual Features:**

- **Color** – The average RGB values of the region (Fig. 6, left).
- **Albedo** – The average reflectance values of the region (Fig. 6, right).
- **Texture** – List of the ratios of high magnitude pixels against the number of activated pixels in the respective Gabor filter\(^{20-23}\) (Fig. 7) response (note: the Gabor filters are taken over several user-selected orientations).

![Figure 6. Left: Average color of fused superpixel regions. Right: Average albedo for fused superpixel regions.](image)

![Figure 7. Gabor filters of different orientations used to determine the texture of identified regions.](image)
**Geometric Features:**

- **Moments** – The first 10 normalized moments and/or the first 7 Hu moments describing the geometric shape of the region contours.
- **Angularity** – The ratio of the approximate perimeter of the best-fit ellipse to the actual region boundary.
- **Compactness** – The ratio of the actual area of the region to the area of its minimum area-enclosing ellipse.
- **Extent** – The ratio of the actual area of the region to the area of its minimum area-bounding rectangle.
- **Circularity** – The ratio of the semi-minor and semi-major axes of the region’s best-fit ellipse.
- **Eccentricity** – A measure of how close to circular the region’s best-fit ellipse is.
- **Size** – The approximate region size given a viewing angle and a distance above the surface, e.g., as mapped from aboard an airship as part of a TSR architecture above an operational area.

Upon completion, the array of enabled feature vectors undergoes a patented normalization cascade\textsuperscript{12,13} to unify the respective ranges of individual feature values, account for the number of dimensions associated with each enabled feature, and ensure that the feature vectors in the array are comparable to each other.

### 2.5 Feature Vector Clustering

The normalized feature vectors are processed using the agglomerative clustering algorithm\textsuperscript{24} described in Figure 8. The current scheme uses the Euclidean distance between their closest entries of any two clusters as the distance between them (i.e., single link agglomerative clustering). Once all regions have been merged into a single cluster, the result may be represented as a tree-like structure known as a dendrogram (Fig. 9). Hereby a lower dendrogram height means that regions were combined earlier and vice versa. Particularly interesting are large height jumps and/or the combination of very imbalanced sub-trees: Figure 10 shows agglomerative clustering of an example image (Fig. 4, left). It should be pointed out that the cluster linking/merging is dependent on the choice of distance measure, i.e., linkage criterion.\textsuperscript{24}

![Figure 8. Procedure for agglomerative clustering.](https://www.spiedigitallibrary.org/conference-proceedings-of-spie)
Figure 9. **Left:** 2D feature points; clusters are created in the order: green, orange, blue, purple, yellow, gray. **Right:** Corresponding dendrogram created using single link agglomerative clustering.

Figure 10. **Left:** Initial cluster assignments before single link agglomerative clustering of example image (Fig. 4, left). **Right:** Cluster assignments resulting from truncating the dendrogram as shown in Figure 11.

Figure 11. Resulting dendrogram for example image (Fig. 4, left), calculated by AGFA™ for the entire enabled feature space. The truncation height is marked with a red line.
The resulting dendrogram may be truncated at a specific user-defined height (i.e., termination criterion) to produce a respective number of clusters. Figure 11 shows the resulting dendrogram for the clustered example image (Fig. 4, left), as well as the resulting cluster assignment when truncating the dendrogram at a specific height (Fig. 12, bottom right).

A dendrogram is also generated for each enabled feature, respectively. These feature dendrograms are not truncated, but are instead considered as independent descriptors for each feature space.

2.6 Flag Computations to indicate relative and absolute anomalies

The following flags are computed to assess the extent to which clusters, resulting from truncating the dendrogram created from the normalized feature vectors, are anomalous relative to each other (Fig. 13):

- **Distance Flag**\textsuperscript{12,13} – Two clusters are anomalous relative to each other if the Euclidean distance between the centroids in feature space of their respective clusters exceeds the sum of the magnitudes of their principal components obtained via principal components analysis\textsuperscript{25} (upper triangle in Figure 12, top right).
• **Number Flag**
  - Two clusters are anomalous relative to each other if one cluster contains less than $\alpha$ times the number of elements compared to the other cluster (lower triangle in Figure 12, top right). Note: $\alpha$ is a user-defined parameter.

If either the distance flag or the number flag exhibits an anomaly, this defines a *relative anomaly*. If both flags show an anomaly, this defines an *absolute anomaly*.

The following additional flags are computed to assess the extent to which anomalies exist within the individual dendrograms associated with each respective enabled feature/feature space.

• **Feature Flag**
  - An anomaly in an individual feature space is defined by the mean of merging heights in its dendrogram being smaller than the standard deviation. To illustrate this: if the merging heights were almost equidistant, the mean would be about the merging height and the standard deviation would be close to zero, hence no anomaly present (similar to Fig. 9, right). If there is however a varied distribution of merging heights (Fig. 11), then the standard deviation would likely be close to the mean merging height, if not larger, hence indicating an anomaly in that feature space. Figure 12, top left shows the individual feature flags for the example image in Figure 4, left.

![Distance Flag Diagram](https://www.spiedigitallibrary.org/conference-proceedings-of-spie)

**Distance Flag:**
- IF $(|C_1 - C_2| > |PCA_{1\text{max}}| + |PCA_{2\text{max}}|)$
  THEN flag as anomaly
- ELSE flag as normal

**Population Flag:**
- with $0 < \alpha < 1$ (e.g., $\sim 10\%$) and $|x| = \text{population of cluster } x$
- IF $(|C_1| < \alpha * |C_2| \text{ or vice-versa})$
  THEN flag as anomaly
- ELSE flag as normal

Figure 13. Illustration of the anomaly detection criteria\textsuperscript{12,13} (from [9]) pertaining to the clustered feature vectors resulting from the principal components analysis.\textsuperscript{25}

### 3. APPLICATION EXAMPLES OF AGFA\textsuperscript{TM}

In the following we demonstrate the capabilities of AGFA\textsuperscript{TM} when applied to a set of planetary images taken from various vantage points (representing space, air, ground), i.e., at different scales. These images may be representative of past, current, and future planetary missions (especially Tier-Scalable Reconnaissance missions), to demonstrate the automated and objective (i.e., unbiased) anomaly detection and target prioritization (i.e., region-of-interest delineation) capabilities of AGFA\textsuperscript{TM}.
3.1 Global Vantage/Scale (e.g., from an orbiter perspective)

Figure 14. Top Left: Valles Marineris, imaged by Viking Orbiters 1 & 2, processed with AGFA\textsuperscript{TM} [Image credit: NASA/JPL/USGS]. Top Center: Initial superpixel segmentation. Top Right: Resulting regions to be analyzed after superpixel fusion. Bottom Left: Features used in anomaly detection. Bottom Center: Agglomerative clustering initialization (singleton clusters). Bottom Right: Projection of cluster assignments resulting from truncating the dendrogram as shown in Figure 15.

Figure 15. Resulting dendrogram for Valles Marineris, calculated by AGFA\textsuperscript{TM} for the entire enabled feature space. The truncation height is marked with a red line.
Individual Feature Flags

<table>
<thead>
<tr>
<th>Anomalous Features</th>
<th>Non-Anomalous Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>Size</td>
<td>Color</td>
</tr>
<tr>
<td>Compactness</td>
<td></td>
</tr>
<tr>
<td>Angularity</td>
<td></td>
</tr>
</tbody>
</table>

Cluster Color

<table>
<thead>
<tr>
<th>Blue</th>
<th>Green</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blue</td>
<td></td>
</tr>
<tr>
<td>Green</td>
<td></td>
</tr>
</tbody>
</table>

Figure 16. Top Left: Breakdown of anomalous features resulting from examining single feature dendrograms. Top Right: Distance flag (upper triangle) and number flag (lower triangle). Red boxes indicate anomalies. Bottom Left: Raw image processed directly by AGFA™. Bottom Right: Projection of cluster assignments resulting from truncating the dendrogram as shown in Figure 15.

Discussion

Figure 14 illustrates the key steps of the AGFA™ detection of Valles Marineris as captured by Viking Orbiters 1 & 2. Of particular note is the order of clustering assigned to the fused superpixel regions using color, size, angularity, and compactness as the features of interest. The final clustering shown illustrates that the segmented crater region is the final group (i.e., most unique in the normalized feature space) assigned during the construction of the complete dendrogram hierarchy.

Post-processing on the two clusters obtained from truncating the dendrogram (Fig. 15) indicates that the distance and number flags (Fig. 16, top right) between them are both active (i.e., red), thus indicating the presence of an absolute anomaly. This example is critical in illustrating the potential of clustering in the proposed unified feature space, which may contain both visual and geometric region features. Furthermore, although an automated choice of threshold may have selected a lower truncation height for the dendrogram, thereby creating several additional clusters, it shows that regions that are highly unique will typically remain their own cluster. Furthermore, the individual flags for each of the features used (Fig. 16, top left) reveal the non-parametric detection of relative anomalies in extent and albedo.
3.2 Regional Vantage/Scale (e.g., from a blimp perspective)

- Color
- Albedo
- Extent

Figure 17. Top Left: Mars crater with water ice deposit, imaged by Mars Express, processed with AGFA\textsuperscript{TM} [Image credit: ESA/DLR/FU Berlin (G. Neukum)]. Top Center: Initial superpixel segmentation. Top Right: Resulting regions to be analyzed after superpixel fusion. Bottom Left: Features used in anomaly detection. Bottom Center: Agglomerative clustering initialization (singleton clusters). Bottom Right: Projection of cluster assignments resulting from truncating the dendrogram as shown in Figure 18.

Figure 18. Resulting dendrogram for Mars crater, calculated by AGFA\textsuperscript{TM} for the entire enabled feature space. The truncation height is marked with a red line.
Discussion

Figure 17 illustrates the key steps of the AGFA™ detection of a deposit of water ice as captured by Mars Express. An important point to consider in this example is that although the choice in height when truncating the dendrogram (Fig. 18) creates five clusters (based on region color, albedo, and extent), highly similar fused superpixel groups are combined effectively, thus illustrating the system’s potential in effectively fusing over-segmented superpixels in addition to grouping disjoint, non-adjacent objects/regions with similar feature descriptors.

Post-processing the five clusters reveals that green, brown, pink, and yellow regions all have active (i.e., red) number flags relative to the red cluster (i.e., background) (Fig. 19, top right). Principal components analysis reveals that each of these clusters is anomalous relative to each other except for the yellow and red clusters. Furthermore, the individual flags for each of the features used (Fig. 19, top left) reveal the non-parametric detection of relative anomalies in extent and albedo.

Figure 19. Top Left: Breakdown of anomalous features resulting from examining single feature dendrograms. Top Right: Distance flags (upper triangle) and number flags (lower triangle). Red boxes indicate anomalies, and green boxes indicate no anomalies. Bottom Left: Raw image processed directly by AGFA™. Bottom Right: Projection of cluster assignments resulting from truncating the dendrogram as shown in Figure 18.
3.3 Local Vantage/Scale (e.g., from a rover/lander perspective)

- Color
- Albedo

Figure 20. Top Left: Scoop site on Mars, imaged by Phoenix Mars Lander, processed with AGFA\textsuperscript{TM} [Image credit: NASA, JPL-Caltech, UoFA, Texas A&M]. Top Center: Initial superpixel segmentation. Top Right: Resulting regions to be analyzed after superpixel fusion. Bottom Left: Features used in anomaly detection. Bottom Center: Agglomerative clustering initialization (singleton clusters). Bottom Right: Projection of cluster assignments resulting from truncating the dendrogram as shown in Figure 21.

Figure 21. Resulting dendrogram for Mars crater, calculated by AGFA\textsuperscript{TM} for the entire enabled feature space. The truncation height is marked with a red line.
Discussion

Figure 20 illustrates the key steps of the AGFA™ extraction of white substance, including water ice, as captured by the Phoenix Mars Lander, using color and albedo as the features for clustering. The choice of the truncation height for the dendrogram (Fig. 21) creates six different clusters. Post-processing reveals that the number flag of each of these clusters is anomalous (i.e., red) relative to the red cluster (background) and that the distance flags for all resulting clusters are anomalous (i.e., red) with respect to each other (Fig. 22, top right). Furthermore, the individual flags for each of the features used (Fig. 22, top left) reveal the non-parametric detection of relative anomalies in both color and albedo.

This example also illustrates a potential pitfall of using single-linkage agglomerative clustering. The next merging step just above the cutting point in the dendrogram (Fig. 21) fuses the ice into the background. This is a consequence of the previous clustering of patches containing pieces of ice and background terrain, thereby exhibiting characteristics of both. As a potential mitigation of this issue, the choice of a different linkage criterion24 may be useful for preventing mergers like this from occurring too early in the clustering process.

4. CONCLUSIONS – NEXT STEPS – OUTLOOK

AGFA™ is a first step towards unbiased anomaly detection in unknown environments, such as, but not limited to, planetary surfaces (e.g., Mars, Enceladus, Europa, Titan, etc.) and subsurfaces (e.g., subsurface ocean on Europa), and, of course, Earth. AGFA™ performs anomaly detection solely based on the feature spaces delivered by multi-modal, multi-sensor equipped (robotic) space missions of the future. As such, this is not the typical automated target recognition (ATR) scenario, but rather the “expect the unexpected” type mission setting.
AGFA™ is readily extensible, for example:

- To include additional feature spaces, such as multi- or hyperspectral data;
- To incorporate advances in image processing algorithms for the initial image segmentation/object recognition phase at the front end of AGFA™;
- To incorporate advances in image analysis algorithms for enhanced feature extraction into AGFA™;
- To incorporate advances in both supervised and unsupervised clustering techniques into AGFA™;
- To involve additional machine learning and deep learning algorithms at the back end of AGFA™.

The development of AGFA™ will continue with particular emphasis on strengthening the initial image segmentation step. Moreover, a target prioritization framework, detailed in [15], will be integrated with AGFA™. Ultimately, AGFA™ will be integrated into a Tier-Scalable Reconnaissance robotic testbed8,16 and deployed outdoors in geologically relevant environments to test and validate its performance, with the aspiration of becoming part of future autonomous robotic space exploration missions.

5. ACKNOWLEDGEMENTS

The work described in this publication was carried out in part at the California Institute of Technology (2003 – 2016) and at the University of Arizona (2009 – present) with partial support from the Edward & Maria Keonjian Endowment at the University of Arizona. Author AJ-WB has been supported by NASA traineeship grant NNX15AJ17H via Arizona Space Grant Consortium (AZSGC).

6. AUTHOR DISCLOSURE STATEMENT

Author WF may have a financial interest in the Automated Global Feature Analyzer™ (AGFA™) as he is named as an inventor on several issued and pending Caltech patents. Authors AJ-WB and MAT have no financial interest in AGFA™. Authors WF and MAT may have a financial interest in tier-scalable reconnaissance mission architectures as they are named as inventors on several issued Caltech patents. Author AJ-WB has no financial interest in TSR.

REFERENCES

[5] US 6,990,406: “MULTI-AGENT AUTONOMOUS SYSTEM” (Fink, Dohm, Tarbell; Caltech)
[6] US 7,734,063: “MULTI-AGENT AUTONOMOUS SYSTEM” (Fink, Dohm, Tarbell; Caltech)
[7] US 7,742,845: “MULTI-AGENT AUTONOMOUS SYSTEM AND METHOD” (Fink, Dohm, Tarbell; Caltech)


[12] US 9,122,956 (and PCT): “Automated feature analysis, comparison, and anomaly detection” (Fink; Caltech)

[13] US 9,424,489: “Automated feature analysis, comparison, and anomaly detection” (Fink; Caltech)


