

Using Clustering and Metric Learning to Improve Science Return of Remote Sensed Imagery

DAVID S. HAYDEN, STEVE CHIEN, DAVID R. THOMPSON, and REBECCA CASTAÑO,
Jet Propulsion Laboratory, California Institute of Technology

Current and proposed remote space missions, such as the proposed aerial exploration of Titan by an aerobot, often can collect more data than can be communicated back to Earth. Autonomous selective downlink algorithms can choose informative subsets of data to improve the science value of these bandwidth-limited transmissions. This requires statistical descriptors of the data that reflect very abstract and subtle distinctions in science content. We propose a metric learning strategy that teaches algorithms how best to cluster new data based on training examples supplied by domain scientists. We demonstrate that clustering informed by metric learning produces results that more closely match multiple scientists' labelings of aerial data than do clusterings based on random or periodic sampling. A new metric-learning strategy accommodates training sets produced by multiple scientists with different and potentially inconsistent mission objectives. Our methods are fit for current spacecraft processors (e.g., RAD750) and would further benefit from more advanced spacecraft processor architectures, such as OPERA.

Categories and Subject Descriptors: I.2.9 [Artificial Intelligence]: Robotics—*Autonomous vehicles*

General Terms: Algorithms, Experimentation, Human Factors

Additional Key Words and Phrases: Onboard data analysis, selective data return, information retrieval, space exploration, clustering

ACM Reference Format:

Hayden, D. S., Chien, S., Thompson, D. R., and Castaño, R. 2012. Using clustering and metric learning to improve science return of remote sensed imagery. *ACM Trans. Intell. Syst. Technol.* 3, 3, Article 51 (May 2012), 19 pages.

DOI = 10.1145/2168752.2168765 <http://doi.acm.org/10.1145/2168752.2168765>

1. INTRODUCTION

In general, data from spacecraft can be collected at a much higher rate than can be transmitted to Earth. Bandwidth bottlenecks due to distance, power, visibility, and competing mission downlinks can significantly limit the lifetime science return of a mission. As a result, scientists must exercise great care in deciding which data to collect and transmit to Earth for analysis. For previously observed or static

Portions of this work were performed by the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration. Additional support came from the National Science Foundation through a Graduate Research Fellowship to D. S. Hayden.

Author's address: D. S. Hayden; email: dshayden@mit.edu.

©2012 Association for Computing Machinery. ACM acknowledges that this contribution was authored or co-authored by a contractor or affiliate of the [U.S.] Government. As such, the Government retains a nonexclusive, royalty-free right to publish or reproduce this article, or to allow others to do so, for Government purposes only.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies show this notice on the first page or initial screen of a display along with the full citation. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, to redistribute to lists, or to use any component of this work in other works requires prior specific permission and/or a fee. Permissions may be requested from the Publications Dept., ACM, Inc., 2 Penn Plaza, Suite 701, New York, NY 10121-0701, USA, fax +1 (212) 869-0481, or permissions@acm.org.

© 2012 ACM 2157-6904/2012/05-ART51 \$10.00

DOI 10.1145/2168752.2168765 <http://doi.acm.org/10.1145/2168752.2168765>

environments, such as most planetary surfaces imaged from orbit, data collection may be scripted so that specific targets are imaged at prearranged times. For unknown or dynamic environments, scripted data collection strategies may include sampling at random or periodic intervals. In both cases, relatively little is known about collected data until it has been downlinked. Returned images may exhibit substantial redundancy, be marred by artifacts, or may not be relevant to current science objectives [Thompson et al. 2008]. Additionally, scripted observations leave the collection of remarkable or unexpected data to chance and could miss key features.

These bandwidth constraints motivate onboard data analysis that could recognize key image content and prioritize the most informative observations for downlink. Previous work in onboard data understanding has been practiced on rover, satellite, and submersible platforms, and is well underway for UAVs [Clough 2002]. The Mars Exploration Rovers can autonomously search for science targets, including clouds and dust devils [Castano et al. 2008], as well as for rocks of specific size, albedo, and shape [Estlin et al. 2010]. For the Earth Observing 1 Satellite, hazardous events such as fires, floods, and volcanic activities are detected and pertinent data downlinked [Chien et al. 2005]. Terrestrial autonomy includes underwater exploration [Kinsey et al. 2006], mine and cave mapping [Thrun et al. 2005], and also exploration of Mars-like environments, such as the Atacama [Wettergreen et al. 2008] and Mojave [Calderon et al. 2008] deserts.

In many cases, it would be useful to complement target detection with *representative sampling* [Castaño et al. 2003; Gilmore et al. 2000]. Representative sampling characterizes the entire dataset by selecting archetypal exemplars from its principal classes, where these classes are not known or specified in advance. Representative sampling is formally equivalent to a vector quantization strategy for data compression [Gersho and Gray 1992]. In practice, one can treat it as a clustering problem in which a spacecraft groups images according to content and downlinks characteristic examples of each cluster [Castaño et al. 2007; Hayden et al. 2010]. Figure 1 demonstrates a graphical example.

Several factors complicate automatic clustering. First, effective performance requires describing the data with numerical features that can be correlated well with its science content. These features should be invariant to irrelevant differences from changing imaging conditions or noise. Finding such features and implementing them in flight software is a difficult process that may need to be repeated for changing mission objectives. Moreover, interfeature redundancy and correlation can adversely affect clustering performance. Previous work shows that clustering with different feature subspaces can dramatically alter selective data-return performance [Hayden et al. 2010].

We desire a general method for learning an optimal image representation based on training data. Specifically, we generate many features and then learn a distance metric that places similar images close together and dissimilar images far apart. As mission objectives evolve and additional training data becomes available, new metrics can be trained on Earth. Then, operators can transmit this information to the spacecraft in the form of relatively compact parameters, such as linear transformation matrices. Computing distances with the learned metric is equivalent to projecting the original feature space and subsequently taking Euclidean distances. The resulting clusters should better reflect the distinctions in science content most relevant to mission objectives.

Onboard remote science analysis presents an additional complication. Traditional metric learning and feature selection require an internally consistent and authoritative set of training data. But in remote exploration, such ground truth standards may not exist. It will almost certainly not be possible to find a single partitioning of the

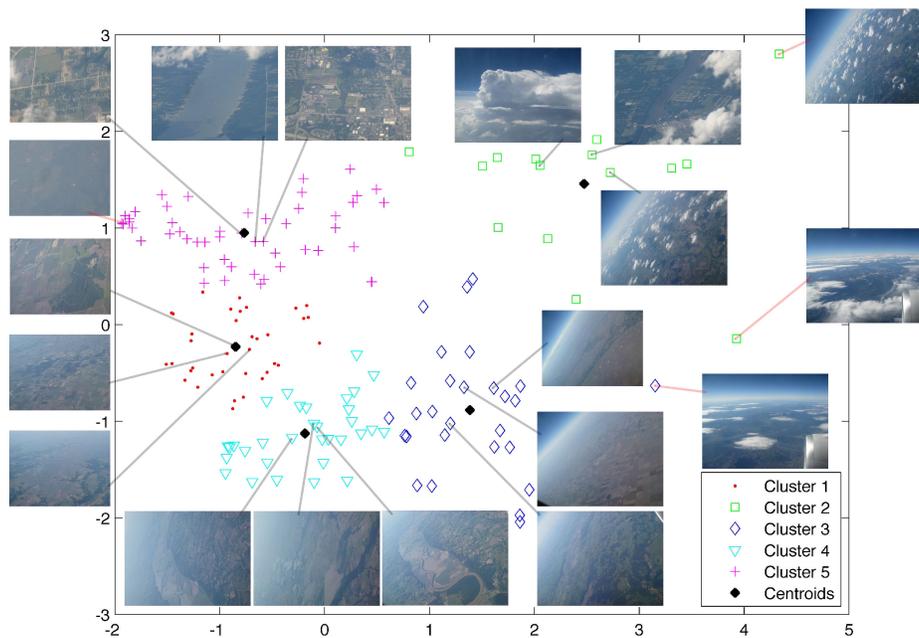


Fig. 1. Aerial data automatically clustered into five groups using the features in Section 3. Three representative images from each cluster are displayed, as are two outlier images (those furthest from any cluster centroid).

data that satisfies all mission science goals simultaneously. It will probably be the case that there are many scientists, each with their own separate science goals, key distinctions, and clusterings. Relaxing the consistency assumption lets the learning process fully utilize a training set from multiple scientists and balance any competing objectives that they have assumed in their training labels.

In this article, we explore efficient metric-learning approaches to selective data return. We demonstrate that clustering based on metric learning outperforms random or periodic collection strategies at producing partitions to match those made by scientists. We especially focus on the problem of satisfying multiple, competing mission objectives. In particular, we extend Linear Discriminant Analysis [Duda et al. 2000] to handle multi-class data with several inconsistent labelings and show that it outperforms modern metric-learning approaches in satisfying multiple competing science objectives. Throughout, we address practical and logistical constraints of implementing a clustering framework for selective data return.

Central to our work is the assumption that the utility of an automatic clustering is determined by how closely it matches clusterings made manually by scientists. While intuitive, this is not formally justified, and in this article, it is only demonstrated on a small terrestrial dataset.

In Section 2, we discuss clustering and metric-learning algorithms in the context of selective data return. We also introduce Multi-domain Multiclass Linear Discriminant Analysis (MDMC-LDA), a new metric-learning algorithm for handling multiple competing science objectives. In Section 3, we discuss our terrestrial dataset, labeling strategy, image features, and clustering performance evaluation strategy. Sections 4 and 5 detail our findings.

2. APPROACH

2.1. Clustering

Clustering provides scientists with considerable control over how downlink bandwidth is used. If scientists desire a summary of observed data, they can downlink representatives from each cluster. Alternatively, they can reallocate bandwidth to favor particular clusters that prove consistently interesting. Clustering in a metric space affords additional flexibility by allowing data to be rank-ordered by representativeness. Data closest to the centroid can be considered more representative, while an image far from any centroid may be an interesting outlier. Geolocating cluster representatives and outliers into a terrain map can facilitate subsequent mission planning.

Of the many clustering techniques available [Xu and Wunsch 2005], we focus on k-means [MacQueen et al. 1967] because it is fast, simple, and well understood. For dataset $X = \{x_1, \dots, x_N\}$, $x_i \in R^d$, k-means randomly defines k centroids $\{c_1, \dots, c_k\}$, $c_i \in R^d$. It then iteratively reassigns each x to cluster C_i if $\|x - c_i\| \leq \|x - c_j\|$ for $j \neq i$ and recomputes each centroid as $c_i = \frac{1}{|C_i|} \sum_{x \in C_i} x$. K-means terminates to a local minimum of the sum-of-squares error of each x and its closest centroid c_i when no c_i change value and no x are assigned to a different cluster.

By perturbing each x to avoid degenerate cases, Arthur et al. [2009] have shown that k-means has polynomial runtime in n . In practice, the typical implementation [Lloyd 1982] often has linear or sublinear runtime [Duda et al. 2000]. If a suboptimal solution is sufficient, one can alter the algorithm to guarantee such performance. Improvements of several orders of magnitude have also been demonstrated: Elkan [2003] exploits the triangle inequality, while Saegusa and Maruyama [2007] demonstrate an FPGA implementation running 30 times per second with $d = 3$, $n = 512^2$, and $k = 256$.

K-means requires that the number of clusters be determined in advance. This can be seeded by initial tests on the ground then adjusted throughout the course of a mission to determine if other values seem more useful. K-means also requires that careful consideration be given to the image descriptors used to represent each image. In exploration scenarios, an ideal clustering partitions the images into groups that reflect current objectives (for example, each cluster could contain images dominated by a particular topography). Constructing an optimal feature space is a more subtle challenge that we will investigate further in Section 2.2.

2.2. Metric Learning

Metric learning can tune clustering behavior towards current or dynamic mission objectives without requiring modifications to flight software. Scientists would manually cluster data returned from the spacecraft. One can use these clusterings to construct a linear or nonlinear feature transformation that emphasizes their preferred distinctions. The scientists would transmit the parameters of this transformation back to the spacecraft¹, which would then use it to form a mission-relevant feature space for cluster analysis of future observations.

Following Xing et al. [2002], we formulate the metric-learning task as follows. Given two sets of constraints in the form of pairs of similar and dissimilar data, metric-learning methods seek a projection in which similar data are close together and dissimilar data are far apart. We describe proximity with the L^2 -norm: for d -dimensional vectors u and v , the distance between them is defined as $\|(u - v)^T I(u - v)\|_2$, where I is the $d \times d$ identity matrix. With metric learning, the identity matrix is replaced with a positive semidefinite or positive definite transformation matrix A that parameterizes

¹The size of the transformation would be the square of the feature space dimensionality in floats—less than 8MB for an unlikely 1,000 dimensions.

a family of Mahalanobis distances in R^d . A diagonal matrix A simply weights each dimension. Otherwise, the transformation is equivalent to a rescaling and rotation of the data. Computing the transformation for each distance computation is equivalent to projecting the feature space and subsequently measuring all distances with the standard Euclidean metric.

Yang [2007] provides an overview of metric-learning approaches. The foundational work of Xing et al. [2002] describes metric learning as a convex optimization problem which minimizes the sums of squared distances between similar data and pushes sums of squared distances between dissimilar data beyond a threshold. Other more recent approaches include Neighborhood Components Analysis [Goldberger et al. 2005] which minimizes the leave-one-out k-nearest-neighbor classification error. Largest-Margin Nearest Neighbor [Weinberger et al. 2006] also minimizes nearest-neighbor error but utilizes a largest margin framework that results in a convex optimization problem. Other metric-learning methods explore nonlinear transformations [Chen et al. 2007], active learning and Bayesian frameworks [Yang and Jin 2007], and online methods [Jain et al. 2009]. To establish a baseline, we restrict our attention to linear methods.

2.3. MDA

Classical multiclass linear discriminant analysis (LDA or MDA) projects a d -dimensional data $X = \{x_1, \dots, x_N\}$ with labels $1 \leq l_i \leq C: i = 1, \dots, N$ into a $(C - 1)$ -dimensional subspace that better separates each of the C classes. The Bayes classification error is optimally minimized when the classes are Gaussian and homoscedastic (having equal covariance matrices). More thorough coverage of MDA can be found in Friedman [1989] and Duda et al. [2000].

Our notation reflects the fact that the set of training data consists of just one self-consistent labeling. Let X_{1j} be the subset of X corresponding to the single labeling's j th class. Let n_{1j} be the number of vectors in the j th class. We define the within-class scatter matrix as

$$S_W = \frac{1}{C} \sum_{j=1}^C \text{cov}(X_{1j}), \quad (1)$$

where $\text{cov}(X_{1j})$ is the covariance matrix of the matrix whose columns are the vectors $x \in X_{1j}$. Let $m = \frac{1}{N} \sum_{x \in X} x$ be the total mean of X and

$$m_{1j} = \frac{1}{n_{1j}} \sum_{x \in X_{1j}} x \quad (2)$$

be the mean of the j th class. Then, we define the between-class scatter matrix as

$$S_B = \frac{1}{N} \sum_{j=1}^C n_{1j} (m_{1j} - m)^T (m_{1j} - m). \quad (3)$$

MDA simultaneously minimizes the projected within-class scatter and maximizes the between-class scatter. This is equivalent to finding a (non-unique) W that maximizes

$$\mathcal{J}(W) = \frac{|W^T S_B W|}{|W^T S_W W|}. \quad (4)$$

The columns of the optimal projection W are the top eigenvectors of $S_W^{-1} S_B$ [Duda et al. 2000]. Note that if dimensionality is high or there are few training samples,

then S_W may be singular. We avoid this condition with a simple regularization scheme where the columns of W are generated from $(S_W^{-1} + \alpha I)S_B$. Designers can set α to a small value or optimize it with cross validation.

Many variations on MDA exist. Duin and Loog [2004], Hamsici and Martinez [2007], and Qin et al. [2005] relax the homoscedastic requirement. Ye [2006], Chen et al. [2000], and Howland et al. [2004] do not require nonsingularity. Friedman [1989] and Guo et al. [2007] provide regularization strategies.

2.4. MDMC-LDA

Space exploration scenarios will often involve multiple teams of scientists with independent and possibly competing objectives. This may make MDA impractical since it requires a single fixed number of classes and a single set of labels. Finding a consensus labeling among all teams might require a challenging negotiation. At worst, the basic task may be ill-posed if the labelers consider different aspects of the data. A more reasonable strategy would be to prompt each team to label relevant training data according to its individual objectives, and then learn a projection which mediated between them. Our algorithm, Multi-Domain, Multiclass Linear Discriminant Analysis (MDMC-LDA) extends MDA to account for multiple, possibly inconsistent labels on multiple, possibly overlapping datasets.

We define a *domain* as a single independent labeling of the training data. The i th domain is characterized by a set of d -dimensional vectors with integer labels ranging from $[1, C_i]$. We take X_i to be the set of its vectors, with $n_i = |X_i|$. The labels of all domains are represented by $L = \{L_1, \dots, L_D\}$, where $L_i = \{l_{i1}, \dots, l_{in_i}\}$. Similarly, X_{ij} is the vectors in the j th class of the i th domain with $n_{ij} = |X_{ij}|$. Finally, $N = \sum_{i=1}^D n_i$ represents the total number of data points, each counted as many times as there are labels supplied for it (alternatively, we could simply say this is the total number of labels).

Our modification of MDA does not directly compare images across domains, since these semantic relationships are undefined. Two classes in different domains might represent the same semantic category, despite having different labels, in which case it would not be desirable to separate them. However, taken independently, the domains supply additional training data that can improve performance. Specifically, we treat each domain as an independent sample of the between-class and within-class sample matrices. The expectation of these matrices, taken over all domains, provides a semantically-consistent solution. We define the within-class scatter as

$$\begin{aligned} S_W &= E_D \left[\frac{1}{C} \sum_{j=1}^C \text{cov}(X_{ij}) \right] \\ &= \frac{1}{D} \sum_{i=1}^D \frac{1}{C_i} \sum_{j=1}^{C_i} \text{cov}(X_{ij}). \end{aligned} \quad (5)$$

Here $C_i = \max L_i$ and $\text{cov}(X_{ij})$ is the covariance matrix of the matrix whose columns are the vectors $x \in X_{ij}$. We define the mean of the i th domain and also the mean of the j th class in the i th domain as

$$m_i = \frac{1}{n_i} \sum_{x \in X_i} x, \quad (6)$$

$$m_{ij} = \frac{1}{n_{ij}} \sum_{x \in X_{ij}} x. \quad (7)$$

Similarly, we find a separate between-class scatter matrix for independent domains and average them to produce a final point estimate. This ignores proximity relationships across different domains.

$$\begin{aligned} S_B &= E_D \left[\frac{1}{D} \sum_{j=1}^{C_i} n_i (m_{ij} - m_i)^T (m_{ij} - m_i) \right] \\ &= \frac{1}{N} \sum_{i=1}^D \sum_{j=1}^{C_i} n_i (m_{ij} - m_i)^T (m_{ij} - m_i). \end{aligned} \quad (8)$$

As in Equation (4), we take the solution to be a W that maximizes the ratio of between-class and within-class scatter.

$$\mathcal{J}(W) = \frac{|W^T S_B W|}{|W^T S_W W|}. \quad (9)$$

We again generate the columns of W by the top eigenvectors of $(S_W^{-1} + \alpha I)S_B$, where the regularization parameter α is determined through cross validation. If $D = 1$, then MDMC-LDA degenerates to MDA. Notationally, the i subscripts in MDMC-LDA can be replaced with 1, and the solution W will be optimal if data are Gaussian and homoscedastic. For $D > 1$, we do not expect these conditions will ever be satisfied in the case of overlapping labels. Informally, though, overlapping labels should push data with corresponding labels into proximity so that clusters can more naturally form that better satisfy all objectives. We thus hypothesize that the projected space will improve overall agreement between multiple labelers, although agreement with any one might be lower than if MDA were trained on that particular labeling.

3. EXPERIMENTAL METHOD

To demonstrate our clustering approach to onboard data analysis, we use 19 image descriptors that can be collected in subpolynomial time and use these to cluster images from an aerial dataset. We investigate correspondence between these automatic clusterings and manual clusterings created by planetary geologists. We also consider clusterings that would result from nonadaptive or uninformed data return strategies, such as random sampling or periodic sampling at regular time intervals. We explore various alternative metric-learning approaches to improve performance, including the new MDMC-LDA algorithm.

3.1. Aerial Dataset

Most studies in onboard data understanding have focused on rover and satellite platforms. Here, we investigate aerial exploration scenarios, such as those proposed for Titan, by a blimp, or for Venus, by a fixed-wing craft [NASA 2006]. In the case of Titan, an autonomous blimp would likely travel for many kilometers—sometimes for more than a week—between downlink opportunities and collect a vast number of traverse images over previously unseen and diverse terrain [Elfes et al. 2008]. Selective data return could help discriminate between its morphological and atmospheric features, which are known to include hydrocarbon lakes, dried riverbeds, shorelines, mountainous and smooth desert-like terrain, sand dunes, clouds, and the occasional crater. Figure 2 shows some examples.

Rather than utilizing terrestrial or Martian satellite imagery, which may not correspond well with expected Titan imagery in color, resolution, or feature diversity, we simulated an aerial imaging sequence on Titan using 162 terrestrial scenes collected with a consumer camera at two megapixels during a commercial flight from New York

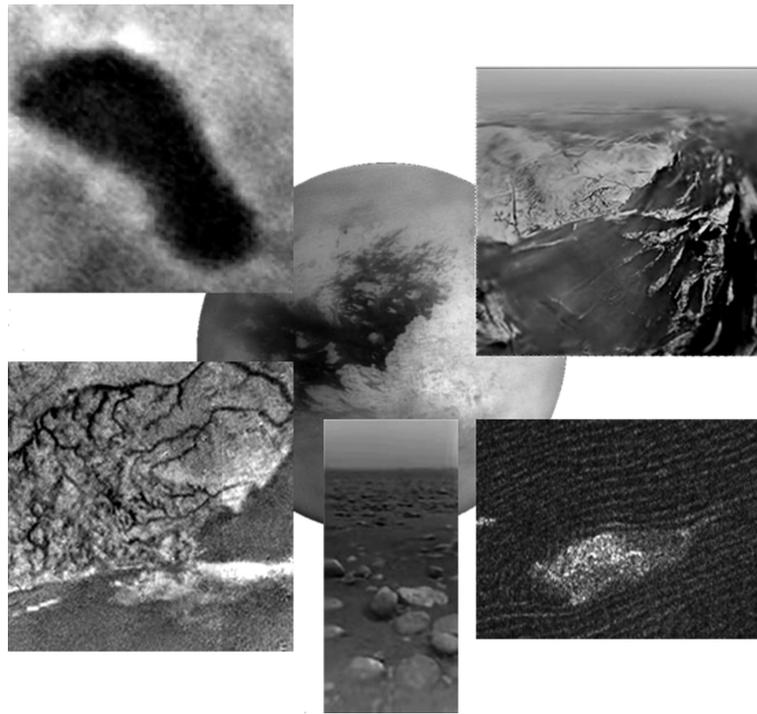


Fig. 2. Images of Titan's morphological diversity as taken from the Cassini satellite and Huygens probe.

to Los Angeles. Figure 1 demonstrates that the dataset primarily contains images dominated by sky, horizon, or undeveloped land. Some contain clouds, water bodies, patches of desert, or patches of vegetation.

To construct ground truth data for a discovery-oriented mission objective, we prompted four planetary geologists to manually group the dataset. All had some background in remote sensing, but their areas of emphasis differed (e.g., volcanology, paleobiology, geology). Images were presented in a graphical interface simultaneously and in random order. Scientists were allowed to name, rename, create, and delete groups throughout the session. Directions were provided in the form of the following prompt.

Suppose that the following images were taken of an environment for which we have little knowledge or data. Furthermore, suppose that you may not be able to receive all images. Please sort these images into five groups in such a way that if you could only receive a small number of images from each group, you could reasonably infer the content of the remaining images in that group.

Fixing the number of groups to five controlled for strong variations that may have resulted from scientists pursuing objectives of radically varying complexity.

3.2. Evaluation

We strive to automatically group images in a way that simultaneously agrees well with all scientist clustering strategies. Perfect correspondence is unlikely, but bandwidth utilization—and thus science return—is still improved if autonomous strategies can group data more effectively than nonadaptive strategies.

Two uninformed strategies for downlinking images are *random sampling*, where a random subset of an image sequence is returned, and *periodic sampling*, where every i th image of a sequence is returned. The former strategy amounts to a random clustering, while the latter is equivalent to clustering based on collection order.² We will subsequently refer to these strategies as random clustering and periodic or time-based clustering.

To compare correspondence between clusterings, we use the information-theoretic adjusted mutual information, which we briefly derive here. Given dataset $S = \{s_1, \dots, s_N\}$ and clusterings

$$U = \{U_1, \dots, U_R\} \quad \text{and} \quad V = \{V_1, \dots, V_C\},$$

where $\cap_{i=1}^R U_i = \emptyset$, and $\cup_{i=1}^R U_i = S$ (e.g., the U_i are a partitioning, or clustering, of S , and similarly for V). Then, the probability that a random data $s \in S$ is also contained in some cluster U_i or in cluster V_j is given by the respective equations

$$P_u(i) = \frac{|U_i|}{N} \quad \text{and} \quad P_v(j) = \frac{|V_j|}{N}.$$

The probability that s is found in both clusters is given by

$$P(i, j) = \frac{|U_i \cap V_j|}{N}.$$

The mutual information between the two labelings is defined as

$$MI(U, V) = \sum_{i=1}^R \sum_{j=1}^C \frac{P(i, j) \log P(i, j)}{P_u(i) P_v(j)}.$$

Mutual information quantifies how much knowing about one clustering tells us about the other. Though it is symmetric and nonnegative, it is not upper-bounded by a constant, and so is of limited utility as a general metric for comparing clusterings. Furthermore, Vinh et al. [2009] demonstrate that mutual information does not take a constant value when comparing random clusterings and tends to grow with the number of clusters. They use a hypergeometric model of randomness to derive an expected value for two random clusterings. This permits a correction similar to the adjusted rand index [Hubert and Arabie 1985] that ensures random clusterings produce a constant value. This correction yields the adjusted mutual information (AMI).

$$AMI(U, V) = \frac{MI(U, V) - E(MI(U, V))}{\max(H(U), H(V)) - E(MI(U, V))}. \quad (10)$$

The entropies of clusterings U, V denote the uncertainty in a data point's cluster membership.

$$H(U) = - \sum_{i=1}^R P(i) \log P(i) \quad \text{and} \quad H(V) = - \sum_{j=1}^C P(j) \log P(j).$$

The denominator in Equation (10) corrects for randomness and serves as a normalization, as otherwise $MI(U, V) \leq \min(H(U), H(V))$. Furthermore, $AMI(U, V) = 0$ only

²To see this, note that for N images and k clusters, each cluster will have N/k images, and cluster centroids will be located at every $N/2k$ images in the sequence. Varying the number of clusters k and returning the closest image to each centroid selects every i th image, for any i .

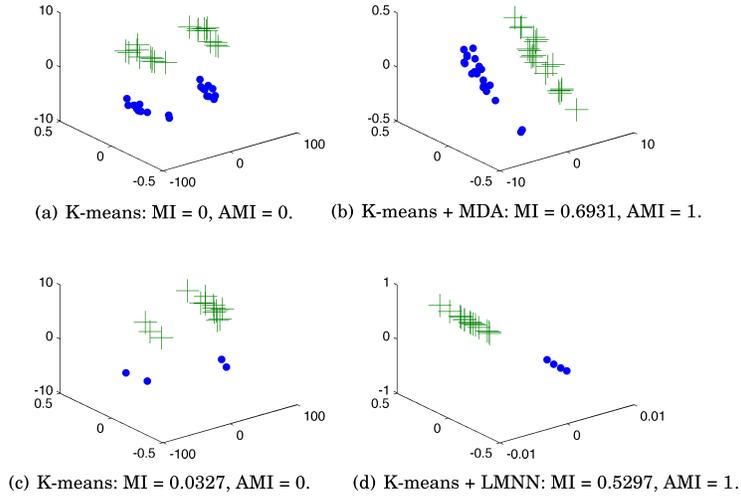


Fig. 3. Two-class clusterings of synthetic data.

when equal to its expected value (e.g., that expected by comparing two random clusterings), and $AMI(U, V) = 1$ when clusterings U, V are identical. In the experiments that follow, we use the AMI as a general measure of correspondence between clusterings.

Figure 3 demonstrates the advantages of both our metric-learning approach and our evaluation metric. Each figure shows two-class clusterings of synthetic data. On the left, (a) and (c) show four gaussians arranged into two classes such that k-means is unable to correctly cluster them. On the right, metric-learning algorithms bring all points of each class into proximity, enabling the desired clusterings. Note that the two classes in (a) are equal in size while the two classes in (b) are not. This affects the entropy of each class, which causes mutual information to give different values for (b) and (d), despite both being clustered perfectly. AMI permits a fair comparison between the two cases.

3.3. Features

Our primary concern is that image descriptors be fast enough for operation onboard power-restricted spacecraft computers (discussed in Section 5). To that end, we do not examine the more common but computationally expensive and higher-dimensional approaches, such as those based on SIFT features [Lowe 1999], Gabor filters [Grigorescu et al. 2002], or correlograms [Huang et al. 1997]. Instead, we present 19 low-level, computationally efficient features that summarize texture, color, and the temporal ordering of each image. Except where noted, all features can be computed in linear time with respect to image size and can be trivially parallelized. Each feature is defined in Table I; relevant notation and definitions are established in subsections that correspond to the Feature Type column.

3.3.1. Texture. Let I be an $m \times d \times 3$ image and I' be its $m \times d$ grayscale. Let I'' be the resulting binary image from performing convolution with a Sobel operator on image I' . Let G_y, G_x be the $m \times d$ matrices representing vertical and horizontal gradient responses, respectively. Let $\nabla L = \sqrt{G_y^2 + G_x^2}$ be the $m \times d$ matrix representing gradient magnitudes, and $\theta = \text{atan2}(G_y, G_x)$ be the $m \times d$ matrix representing gradient orientations. Then, edge density and magnitude can help distinguish smooth from

Table I. Low-Level Image Descriptors

Feature Type	Feature Name	Formula
Texture	Edge Density	$\frac{1}{md} \sum I'$
	Mean Gradient Magnitude	$\frac{1}{md} \sum \nabla L$
	Gradient Magnitude Entropy	$H(\nabla L)$
	Gradient Orientation Entropy	$H(\theta)$
	Quadrant-1 Frequency Energy	$\sum_{x>0,y>0} N$
	Quadrant-2 Frequency Energy	$\sum_{x<0,y>0} N$
Color	Standard Deviation of Band Intensity	$\sqrt{\frac{1}{md} \sum (P_j - \mu)^2}$ for $j = 1 \dots 3$
	Mean Band Intensity	$\frac{1}{md} P_j$ for $j = 1 \dots 3$
	Min Band Intensity	$\min P_j$ for $j = 1 \dots 3$
	Max Band Intensity	$\max P_j$ for $j = 1 \dots 3$
Time	Acquisition Order	I_t

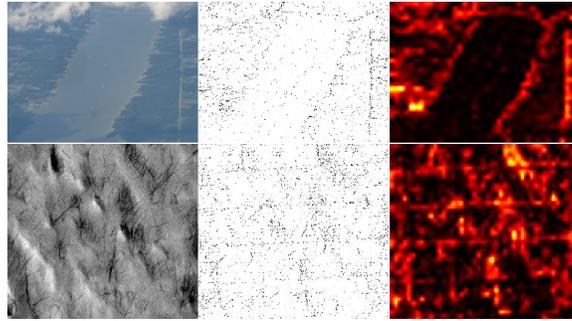


Fig. 4. Statistics such as edge density (visualized by heat maps on the right) taken on an edge image (middle) can help distinguish smooth (top-left) from rough (bottom-left) terrain.

rough terrain, while gradient magnitude and orientation can provide information on texture regularity, which can help distinguish periodic textures (e.g., sand dunes, large expanses of sediment). Figure 4 shows a visualization of the edge densities between images with smooth and rough terrain.

Let f be the resulting $m \times d$ matrix after the 2D Fourier transform on I' , and let $G(f) = |f|^2$ be the $m \times d$ power spectrum of f . Let the normalized power spectrum be $N = G(f) / \sum G(f)$. Then, taking the energy in the first and second quadrants (which are symmetric to the fourth and third) can provide good texture discrimination.

The frequency energy statistics were motivated by Liu and Jernigan [1990], which were found by a survey of 28 low-level frequency space statistics to best discriminate a subset of the Brodatz textures. The frequency features are an exception to the linear-time constraint since they require a fast Fourier transform. This is an $n \log n$ operation, though there are many efficient hardware and software implementations (we also touch on this in Section 5).

3.3.2. Color. Let P_j be the $m \times d$ matrix of pixels in band j of I . Then histogram summarization statistics taken on each band can help distinguish verdant from mountainous terrain, clouded from cloudless skies, and dried river basins from rivers. Figure 5 show an example.

3.3.3. Temporal. For a sequence of images collected in serial, let the acquisition order I_t be an integer that represents the temporal ordering of each image.

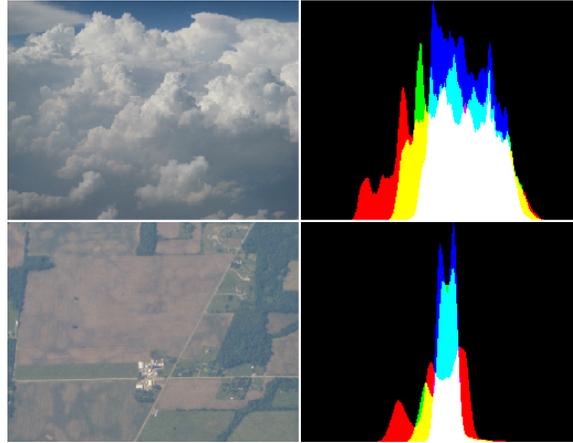


Fig. 5. Statistics taken on image histograms (right) can make coarse distinctions in terrain type, for instance, clouded skies (top-left) and undeveloped land (bottom-left).

4. RESULTS

We begin by examining the clusterings performed by each scientist. Figure 6 depicts class distributions and names for each scientist.

Priorities differed among the scientists. For instance, Scientists 3 and 4 made four distinctions in atmospheric qualities of the imagery, while Scientist 1 made two distinctions, and Scientist 2 placed them into a single group. Roughly 40% of labels from Scientists 1 and 4 focused on atmospheric distinctions, compared to 68% for Scientist 3 and 14% for Scientist 2. All scientists, except for the third, created a group based on the presence of water, but these comprised only 6% of Scientist 4's labels, compared to at least 30% for Scientists 1 and 2.

Many of the scientists' clusters corresponded for at least some images. For example, scientists consistently ascribed similar labels to images, such as (a) and (b), where the horizon was roughly centered and cloud cover was evident. Images, such as (c) and (d), that contained water bodies were often labeled consistently but only when there was no horizon or clouds present. The labeling consensus was weaker in the presence of sky or clouds. While Scientist 2 typically labeled an image as "Hydrology" if it contained even a very small lake, Scientists 3 and 4 would often favor horizon and cloud labels as they began to become more prevalent, as can be seen with images (g) and (h). Unsurprisingly, Scientist 1's "Desert" label sometimes corresponded with Scientist 2's "Mountain" label, and both were often contained within the "Land" labels of Scientists 3 and 4 (as in images (e) and (f)).

These differences imply that scientists were pursuing different objectives. In follow-up discussions, some suggested distinctions in atmospheric qualities had been most important, while others preferred to disambiguate features of the terrain. Table II quantifies the agreements between scientists by showing the adjusted mutual information between scientists' clusterings. Scientists 1 and 3 had the highest agreement, at 0.4232. Scientist 2 appears as an outlier, with the lowest mean agreement among other scientists of 0.184, as compared to agreements of around 0.3 among the remaining scientists. The overall mean agreement of 0.285 is the expected mean AMI of any human strategy against the rest; it establishes a milestone for how well autonomous clusterings might simultaneously correspond to scientists' clusterings.

We compared the correspondence of manual clusterings with various automated approaches, including static data return strategies like random or periodic sampling. We

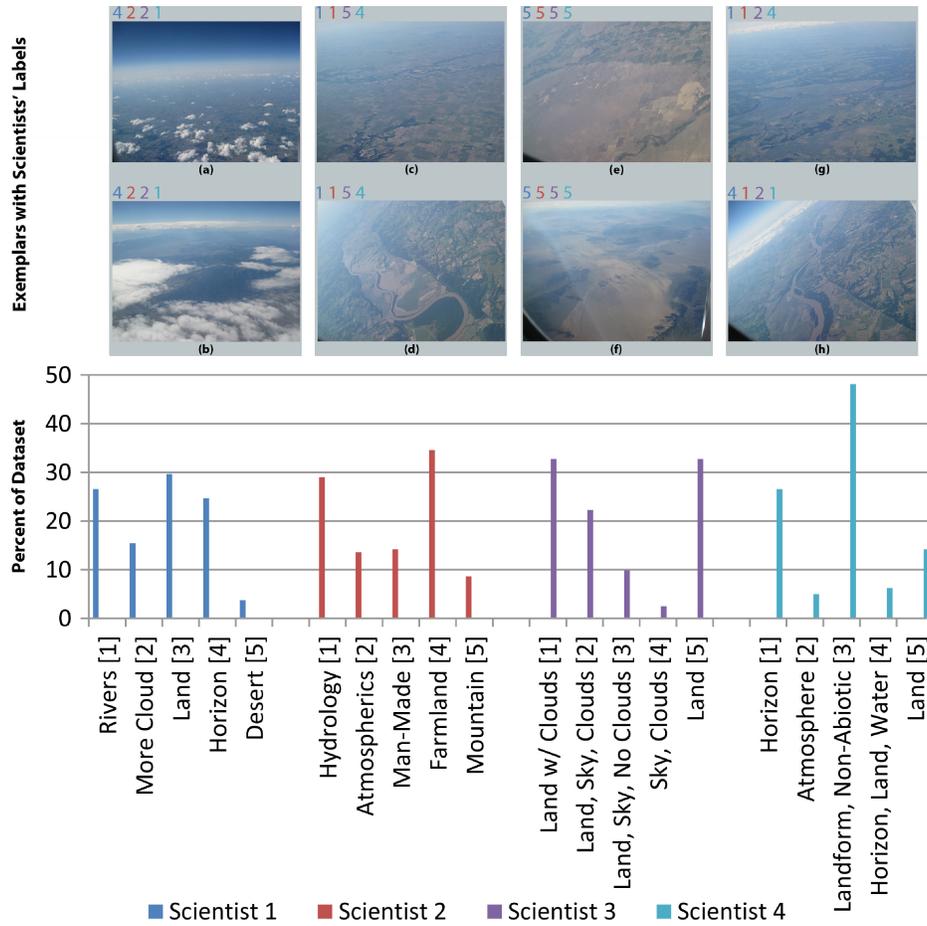


Fig. 6. Bottom: names and distributions of labels provided by four planetary geologists. Top: images that highlight partial correspondence among scientists' labelings (a)–(f), in addition to some differing objectives (g)–(h). Note, the color-coated integers above each image denote the class labels assigned by each scientist. Labels can be referenced by their trailing index.

Table II. Clustering Agreement Among Scientists as Measured by Adjusted Mutual Information

	Scientist 1	Scientist 2	Scientist 3	Scientist 4
Scientist 1	1	0.2492	0.3627	0.4232
Scientist 2	0.2492	1	0.1102	0.1932
Scientist 3	0.3627	0.1102	1	0.3696
Scientist 4	0.4232	0.1932	0.3696	1
Mean Agreement	0.3450	0.1840	0.2810	0.3290

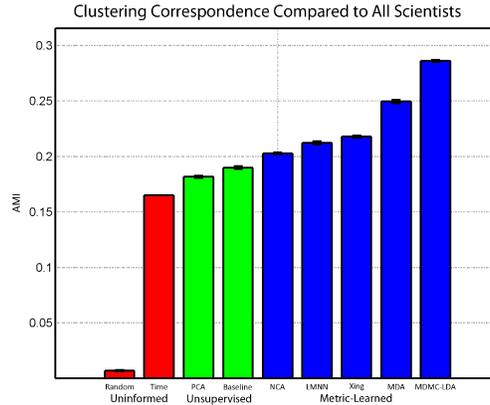


Fig. 7. Mean adjusted mutual information between all scientists’ clusterings and uninformed strategies (red), unsupervised autonomous strategies (green), and metric-learned autonomous strategies (blue). Although the uninformed and unsupervised strategies do not benefit from training data, all clusterings are compared over 1,000 trials using the same test set, which comprised 75% of the data.

also considered clustering approaches using the entire feature set, as well as feature spaces learned by metric-learning strategies based on scientist-labeled training data. We trained the metric-learning strategies on a uniformly sampled 25% training set, starting four times at image indices 1 to 4 for multifold cross validation. For each trial, we performed an automated k-means clustering on the remaining 75% of the data and evaluated the result using the AMI score. Because k-means results can be sensitive to initialization, we performed 1,000 randomized runs for each cross-validation subset.

Figure 7 compares mean clustering agreement between all scientists and strategies based on uninformed and autonomous methods. The red uninformed approaches represent random and periodic sampling. The unsupervised approaches in green represent strategies that do not benefit from training data. The baseline approach is simple clustering with all 19 features from Section 3.3. The PCA approach uses principal components analysis to reduce dimensionality. The metric-learning strategies compare neighborhood components analysis (NCA), largest-margin nearest neighbor (LMNN), multiclass discriminant analysis (MDA), and multi-domain multiclass linear discriminant analysis (MDMC-LDA). MDMC-LDA used training labels from all scientists. The remaining metric-learning approaches were restricted to learning from a single set of labels. Experiments validated our intuition that learning would be most productive on the labels of the first scientist, who had the highest overall agreement with others. Dimensionality was reduced to $d = 2$ for PCA, NCA, LMNN, MDA, and MDMC-LDA. Comparisons were averaged over 1,000 trials.

Periodic sampling produced a mean agreement of 0.1648. Unsupervised methods marginally outperforms it at 0.1898 for clustering in 19 dimensions, and 0.1816 for clustering with PCA. Metric-learning strategies show obvious benefit over periodic sampling, with correspondence improving from 23% for NCA to 73.5% for MDMC-LDA. It is interesting to note that the more classical MDA shows marked improvement over contemporary approaches.

Figure 8 compares mean clustering agreement relative to the first scientist. The same trends are evident: metric learning significantly outperforms unsupervised strategies, and both outperform uninformed strategies. All methods show improved correspondence. Remarkably and counterintuitively, MDMC-LDA significantly outperforms all other approaches despite having to balance competing objectives. With

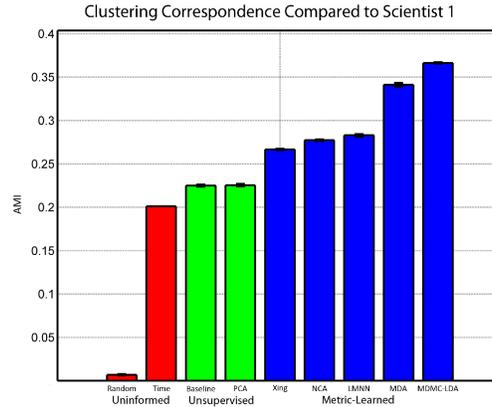


Fig. 8. Mean adjusted mutual information between Scientist 1’s clustering and uninformed strategies (red), unsupervised autonomous strategies (green), and metric-learned autonomous strategies (blue). Although the uninformed and unsupervised strategies do not benefit from training data, all clusterings are compared over 1,000 trials using the same test set, which comprised 75% of the data.

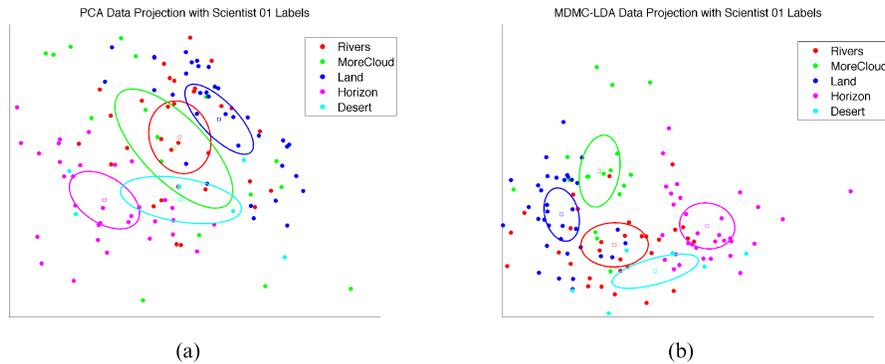


Fig. 9. Effects of metric learning on test data. (a) Shows an unsupervised projection; clusters are overlapping and indistinct. (b) Shows a metric-learned projection. Scientists’ clusters are better-separated in this synthetic feature space, suggesting these features better capture the science content of interest.

$AMI = 0.3662$, MDMC-LDA produces a clustering that more closely matches this scientist than do the clusterings of all but one other scientist.

Ideal clustering agreement would approach 1.0, particularly when autonomous clusterings are attempting to match a single objective. The relatively low scientist agreement in our results suggests that such high correspondence will not be possible in the face of multiple mission objectives. Even so, with correspondence nearly double that of alternative strategies, clustering informed by metric learning could provide substantial gains in science return over a mission’s lifetime. Figure 9 provides some intuition on what clusters look like in a learned space. Figure 9 (a) displays a 2D projection of the test set using an unsupervised strategy (PCA) based only on the intrinsic structure of the data. Figure 9 (b) displays the same data projected with MDMC-LDA. Colors correspond to labels provided by the first scientist. Except in the case of the sparse ‘Desert’ cluster (comprising 3.7% of the data), cluster compactness and consistency is substantially improved when data is projected into a space that accounts for the scientist’s preferences. If a cluster representative were downlinked from each cluster (the image closest to each centroid), then three classes would be accurately identified. And

even in the case of the red ‘River’ cluster, we would see reasonable behavior in that a ‘Land’ image would be downlinked. In other words, the learned feature space emphasizes distinctions in science content that is not obvious from the distribution of base feature vectors.

5. DISCUSSION

Many current space-based missions, such as the exploration of Mars by the Spirit and Opportunity rovers, or the imaging of Earth or Mars by orbiters, involve research groups from around the world, each competing over limited bandwidth. Orbiter targets are scheduled a year or more in advance. Rover actions and downlink priorities are debated by teleconference on a daily or weekly basis. In these static and largely observable environments, competing requests can be individually serviced. In essence, bandwidth can be distributed among different groups, even if it’s not to the satisfaction of all involved.

This would not be so easy in a dynamic or previously unobserved environment, such as Titan. While groups could influence the trajectory of an aerobot, it seems unlikely that downlink bandwidth could be precisely allocated without target detection capabilities that catered to the needs of individual groups. These objectives would be difficult to anticipate before launch, but they would almost certainly be in competition in the same way that scientists’ labels on our terrestrial dataset were competing. Supporting the changing objectives of each group with target detectors or other custom autonomy code would be hugely expensive, if not completely impractical, in large part due to the complexities involved in coding flight software [Dvorak et al. 2009].

Metric learning and clustering provide a mechanism for supporting changing mission objectives by allowing scientists on the ground to further inform downlink decisions through small uplinks to the spacecraft that do not require modification to flight software. Although onboard feature extraction will not undergo many changes over a mission, scientists on the ground are free to experiment with training data and metric-learning techniques without concern for computational or flight hardware complexity. Our MDMC-LDA approach shows particular promise because it can balance multiple, competing objectives, perhaps without making great concessions to any individual objective. Future work will investigate this behavior—empirically on different datasets and theoretically on how objectives are balanced.

Our approach (feature extraction, clustering, metric learning) has been carefully chosen so that it is feasible on current spacecraft hardware (notably, BAE Systems’ RAD750³, in use on dozens of spacecraft). We accomplished this by keeping to fast, subpolynomial runtime. Future spaceflight processors will have greater computational power, with parallelism playing an increasing role. Potential architectures include OPERA⁴, the Xilinx Virtex 5 Pro⁵, and Coherent Logix’s Hyper X.⁶ For our current approach, all feature extractors are trivially parallelized except for the frequency-domain texture descriptors, which rely on a Fourier transform. Work on ever-faster, even parallel implementations of the fast Fourier transform remains active, however. The MIMD (multiple instruction, multiple data) parallelization of the FFT in

³Details on the RAD750 can be found at www.baesystems.com. Here, we note that it runs at up to 200MHz, has 128MB RAM, and is capable of approximately 400 MIPS—several orders of magnitude slower and about a decade behind modern desktops at the time of this writing.

⁴OPERA (Onboard Processing Expandable Reconfigurable Architecture) was presented at SPC08, the First Workshop on Fault-Tolerant Spaceborne Computing Employing New Technologies, hosted at Sandia Labs. At the time of this writing, slides can be downloaded at www.zettaflops.org.

⁵<http://www.xilinx.com/>

⁶<http://coherentlogix.com>

Averbuch et al. [1990] would be appropriate for the OPERA platform. In addition to the optimizations we previously cited, k-means can also be parallelized (see for instance, Stoffel and Belkoniene [1999]).

With our current approach being amenable to parallelization, there may be headroom for more sophisticated clustering and feature extraction approaches as spacecraft hardware advances. The scientist labeling strategies that we saw for our terrestrial imagery demonstrated that a single image might be considered interesting based on different criterion, such as dominant terrain type (e.g., sky, desert, mountain), or presence of a specific, possibly small feature (e.g., clouds, water). Our approach would likely be more flexible to these objectives if higher-granularity summarization statistics were employed, such as textons [Julesz 1981], wavelets [Manjunath and Ma 2002], or color-based features [Burghouts and Geusebroek 2009]. These descriptors often require alternative similarity measures (e.g., earth-movers distance [Rubner et al. 2000]), which are not readily implementable with k-means but could be with kernel-based methods. Additional spatial resolution can be obtained by taking statistics in the neighborhoods of interest points detected based on saliency, or more expensively, within individually detected regions or segments [Tuytelaars and Mikolajczyk 2008]. Future work will explore these methods on additional datasets and with larger numbers of scientists.

6. CONCLUSION

The disparity between the amount of data that can be collected during the course of future space exploration missions and the amount that can be downlinked back to Earth is likely only going to increase. Onboard clustering provides scientists and mission planners with improved control over downlink bandwidth through selective data-return strategies, including representative sampling. Metric learning can tune clusterings for specific, even changing mission objectives through small uploads that do not require changes to flight software. Our new MDMC-LDA algorithm extends metric learning so that multiple, possibly competing objectives can be simultaneously accounted for. Using an aerial dataset and algorithms and features that could be implemented on current spaceflight hardware (and parallelized for future hardware), we show that autonomous clustering more closely corresponds with the way that four planetary geologists grouped images than do nonadaptive random or periodic sampling strategies. Metric-learning approaches further improved correspondence, with MDMC-LDA substantially outperforming contemporary methods.

REFERENCES

- ARTHUR, D., MANTHEY, B., AND ROGLIN, H. 2009. k-Means has polynomial smoothed complexity. In *Proceedings of the 50th Annual IEEE Symposium on Foundations of Computer Science*. IEEE, 405–414.
- AVERBUCH, A., GABBER, E., GORDISSKY, B., AND MEDAN, Y. 1990. A parallel FFT on an MIMD machine* 1. *Parall. Comput.* 15, 1–3, 61–74.
- BURGHOUTS, G. AND GEUSEBROEK, J. 2009. Performance evaluation of local colour invariants. *Comput. Vision Image Understand.* 113, 1, 48–62.
- CALDERON, F., THOMPSON, D. R., AND WETTERGREEN, D. 2008. Autonomous rover reflectance spectroscopy with dozens of targets. In *Proceedings of the International Symposium on Artificial Intelligence, Robotics and Automation in Space (iSAIRAS)*.
- CASTANO, A., FUKUNAGA, A., BIESIADECKI, J., NEAKRASE, L., WHELLEY, P., GREELEY, R., LEMMON, M., CASTANO, R., AND CHIEN, S. 2008. Automatic detection of dust devils and clouds on Mars. *Machine Vision Appl.* 19, 5–6, 467–482.
- CASTAÑO, R., JUDD, M., ANDERSON, R., AND ESTLIN, T. 2003. Machine learning challenges in Mars rover traverse science. In *Proceedings of the ICML Workshop on Machine Learning Technologies for Autonomous Space*.
- CASTAÑO, R., ESTLIN, T., ANDERSON, R. C., GAINES, D. M., CASTANO, A., BORNSTEIN, B., CHOUINARD, C., AND JUDD, M. 2007. Oasis: Onboard autonomous science investigation system for opportunistic rover science. *J. Field Robotics*, 379–397.

- CHEN, J., ZHAO, Z., YE, J., AND LIU, H. 2007. Nonlinear adaptive distance metric learning for clustering. In *Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 123–132.
- CHEN, L., LIAO, H., KO, M., LIN, J., AND YU, G. 2000. A new LDA-based face recognition system which can solve the small sample size problem. *Pattern Recog.* 33, 10, 1713–1726.
- CHIEN, S., SHERWOOD, R., TRAN, D., CICHY, B., RABIDEAU, G., CASTANO, R., DAVIES, A., MANDL, D., FRYE, S., TROUT, B., ET AL. 2005. Using autonomy flight software to improve science return on Earth Observing One. *J. Aerospace Comput. Infor. Comm.* 2, 4, 196–216.
- CLOUGH, B. 2002. Metrics, schmetrics! How the heck do you determine a UAV's autonomy anyway. In *Proceedings of the Performance Metrics for Intelligent Systems Workshop*.
- DUDA, R., HART, P., AND STORK, D. 2000. *Pattern Classification*. Wiley-Interscience, New York, NY.
- DUIN, R. AND LOOG, M. 2004. Linear dimensionality reduction via a heteroscedastic extension of LDA: The Chernoff criterion. *IEEE Trans. Pattern Anal. Machine Intell.* 26, 6, 732–739.
- DVORAK, D. ET AL. 2009. NASA study on flight software complexity.
- ELFES, A., HALL, J., KULCZYCKI, E., CLOUSE, D., MORFOPOULOS, A., MONTGOMERY, J., CAMERON, J., ANSAR, A., AND MACHUZAK, R. 2008. Autonomy architecture for aerobot exploration of Saturnian moon Titan. *IEEE Aerospace and Electronic Systems Mag.* 23, 7, 16–24.
- ELKAN, C. 2003. Using the triangle inequality to accelerate k-means. In *Proceedings of the International Conference on Machine Learning*. Vol. 20, 147–153.
- ESTLIN, T., BORNSTEIN, B., GAINES, D., THOMPSON, D., CASTAÑO, R., ANDERSON, R., DE GRANVILLE, C., BURL, M., JUDD, M., AND CHIEN, S. 2010. Aegis automated targeting for the MER opportunity rover. In *Proceedings of the International Symposium on Space Artificial Intelligence, Robotics, and Automation for Space*.
- FRIEDMAN, J. 1989. Regularized discriminant analysis. *J. Amer. Stat. Assoc.* 84, 405, 165–175.
- GERSHO, A. AND GRAY, R. 1992. *Vector Quantization and Signal Compression*. Springer, Berlin.
- GILMORE, M., CASTAÑO, R., MANN, T., ANDERSON, R., MJOLSNESS, E., MANDUCHI, R., AND SAUNDERS, R. 2000. Strategies for autonomous rovers at Mars. *J. Geophysical Res.-Planets*.
- GOLDBERGER, J., ROWEIS, S., HINTON, G., AND SALAKHUTDINOV, R. 2005. Neighbourhood components analysis. In *Advances in Neural Information Processing Systems 17*, L. K. Saul, Y. Weiss, and L. Bottou Eds., MIT Press, Cambridge, MA, 513–520.
- GRIGORESCU, S., PETKOV, N., AND KRUIZINGA, P. 2002. Comparison of texture features based on gabor filters. *IEEE Trans. Image Process.* 11, 10, 1160–1167.
- GUO, Y., HASTIE, T., AND TIBSHIRANI, R. 2007. Regularized linear discriminant analysis and its application in microarrays. *Biostatistics* 8, 1, 86.
- HAMSIKI, O. AND MARTINEZ, A. 2007. Bayes optimality in linear discriminant analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* 647–657.
- HAYDEN, D. S., CHIEN, S., THOMPSON, D. R., AND CASTANO, R. 2010. Onboard clustering of aerial data for selective data return. In *Proceedings of the 10th International Symposium on Artificial Intelligence, Robotics, and Automation in Space*.
- HOWLAND, P., JEON, M., AND PARK, H. 2004. Structure preserving dimension reduction for clustered text data based on the generalized singular value decomposition. *SIAM J. Matrix Anal. Appl.* 25, 1, 165–179.
- HUANG, J., KUMAR, S., MITRA, M., ZHU, W., AND ZABIH, R. 1997. Image indexing using color correlograms. In *Proceedings of the Computer Vision and Pattern Recognition*. IEEE Computer Society, 762.
- HUBERT, L. AND ARABIE, P. 1985. Comparing partitions. *J. Classification* 2, 193–218.
- JAIN, P., KULIS, B., DHILLON, I. S., AND GRAUMAN, K. 2009. Online metric learning and fast similarity search. In *Advances in Neural Information Processing Systems 21*, D. Koller, D. Schuurmans, Y. Bengio, and L. Bottou Eds., 761–768.
- JULESZ, B. 1981. Textons, the elements of texture perception, and their interactions. *Nature* 290, 5802, 91.
- KINSEY, J., EUSTICE, R., AND WHITCOMB, L. 2006. A survey of underwater vehicle navigation: Recent advances and new challenges. In *Proceedings of the 7th Conference on Maneuvering and Control of Marine Craft (MCMC'06)*.
- LIU, S. AND JERNIGAN, M. 1990. Texture analysis and discrimination in additive noise. *Comput. Vision Graph. Image Process.* 49, 1, 52–67.
- LLOYD, S. 1982. Least square quantization in pcm. *IEEE Trans. Inform. Theor.* 28.
- LOWE, D. 1999. Object recognition from local scale-invariant features. In *Proceedings of the International Conference on Computer Vision*. IEEE Computer Society, 1150.

- MACQUEEN, J. ET AL. 1967. Some methods for classification and analysis of multivariate observations. In *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability*. Vol. 1, 14.
- MANJUNATH, B. AND MA, W. 2002. Texture features for browsing and retrieval of image data. *IEEE Trans. Pattern Anal. Mach. Intell.* 18, 8, 837–842.
- NASA. 2006. 2006 solar system exploration roadmap. Tech. rep., NASA, Washington, D.C.
- QIN, A., SUGANTHAN, P., AND LOOG, M. 2005. Uncorrelated heteroscedastic LDA based on the weighted pairwise Chernoff criterion. *Pattern Recog.* 38, 4, 613–616.
- RUBNER, Y., TOMASI, C., AND GUIBAS, L. 2000. The earth mover’s distance as a metric for image retrieval. *Int. J. Comput. Vision* 40, 2, 99–121.
- SAEGUSA, T. AND MARUYAMA, T. 2007. An FPGA implementation of k-means clustering for color images based on Kd-tree. In *Proceedings of the International Conference on Field Programmable Logic and Applications (FPL06)*. IEEE, 1–6.
- STOFFEL, K. AND BELKONIENE, A. 1999. Parallel k/h-means clustering for large data sets. In *Proceedings of the Euro-Par’99 Conference on Parallel Processing*. 1451–1454.
- THOMPSON, D., SMITH, T., AND WETTERGREEN, D. 2008. Information-optimal selective data return for autonomous rover traverse science and survey. In *Proceedings of the International Conference on Robotics and Automation (ICRA’08)*. IEEE, 968–973.
- THRUN, S., THAYER, S., WHITTAKER, W., BAKER, C., BURGARD, W., FERGUSON, D., HAHNEL, D., MONTEMERLO, D., MORRIS, A., OMOHUNDRO, Z., ET AL. 2005. Autonomous exploration and mapping of abandoned mines. *IEEE Robot. Automation Mag.* 11, 4, 79–91.
- TUYTELAARS, T. AND MIKOLAJCZYK, K. 2008. Local invariant feature detectors: A survey. *Found. Trends® in Comput. Graph. Vision* 3, 3, 177–280.
- VINH, N. X., EPPS, J., AND BAILEY, J. 2009. Information theoretic measures for clusterings comparison: Is a correction for chance necessary? In *Proceedings of the 26th International Conference on Machine Learning*. L. Bottou and M. Littman Eds., 1073–1080.
- WEINBERGER, K., BLITZER, J., AND SAUL, L. 2006. Distance metric learning for large margin nearest neighbor classification. In *Advances in Neural Information Processing Systems 18*, Y. Weiss, B. Schölkopf, and J. Platt Eds., MIT Press, Cambridge, MA, 1473–1480.
- WETTERGREEN, D., WAGNER, M. D., JONAK, D., BASKARAN, V., DEANS, M., HEYS, S., PANE, D., SMITH, T., TEZA, J., THOMPSON, D. R., TOMPKINS, P., AND WILLIAMS, C. 2008. Long-distance autonomous survey and mapping in the robotic investigation of life in the atacama desert. In *Proceedings of the International Symposium on Artificial Intelligence, Robotics and Automation in Space (iSAIRAS)*.
- XING, E., NG, A., JORDAN, M., AND RUSSELL, S. 2002. Distance metric learning, with application to clustering with side-information. In *Advances in Neural Information Processing Systems 16*, S. Becker, S. Thrun, and K. Obermayer Eds., MIT Press, Cambridge, MA, 521–528.
- XU, R. AND WUNSCH, D. 2005. Survey of clustering algorithms. *IEEE Trans. Neural Netw.* 16, 3, 645–678.
- YANG, L. 2007. An overview of distance metric learning. In *Proceedings of the Computer Vision and Pattern Recognition Conference*.
- YANG, L. AND JIN, R. 2007. Bayesian active distance metric learning. In *Proceedings of the Uncertainty in Artificial Intelligence*.
- YE, J. 2006. Characterization of a family of algorithms for generalized discriminant analysis on undersampled problems. *J. Mach. Learn. Res.* 6, 1, 483.

Received January 2011; revised July 2011; accepted September 2011