Optimizing Change Detection for Planetary Remote Sensing Datasets

Reid Parsons¹,² and Hideaki Miyamoto³

¹Earth and Geographic Sciences Dept., Fitchburg State University, Fitchburg, MA, USA
²Department of Museum Collection Utilization Studies, University of Tokyo, Japan
³Department of Systems Innovation, School of Engineering, University of Tokyo, Japan
E-mail: rparson4@fitchburgstate.edu

Abstract. The fast-growing Remote Sensing dataset for Mars has challenged scientists in the community to make full use of the imagery collected from spacecraft. Cameras aboard long-lived orbiter missions capturing single-band, high-resolution images provide a special opportunity to monitor surface and atmospheric changes associated with geologic, weather and climate processes. A potential method for detecting changes is to employ principle component analysis (PCA) - an existing image processing method normally applied to multi-spectral imagery. When PCA is applied to a time-series stack of single-band images of the same location, subtle changes in pixel brightness are enhanced, and are represented as a series of transformed image components. PCA of a time-series image stack sorts these variations in the dataset according to their amplitudes. The most significant pixel variations are found in the lower principle components (PCs) while more subtle changes in pixel brightness are revealed in higher PCs. This method makes visual identification of subtle, ongoing geologic processes easier as is shown in a series of HiRISE images containing recurring slope lineae (RSL). As a final step, the change detection method proposed herein can estimate the likelihood of each of the original images containing the feature found in the PCA components. This is performed via a PCA-to-image coordinate rotation using the Eigen matrix calculated from the original PCA analysis. This proposed method is called “Change Detection via PCA of Stacked Time-series” (CDPCAST).

1. Introduction
Change detection is an analytical tool of Remote Sensing in which repeat observations of a particular location can be overlain, subtracted, or ratioed to detect differences between images. Many change detection techniques have been developed for multi-spectral sensors (cameras capturing several wavelength bands simultaneously) to monitor landcover and environmental changes on Earth (see, for example, [1, 2, 3]). Mars-observing spacecraft, however, commonly use a high-resolution, single band (capturing one range of light wavelength) a camera sensor resulting in grayscale images. Repeat images taken at the same location on Mars using these cameras as led to several discoveries of ongoing geologic processes operating at the sub-kilometer scale.

The arrival of the first high-resolution imaging sensors began with the Mars Orbital Camera (≈ 2 m/pixel) in 1997 aboard Mars Global Surveyor and provided images which suggested
liquid water was driving ongoing geologic activity in the formation of gullies [4]. Subsequent observations in a variety of locations made by the High Resolution Imaging Science Experiment (HiRISE, 0.25 m/pixel) aboard the Mars Reconnaissance Orbiter (2006 arrival) provided evidence for aqueous activity in forming RSL [5] which were previously interpreted as landslides [6]. RSL undergo a repetitive seasonal cycle in which dark, finger-like albedo changes originate near rock outcrops in early summer [7] and extend down steep, dust-covered slopes (see Figure 1). Eventually, the lineae stop growing, and slowly lighten in color, sometimes becoming brighter than their surroundings [6].

Figure 1. Uncalibrated HiRISE time-series image stack located on the southeastern wall of the 16 km diameter Palikir Crater (41.65°S, 202.71°E). Image IDs are given in yellow. The calendar date of acquisition are listed hereafter along with the Martian solar longitude ($L_s$) where $L_s = 180^\circ$, $L_s = 270^\circ$, and $L_s = 360^\circ$ corresponds to the spring equinox, summer solstice, and fall equinox in the southern hemisphere: a) 2007/11/02 ($L_s = 341^\circ$) b) 2009/01/02 ($L_s = 185^\circ$) c) 2010/02/26 ($L_s = 57^\circ$) d) 2011/03/03 ($L_s = 247^\circ$) e) 2011/03/30 ($L_s = 265^\circ$) f) 2011/04/27 ($L_s = 282^\circ$) g) 2011/05/30 ($L_s = 302^\circ$) h) 2011/06/10 ($L_s = 309^\circ$) i) 2011/06/27 ($L_s = 318^\circ$).

RSL are one example of features which continue to alter the Martian surface, but other features include new, small impacts [8], aeolian dune migration [9], sublimation of polar ice [10], and gully formation [11]. Generally, the detection of these changes is done manually. The scientist or user of the remote sensing data viewed the data in GIS (Geographic Information Systems) software and visually identified changes by overlaying, and “flipping” between the two images. As planetary remote sensing data sets continue to grow, so does the need for a more efficient means of change detection which is suited toward high-resolution single-band imagery.

The potential advantage of using PCA as a means for detecting change, is that it considers the values of all pixels in all the images to determine the most statistically significant variations in pixel values. These variations are then isolated from one another as a hierarchical set of “principle components” of the image dataset. The user can quickly identify patterns or features of interest which are present in one or more of these components. This method, hereafter referred
to as CDPCAST, provides a means for locating changes both spatially and temporally in repeat imagery. Nine HiRISE images collected from Palikir crater on Mars between November 2007 and June 2011 documenting RSL formation (Figure 1) will be used to illustrate the CDPCAST method.

2. Methods

2.1. Preprocessing

Existing methods for detecting changes in repeat imagery (including the one proposed here) is a step-wise process which begins with image preprocessing (Figure 2). Preprocessing involves image calibration (converting camera data into reflectance), orthorectification (removing geometric distortion associated with viewing angle and topography), and co-registration (aligning images of the same location so that they perfectly overlap). Although image co-registration has traditionally been a time-consuming process of manually selecting identical points in a pair of images, progress has been made in automating this procedure [12, 13].

The HiRISE images of Mars used here were not calibrated into reflectance data as it is not a required step in the change detection process. In this case, the same camera was used to collect all images so the relative change in brightness can be detected using the raw data. This example helps to illustrate potential broader application of this technique to other fields using repeat images and different camera data (e.g. medical imaging). In the example here, HiRISE images were orthorectified and co-registered by the HiRISE Operations Center and are provided to the public [14].

In addition to these initial steps, CDPCAST requires that each image in a time series be of the same extent and resolution and aligned (via coregistration) so that they can be “stacked” into a single, time-series image file. This final step was conducted using the open-source QGIS
Application. Figure 1 shows each of the preprocessed, uncalibrated HiRISE images in the final image stack. Note that the extent of these individual images has been cropped to a particular region of interest and region of overlap. The source image ID is given in yellow, axis labels indicate numbers of pixels, and the color bar indicates the pixel values. These images are arranged in order of time taken and each has a resolution of 0.25 m/pixel.

2.2. Principle Component Analysis
PCA is a common data reduction method employed in the field of remote sensing, but is most often applied to multi- or hyper-spectral images in order to distinguish between different materials in an image by accentuating differences in reflectivity at different wavelengths of light. In this standard application, the input image is a stack of images taken at many different wavelengths simultaneously. The CDPCAST method, in contrast, performs a PCA on a stack of images taken at the same wavelength at many different times. Although prior studies have employed PCA as a means of change detection on Earth using images taken at different times [15, 16, 17], the remote sensing community has focused more on alternative methods [1]. On planetary surfaces where geologic activity results in small features (requiring high-spatial resolution) which don’t carry a strong spectral signal (dust-covered surfaces), PCA of single-band images stacked into a time-series could greatly improve change detection.

Figure 3. An illustration of PCA using only the sixth and seventh images from the time-series. a) The seventh image in the stack (Figure 1g) has well-developed, dark RSL features forming (highlighted in red in Figure 3a), but
those pixels were lighter in color in the previous image from this location (Figure 1f). Therefore, plotting the values of the sixth image versus those of the seventh image, gives a cloud of data points which are shown in Figure 3b. The upward trend along the line $y = x$ indicates close correlation of the pixel values in these two images, the exception being the points highlighted in red which correspond to the darker RSL pixels in image seven.

One way to extract differences between these two images involves re-plotting the pixel values on a new set of axes. One axis would closely follow the $y = x$ line in Figure 3b, but would actually follow the direction of maximum variance in the pixel cloud. The second axis runs perpendicular to the first highlighting changes in brightness between the two images. This rotated pixel cloud is shown in Figure 3c, with the thin lines running through the cloud indicating the orientation of the original axes.

PCA changes the values of the pixels in each image by plotting the pixel data on a set of axes which capture the maximum variance in the dataset. Each pixel in the original image stack can be treated as a 9-dimensional vector since there is a value for that pixel in each of the nine images. However, because much of the pixels remain unchanged between images, there is redundancy in the stacked image data that can be reduced by plotting the pixel data on a new set of axes that capture the most important, or PCs of the image stack.

In essence, the stacked image data can be treated as a matrix of numbers, $S_{tp}$, with rows indicating the number of images in the stack ($t$) and the number of columns indicating the number of individual pixels ($p$) in the image. Determining the orientation of the PC axes on which the data is to be re-plotted involves a series matrix calculations - the first of which is the covariance of $S$.

The covariance matrix ($C$) is given by

$$C = SS^T$$

and has $t$ columns and rows (a $9 \times 9$ matrix for our example data set) and describes how closely related each image in the stack is to every other image in the stack. An Eigen decomposition of $C$ will give the direction (in image coordinate space) along which the maximum variation lies (as Eigen vectors) and the amplitude of the variation in the dataset in that direction (as Eigen values). The Eigen values of $C$ are the the roots of the characteristic polynomial equation:

$$\det(\lambda I_9 - C) = 0$$

where $I_9$ is the $9 \times 9$ identity matrix and will result in nine Eigen values ($\lambda_i$ where $i = [1,..., t = 9]$). Each of the nine Eigen values will have an associated unit length, 9-dimensional, Eigen vector ($e_i$) given by

$$e_i = N(\lambda_i I_9 - C)$$

where the right side of Equation (3) is the null space associated with $\lambda_i$. These Eigen vectors describe the axes of the PCs of the image data set. Thus, the Eigen vector associated with $\lambda_1$ (where the Eigen values have been sorted by their magnitude $\lambda_1 > \lambda_2$...) corresponds to the first PC and spans the maximum variance in the dataset. The value of the variance along $e_1$ is given by $\sqrt{\lambda_1}$.

The key step in performing the transformation from image coordinates to PC coordinates it to construct the Eigen matrix ($E$):

$$E = [e_1, \ldots, e_9]$$

whose columns contain the Eigen vectors. These vectors point in the direction of the PC axes using the image data coordinates. Thus, to perform the transformation from image to PC coordinates, the inverse of $E$ is multiplied by the image data as follows:

$$[S]_{PCA} = E^{-1}S$$
It naturally follows from Equation (5) that multiplying a pixel vector that has coordinates in PC space will convert back to the original image coordinates. This has relevance for locating the original images containing the features discovered in the PCA and is discussed in Section 3.2.

3. Results

Figure 4 shows the result of the PC transform (performed using ENVI® remote sensing image analysis software) on the stacked time-series. Because most of the variation in the images is captured in the first PC, it is colored according to the color bar given in Figure 1 with a pixel value range of 0 (black) to 256 (white). In subsequent components, the color bar shown on the lower-left of Figure 4 is used. Generally, variations covering a large spatial extent or are present in a large fraction of images that cause a big shift in pixel values will be captured in the lower PCA components. Higher components, on the other hand, will depict smaller amplitude variations occurring in fewer images (or over smaller spatial extents). Thus components 2 and 3 (Figure 4b and 4c, respectively) are dominated by variations caused by differences in lighting (sun direction and elevation angle) and possibly dust or frost deposition/erosion. RSL features are identifiable in PCA4 (Figure 4d) as elongated blue “fingers” and become more dominant and appear red in component 5 (Figure 4e). Subsequent components capture smaller variations caused either by changes in lighting, RSL shape (in PCA 8 and 9), or camera noise.

![Figure 4](image-url)

**Figure 4.** Nine components resulting from a PCA transformation of the image time-stack. PCA1 (a) uses the gray color scale used in Figure 1 due to its large variation in pixel values. The range of pixel values get progressively smaller in each PC due to reduction of data set variability. The higher components are capturing more subtle variations between the images in the stack which may be of geologic interest. In this case, we are interested in the RSL features seen in components 4, 5, 8, and 9 (d, e, h, and i).

The user can more easily identify features in the PCA components than the original images and can select those components to determine which images they correspond to. These selected components will be used in the next step in the CDPCAST method.
3.1. Constructing the PCA feature vector

Having identified PCs 5, 6, 8, and 9 as containing the feature of interest, the next objective is to identify which of the original images contain RSL. In Figure 3a, RSL were already determined to have formed in the seventh image (Figure 1g), but there may also be other images in the time-series which contain RSL. Future application of CDPCAST could potentially locate changes in much larger time-series involving tens of images without prior knowledge of which images contain changes.

Before the original images containing the feature can be identified, the user must first construct a “feature vector” whose elements correspond to each of the PCs as indicated in step 3 of Figure 2. Looking at each of the PC images in sequence, the user enters a zero into the feature vector, f, if the feature is not present. Because PCA1 is the average image of the dataset, it is neglected from this procedure as it is the reference from which changes are revealed in the subsequent PCs. Thus, f1 will always be set to zero. The first PC following PCA1 that contains the feature will be given a value of positive one. In the subsequent PCs, if the feature is present but the pixel values are reversed in sign compared to the first PC containing the feature, then a negative one is entered into the corresponding element of f.

3.2. Locating images containing feature

In practice, determining the elements of f can be ambiguous as will be discussed in Section 4. However, our goal is to get a qualitative measure of which images in the original stack are most likely to contain the feature identified in the PCs. Once the feature vector has been determined, we can quantify the potential (p) of an original image containing the feature via the following equation:

$$p = Ef\sqrt{\lambda}$$

(6)

where the elements of p correspond to the sequence of images in the original stack. Equation (6) applies weights the feature vector (f) corresponding to the variance ($\sqrt{\lambda}$) of the PC containing the feature. The result is then transformed back into the original image coordinates using the Eigen matrix (step 4 of Figure 2). The result is a vector (p) whose elements indicate the relative influence each image had in producing the feature vector with larger, positive values indicating more influence than smaller, negative values. However, there is a potential for sign reversal depending on the orientation of the PCs that are unique to each dataset, so the user should also check whether negative or positive values indicate the presence of the feature. The actual calculation of p using the Eigen matrix, feature vector, and variances for the example HiRISE image data is given below:

$$\begin{bmatrix}
2.0 \\
-5.7 \\
-0.6 \\
-3.5 \\
-3.4 \\
-3.3 \\
5.7 \\
5.1 \\
4.4
\end{bmatrix}
= \begin{bmatrix}
0.41 & -0.07 & 0.89 & 0.2 & -0.01 & 0.02 & -0.03 & -0.02 & -0.03 \\
0.36 & -0.04 & 0.01 & -0.78 & -0.34 & -0.33 & 0.14 & -0.03 & 0.03 \\
0.41 & -0.83 & -0.27 & 0.09 & 0.21 & 0.13 & 0.01 & -0.06 & 0.02 \\
0.29 & 0.25 & -0.09 & -0.26 & 0.34 & 0.2 & -0.73 & 0.17 & -0.26 \\
0.31 & 0.29 & -0.06 & -0.13 & 0.33 & 0.49 & 0.52 & 0.28 & 0.32 \\
0.3 & 0.3 & -0.14 & 0.14 & 0.36 & -0.3 & 0.05 & -0.74 & 0.14 \\
0.29 & 0.18 & -0.19 & 0.18 & -0.59 & 0.48 & 0.06 & -0.29 & -0.38 \\
0.3 & 0.17 & -0.18 & 0.32 & 0.07 & -0.5 & 0.26 & 0.45 & -0.47 \\
0.3 & 0.09 & -0.19 & 0.32 & -0.37 & -0.15 & -0.32 & 0.24 & 0.66
\end{bmatrix}
\begin{bmatrix}
0 \\
0 \\
0 \\
1 \\
-1 \\
0 \\
0 \\
1 \\
-1
\end{bmatrix}
= \begin{bmatrix}
220 \\
21 \\
16 \\
9.8 \\
6.3 \\
4.2 \\
2.9 \\
2.7 \\
2.5
\end{bmatrix}$$

(7)

4. Discussion

In order to assess the performance of the qualitative prediction made by Equations (6) and (7) regarding the presence of RSL, we take a closer look at the largest elements of p and the
Figure 5. Images which contain RSL. Image (a) and cropped sections of images (f), (g), (h), and (i) from Figure 4 are reproduced here to highlight those images which contain RSL. Other images do not have RSL. The inset gives the “feature potential” value, $p$, for all images in the stack. The solid line gives $p$ using $f_8 = 1$ as shown in Equation (7) which gives good results for all images except (f). The dashed line sets $f_8 = -1$ and results in better indication of which images contain RSL.

original images in our dataset. In Equation (7) the seventh element of $p$ contains the highest value. Therefore the seventh image in the original stack is correctly predicted to contain RSL (Figure 3a or 5g). Elements eight and nine of $p$ also have large values relative to the other elements of $p$ and those images also contain RSL. A zoomed-in portion of these images (g, h, and i from Figure 4) are reproduced in Figure 5 to illustrate the presence of RSL.

The fourth largest value of $p$ is $p_1 = 2.0$ and is associated with image “a” shown in Figures 4a and 5a. Although less distinct, incipient RSL can be identified downslope of the steep rim of Palikir Crater (arrows in Figure 5a). Using the four largest values of $p$, the method outlined
here is successful in not only predicting the presence of RSL, but also providing an indication of the RSL distinctness. However, there is an image for which this method was not successful. Image “f” in Figure 5 also exhibits incipient RSL, but has low predictive score ($p_6 = -3.3$). The reason for this can be identified by looking at the matrix multiplication shown in Equation (7). During the calculation of $p_6$, the sixth row of $E$ is multiplied element-wise with the two vectors furthest to the right and each product is summed to give $p_6$. Looking at the sixth row of $E$, one notices that the eighth element most greatly affects the value of $p_6$. Stated differently, the sixth image in the original stack is captured mostly by the eighth PC. Taking a closer look at PCA8 (Figure 4h) and comparing it to the sixth image (Figure 1f or 5f) reveals that the incipient RSL locations in image six are colored red in PCA8 - suggesting that the value for $f_8$ should be negative instead of positive. This example illustrates the potential for ambiguity in defining $f$ because there are also parts of PCA8 with blue colored RSL shapes.

The inset plot in Figure 5 indicates the value of the elements of $p$ for images (a) through (i) from Figure 1 using a $f_8$ value of one (solid line) or negative one (dashed line). The $p$ value for image “f” is most affected by this change. Given the observation of incipient RSL in image “f,” a value of negative one for $f_8$ provides a better agreement with the presence of RSL. Alternatively, if $f_8$ were set to zero, the elements of $p$ would take a value halfway between the two lines shown in the inset plot in Figure 5.

5. Conclusions
Determining which images in the time-series stack that are responsible for the variations found in the PCA analysis is the potentially unreliable part of the CDPCAST method, and requires some caution as shown in Section 4. The method is reliable in predicting which images contain the most prominent RSL, but images containing less distinct manifestations of the feature of interest may not be picked out by Equation (6). Therefore the values of the elements of $p$ should not be interpreted as definitive predictions of a feature’s presence in an image, but merely point the user toward the most likely candidates.

Optimizing change detection for the rapidly growing Martian remote sensing dataset will further our understanding of ongoing geologic processes and take full advantage of the spacecraft imagery that has been collected. The method proposed here applies an existing image processing technique (PCA) to stacked images in a time series to enhance the change detection processes. PCA is well-suited for change detection in high-resolution, single-band imagery of planetary surfaces when they are stacked into a multi-temporal image file. Although variations in lighting between images does cause significant variation between images, the higher PCA components help to identify lower-amplitude brightness changes which may result from ongoing geologic processes (Figure 4h) which might otherwise be difficult to detect. Future development of the CDPCAST method together with automation of image preprocessing steps will greatly enhance the planetary science community’s ability to monitor surface changes on terrestrial Solar System objects.

References
[1] Singh A 1989 *International Journal of Remote Sensing* **10** 989–1003
[2] Mas J F 1999 *International Journal of Remote Sensing* **20** 139–152
[3] Brewer C K, Winne J C, Redmond R L, Opitz D W and Mangrich M V 2005 *Photogrammetric Engineering & Remote Sensing* **71** 1311–1320
[4] Malin M C and Edgett K S 2000 *Science* **288** 2330–2335
[5] McEwen A S, Ojha L, Dundas C M, Mattson S S, Byrne S, Wray J J, Cull S C, Murchie S L, Thomas N and Gulick V C 2011 *Science* **333** 740–743 ISSN 0036-8075 (Preprint [http://science.sciencemag.org/content/333/6043/740.full.pdf](http://science.sciencemag.org/content/333/6043/740.full.pdf)) URL [http://science.sciencemag.org/content/333/6043/740](http://science.sciencemag.org/content/333/6043/740)
[6] Sullivan R, Thomas P, Veverka J, Malin M and Edgett K S 2001 *Journal of Geophysical Research: Planets* **106** 23607–23633 ISSN 2156-2202 URL [http://dx.doi.org/10.1029/2000JE001296](http://dx.doi.org/10.1029/2000JE001296)
[7] Ojha L, McEwen A, Dundas C, Byrne S, Mattson S, Wray J, Masse M and Schaefer E 2014 *Icarus* **231** 365 – 376 ISSN 0019-1035 URL [http://www.sciencedirect.com/science/article/pii/S0019103513005393](http://www.sciencedirect.com/science/article/pii/S0019103513005393)

[8] Daubar I, McEwen A, Byrne S, Kennedy M and Ivanov B 2013 *Icarus* **225** 506 – 516 ISSN 0019-1035 URL [http://www.sciencedirect.com/science/article/pii/S0019103513001693](http://www.sciencedirect.com/science/article/pii/S0019103513001693)

[9] Bourke M, Edgett K and Cantor B 2008 *Geomorphology* **94** 247–255

[10] Thomas P, Malin M, James P, Cantor B, Williams R and Giersch P 2005 *Icarus* **174** 535 – 559 ISSN 0019-1035 mars Polar Science III URL [http://www.sciencedirect.com/science/article/pii/S0019103504003203](http://www.sciencedirect.com/science/article/pii/S0019103504003203)

[11] Diniega S, Byrne S, Dundas C M and McEwen A 2009 *AGU Fall Meeting Abstracts*

[12] Di K, Liu Y, Hu W, Yue Z and Liu Z 2014 *IOP Conference Series: Earth and Environmental Science* **17** 012015 URL [http://stacks.iop.org/1755-1315/17/i=1/a=012015](http://stacks.iop.org/1755-1315/17/i=1/a=012015)

[13] Scheffler D, Hollstein A, Diedrich H, Segl K and Hostert P 2017 *Remote Sensing* **9** ISSN 2072-4292 URL [http://www.mdpi.com/2072-4292/9/7/676](http://www.mdpi.com/2072-4292/9/7/676)

[14] HiRISE Operations Center Gullies in palikir crater [https://www.uahirise.org/dtm/dtm.php?ID=PSP_005943_1380](https://www.uahirise.org/dtm/dtm.php?ID=PSP_005943_1380) accessed: 2017-07-03

[15] Byrne G F, Crapper P F and Mayo K K 1980 *Remote Sensing of Environment* **10** 175–184

[16] Ingebritsen S E and Lyon R J P 1985 *International Journal of Remote Sensing* **6** 687–696

[17] Rokni K, Ahmad A, Selamat A and Hazini S 2014 *Remote Sensing* **6** 4173–4189