

Using Spectral Decomposition to Detect Dirty Solar Panels and Minimize Impact on Energy Production

¹Ernesto Zamora Ramos, ²Suzanna Ho and ³Evangelos A. Yfantis

Computer Science Department, University of Nevada, Las Vegas, United States;

¹zamorara@unlv.nevada.edu; ²hos4@unlv.nevada.edu; ³yfantis@cs.unlv.edu

ABSTRACT

Dirt and dust deposits on the surface of a solar panel array obstruct the amount of light that can reach the photovoltaic cells, reducing the amount of electricity produced. Solar panels are cleaned when the energy drop has already occurred and is detected. This work presents an algorithm designed to detect dirty solar panels. It is based on the spectral decomposition of the scattered light reflected off the panels' surface by analyzing color images of the surface obtained using digital cameras. It applies the statistical classification method of Mahalanobis distance to separate images where it detects the excess reflected light, classifying them as having a high probability of representing dirty solar panels. It aims to minimize the loss of energy by warning solar plants operators to clean panels before the energy drop becomes significant.

Keywords: Solar Panel, Photovoltaic Cell, Pattern Recognition, Mahalanobis Distance, Classifier.

1 Introduction

Solar power plants are currently growing in number across the globe. They are a source of renewable, clean energy that can be used to significantly reduce the ecological impact and increase the efficiency of production of electric energy.

Solar power plants incorporate large arrays of solar panels. However, today, many individuals have access to solar panels that can be used to produce enough electricity to power a house. Research in the area is abundant right now in pursuit of more efficient ways to collect the solar energy.

Solar panels are collections of interconnected solar cells (also called photovoltaic cells) that absorb the energy of incident light, converting it into an electric current through a phenomenon called "photovoltaic effect."

The photovoltaic (PV) effect is directly related to the photoelectric effect. In summary, the electrons on certain materials can be excited by incident light. Semiconductor materials, such as silicon, are usually used. When an electron in the valence band of a crystal's atom absorbs enough energy from incident photons, it jumps to the conductive band and becomes free, ionizing the source atom with a positive charge. Under the presence of an electric field, the separated electrons and ions are attracted to the opposite charged plates, creating an electromotive force. If a circuit is connected to these plates, an electric current flows. As light continues to excite the material, the ionization is maintained and the electricity continues to flow [1].

Clearly, if more light reaches and gets absorbed by a PV cell, then, more atoms get ionized in the crystal and more electrons become free. As a result, the potential of the electromotive force created by the

separation of more negative and positive charges increases as well as the electricity flowing through the circuit.

The basic structure of a PV cell is designed to allow the maximum light possible to reach the excitable material, maximize the absorption of photons and minimize reflection. Solar cells rely on a layer of antireflection coating on the front of the cell to reduce reflection of the incident light. On simple cells, light rays enter through the front surface and, if not absorbed, leave through the rear. More sophisticated designs extend the path of light inside the cell to improve absorption through a process called "light trapping" [2].

Sunlight is comprised of ultraviolet, visible, and infrared light. While red light has the longest wavelength in the visible spectrum, infrared light is even longer. Therefore, infrared and red light, both having the larger wavelengths compared to the other components of the incident light, penetrate the glass more readily and produce the most amount of electricity. Meanwhile, blue and violet light suffer the most absorption by the coating of the cell, not the cell itself, and offer little contribution to the production of electric energy in comparison.

As we said earlier, the amount of energy generated by PV cells is directly proportional to the amount of light absorbed. And the more light directly illuminating the cell, more photons reach the material, and more light can be absorbed.

Now, the amount of light striking the solar cells on a solar panel array is dependent on many factors, including the month of the year, day of the month, time of day, weather conditions, and other location-dependent circumstances. Most weather conditions that can limit the amount of light, and thus, the amount of electricity generated, cannot be avoided. Other causes, however, such as light obstruction due to other objects, broken cells and overall cleanliness or dirtiness of the panels, can be dealt with in order to maximize the amount of light reaching the solar cells.

Many solar power plants are established in areas with arid climates due to the low humidity and clear skies year round. Dust and sand storms are common in these climates and the dust gradually settles on the glass surface of solar panels, slowly decreasing the amount of light that reaches the solar cells. The loss of light energy depends on the amount, size, and chemical composition of the dust [3][4]. In terms of time, trees are sparse in arid climates, and during migration in the fall and spring, birds use solar farms as rest areas; therefore, the solar panels become dirty with bird excrement. In general, this is a problem throughout the entire year. Bird droppings are worse than dust, because no light passes through them.

Deposits on a dirty panel reflect, scatter and obstruct the incident light, reducing the amount of photons that can penetrate through and reach the PV cells, consequently decreasing the amount of electricity produced [3].

In this paper, we devise an algorithm specifically to detect dirty solar panels. It is based on the spectral decomposition of the scattered light reflected off the panels' surface by analyzing color images of the surface obtained using digital cameras. It applies statistical classification methods to separate images where it detects the excess reflected light, classifying them as having a high probability of representing dirty solar panels.

It is important to know when a panel is dirty, so that it can be cleaned promptly to minimize the loss of energy. This study aims to warn solar power operators about panels that are becoming dirty before the loss of energy becomes noticeable.

2 Classification Algorithm

The classification algorithm presented here involves sampling, determining the classification vector, and developing, training, and testing the classifier. Our goal is to use videos of solar panel surfaces captured by cameras from a solar panel site and classify the solar panels as clean or dirty. All of the samples we collected are images of solar panels captured by a digital camera [10].

2.1 Observation

By observation, we noted that the surface of solar panels looks dark, due to the limited light reflected off it and the amount of light trapped inside the cells and absorbed by the cells' material. In contrast, a dirty panel's surface looks lighter, because more light is reflected off the surface and scattered by the deposits.

We can see that the corresponding histograms of the tristimulus values (color channels red, green and blue) of images of clean and dirty panels support our hypothesis. See sample in figure 1.

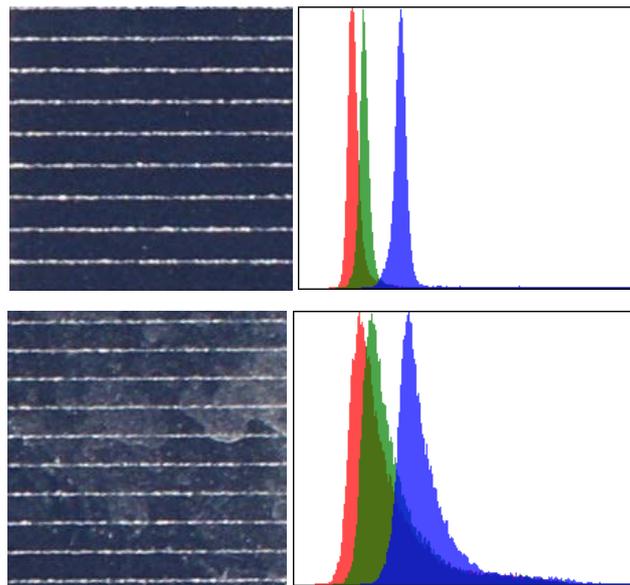


Figure 1. Comparison of a sample for a clean panel with a sample for a dirty panel. Top-left: clean panel; top-right: histogram of the tristimulus values for the clean panel; bottom-left: dirty panel; bottom-right: histogram of the tristimulus values for the dirty panel. Notice that the mean intensity for each color channel is higher for the dirty panel. The variance for each color channel is larger for the dirty panel, indicated by the wider spread of color intensities around the mean intensities.

In general, the mean intensity for each color channel is higher for the dirty panel, indicating a brighter image. This is consistent with our premise that deposits on a dirty panel reflect more light than a clean panel. On the same note, the variance for each color channel is larger for the dirty panel than for the clean panel.

2.2 Sampling

A static digital camera was pointed at a fixed point on a solar panel. The camera capture was utilized to obtain sample data for the classifier development, testing and experimentation phases.

As stated earlier, the intensity on the incident light varies with specific day of the year and time of day. So, for the sampling, we decided to collect data during seven consecutive days, only inside the same timeframe every day, making sure that weather conditions such as wind speed, clouds and temperature remained relatively similar during each collection.

First, we used captured images of a panel when it was clean. Then we artificially applied dirt, dust, or both randomly until we reached the desired power drop to consider the panel dirty, and retrieved the captures. The camera images were divided into samples of dimensions 200×200 pixels. Each obtained image was always cropped in the same way. The goal was to use crops from the same panel, and crops from a combination of panels with similar characteristics. We settled on three groups of training data, where a group contains one clean sample set and one dirty sample set.

1. The First group contains data from the same panel. Each set has 12 samples.
2. The second group builds upon the First group by incorporating data from another panel of similar characteristics, i.e. same PV cell structure. Each set has 20 samples.
3. The third group does not build upon the First and second group. Instead, it incorporates data from two panels of similar characteristics, but with a lighter shade of blue. Each set has 19 samples.

2.3 Classifier development

The goal here is to gauge the similarity of an unknown sample to a known distribution by comparing the set of conditions of the unknown sample to the ideal set of conditions of the known distribution. We can do so by computing the Mahalanobis Distance, which is the relative measure of the data point's distance from a common point [5][6].

Consider a sample set of K clean panels, another sample set of D dirty panels, and an unknown sample x from an arbitrary panel. For this panel, we compute the averages of the tristimulus values (red, green and blue), and the variance–covariance matrix of x . The averages constitute a classification vector. If the panel is clean, then its vector belongs to the clean class; otherwise, it belongs to the dirty class. All the clean classification vectors have a grand average, which is another vector, and they have their own variance–covariance matrix. The grand average of the clean classification vectors defines the centroid, or center of gravity, of the clean class. The variance–covariance matrix of the vectors of the clean class defines how close the vectors are to the centroid. Together, the variance–covariance matrix and the centroid make up the parameters of the classifier [9].

The following theorems describe the statistical basis behind the math of our classifier.

2.3.1 Theorem

Let $\overline{x_{cm}}$ be a classification vector from the clean class c of images $m = 1, 2, \dots, K$ with centroid $\overline{x_c}$ and variance–covariance matrix Σ_c , then the mean of the vector $\overline{x_{cm}} - \overline{x_c}$ is zero, and the variance–covariance matrix of $\overline{x_{cm}} - \overline{x_c}$ is $\frac{K-1}{K}\Sigma_c$. Furthermore, the Mahalanobis distance from $\overline{x_c}$ to $\overline{x_{cm}}$ is

$$d_{cm}^2 = (\overline{x_{cm}} - \overline{x_c})^T \left(\frac{K}{K-1} \right) \Sigma_c^{-1} (\overline{x_{cm}} - \overline{x_c}) \quad (1)$$

Alternatively, the Mahalanobis distance from $\overline{x_d}$ to $\overline{x_{dm}}$ where d is the dirty class is

$$d_{dm}^2 = (\overline{x_{dm}} - \overline{x_d})^T \left(\frac{D}{D-1} \right) \Sigma_d^{-1} (\overline{x_{dm}} - \overline{x_d}) \quad (2)$$

Proof: Let $\mu = E(\overline{x_{cm}})$, then:

$$\begin{aligned} E(\overline{\overline{x}}) &= E\left[\frac{\sum_{m=1}^K \overline{x_{cm}}}{K}\right] \\ &= \frac{1}{K} \sum_{m=1}^K E[\overline{x_{cm}}] \\ &= \frac{1}{K} K \mu \\ &= \mu \end{aligned}$$

Thus $E(\overline{x_{cm}} - \overline{x_c}) = E(\overline{x_{cm}}) - E(\overline{x_c}) = \mu - \mu = 0$, where μ , by definition, is a column vector with three components, namely, the means of the red, green and blue color components.

The variance–covariance matrix Σ of the classification vectors is a 3×3 positive, definite, symmetric matrix and is denoted by:

$$\Sigma = E(\overline{x_{cm}} - \mu)'(\overline{x_{cm}} - \mu)$$

The variance–covariance matrix of the centroid $\overline{x_c}$ is:

$$\begin{aligned} E(\overline{\overline{x_c}} - \mu)'(\overline{\overline{x_c}} - \mu) &= E\left[\frac{\sum_{i=1}^K x_{ci}}{K} - \mu\right]' \left[\frac{\sum_{j=1}^K x_{cj}}{K} - \mu\right] \\ &= \frac{1}{K^2} E\left[\left(\sum_{i=1}^K x_{ci} - \mu\right)' \left(\sum_{j=1}^K x_{cj} - \mu\right)\right] \\ &= \frac{1}{K^2} \left\{ \sum_{i=1}^K E(x_{ci} - \mu)'(x_{ci} - \mu) + \sum_{\substack{i=1 \\ i \neq j}}^K \sum_{\substack{j=1 \\ i \neq j}}^K E(x_{ci} - \mu)(x_{cj} - \mu) \right\} \\ &= \frac{1}{K^2} K \Sigma + \sum_{\substack{i=1 \\ i \neq j}}^K \sum_{\substack{j=1 \\ i \neq j}}^K 0 \\ &= \frac{\Sigma}{K} \end{aligned}$$

The variance–covariance matrix of $\overline{x_{cm}} - \overline{x_c}$ is:

$$\begin{aligned} E(\overline{x_{cm}} - \overline{x_c})'(\overline{x_{cm}} - \overline{x_c}) &= E[(\overline{x_{cm}} - \mu)(\overline{x_c} - \mu)]'[(\overline{x_{cm}} - \mu)(\overline{x_c} - \mu)] \\ &= E(\overline{x_{cm}} - \mu)'(\overline{x_{cm}} - \mu) - E(\overline{x_{cm}} - \mu)'(\overline{x_c} - \mu) + \\ &\quad - E(\overline{x_c} - \mu)'(\overline{x_{cm}} - \mu) + E(\overline{x_c} - \mu)'(\overline{x_c} - \mu) \\ &= \Sigma - \frac{1}{K} \Sigma - \frac{1}{K} \Sigma + \frac{1}{K} \Sigma \\ &= \frac{K-1}{K} \Sigma \end{aligned}$$

From this result we infer that the Mahalanobis distance from the centroid \bar{x}_c to vector \bar{x}_{cm} is:

$$d_{cm}^2 = (\bar{x}_{cm} - \bar{x}_c)^T \left(\frac{K}{K-1} \right) \Sigma_c^{-1} (\bar{x}_{cm} - \bar{x}_c) \quad \square$$

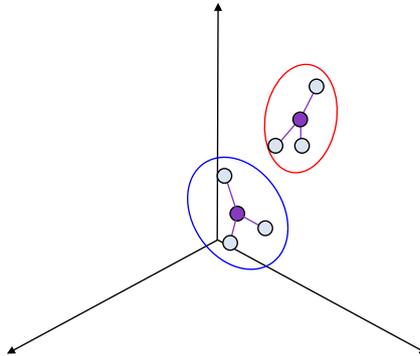


Figure 2. Illustration for theorem 2.3.1. The clean space is blue, and any points within that space are clean classification vectors. The dirty space is red, and any points within that space are dirty classification vectors. The centroids are purple.

Theorem 2.3.1 describes how we can compute the Mahalanobis distance of the m -th vector of a class to the centroid of that class, and it is illustrated in figure 2. The clean classification vectors form a space with the centroid \bar{x}_c , which is a subspace of the 3-D space defined by the (R, G, B) values, and its shape is similar to an ellipsoid, where each axis has a different size. Every classification vector \bar{x}_{cm} of the clean class that does not belong to the intersection between the clean class and the dirty class has a smaller distance from the centroid of the clean space than from the centroid of the dirty space. Similarly, the dirty classification vectors form a space with the centroid \bar{x}_d , and its subspace has the same characteristics as the clean subspace. Every classification vector \bar{x}_{dm} of the dirty class that does not belong to the intersection of the clean class with the dirty class, has a smaller distance from the centroid of the dirty space than from the centroid of the clean space.

Classification vectors belonging to the intersection of the spaces could have larger distances from the corresponding centroid than from the centroid of the other class. The issue of intersections is addressed later in theorem 2.4.1.

The following theorems describe the statistical basis behind the math of our classifier.

2.3.2 Theorem

Let \bar{x}_{cm} , $m = 1, 2, \dots, K$, be a classification of the clean space with mean vector μ_c , variance–covariance matrix Σ_c , and centroid \bar{x}_c . Let \bar{x} be a new classification vector. If \bar{x} belongs to the clean space, then $E(\bar{x} - \bar{x}_c) = 0$, and the variance–covariance matrix of \bar{x} is $\frac{K-K}{K} \Sigma_c$. Furthermore, the Mahalanobis distance between \bar{x} and \bar{x}_c is:

$$d_1^2 = (\bar{x} - \bar{x}_c)' \left(\frac{K}{K+1} \right) \Sigma_c^{-1} (\bar{x} - \bar{x}_c).$$

Proof: Let $E(\bar{x}) = \mu_c$, then, from theorem 2.3.1, $E(\bar{x}_c) = \mu_c$ and $E(\bar{x} - \bar{x}_c) = \mu_c - \mu_c = 0$.

The variance–covariance matrix of $\bar{x} - \bar{x}_c$ is:

$$\begin{aligned} E(\bar{x} - \bar{x}_c)'(\bar{x} - \bar{x}_c) &= E[(\bar{x} - \mu_c)(\bar{x}_c - \mu_c)]'[(\bar{x} - \mu_c)(\bar{x}_c - \mu_c)] \\ &= E(\bar{x} - \mu_c)'(\bar{x} - \mu_c) - E(\bar{x} - \mu_c)'(\bar{x}_c - \mu_c) + \\ &\quad - E(\bar{x}_c - \mu_c)'(\bar{x} - \mu_c) + E(\bar{x}_c - \mu_c)'(\bar{x}_c - \mu_c) \\ &= \Sigma_c - 0 - 0 + \frac{1}{K} \Sigma \\ &= \frac{K+1}{K} \Sigma \end{aligned}$$

From this result we infer that the Mahalanobis distance from the centroid \bar{x}_c to vector \bar{x} is:

$$d_1^2 = (\bar{x} - \bar{x}_c)' \left(\frac{K}{K+1} \right) \Sigma_c^{-1} (\bar{x} - \bar{x}_c) \quad \square$$

Theorem 2.3.2 describes how we can classify a new, arbitrary vector. The idea has been summarized on figure 3.

2.4 Misclassification Error

In the previous section, we designed our classifier based on the Mahalanobis distance. In summary, we trained the classifier using the sample data that has been classified manually beforehand. The training defines two ellipsoids in space, each representing the respective class of clean panel or dirty panel. When a new, arbitrary sample is to be classified by the trained classifier, we extract the classification vector for the sample and compute the Mahalanobis distance from the new vector to each centroid. Whichever distance is smaller, we say that the new vector, and thus, the new sample belong to that class. This classification decision is justified by theorem 2.4.1.

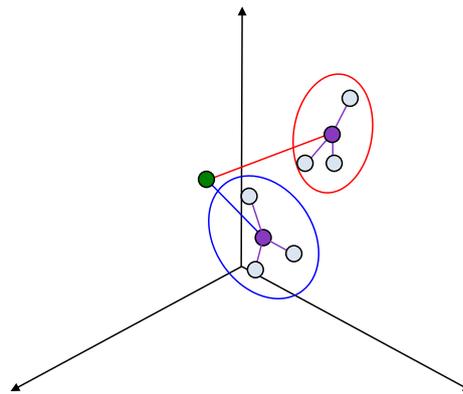


Figure 3. Illustration for theorem 2.3.2. The blue ellipsoid represents the clean class space, and the red ellipsoid represents the dirty class space. The purple points are the centroids of each class. The green point is the unknown vector. It belongs to the clean class since its relative distance to the clean space's centroid is smaller than its relative distance to the dirty space's centroid.

2.4.1 Theorem

Let \bar{x}_{cm} , where $m = 1, 2, \dots, K$, be a classification vector of the clean space with mean vector μ_c , variance–covariance matrix Σ_c , and centroid \bar{x}_c . Let \bar{x}_{dm} , where $m = 1, 2, \dots, D$, be a classification

vector of the dirty space with mean vector μ_d , variance-covariance matrix Σ_d which is not significantly different from Σ_c , and centroid \bar{x}_d . A better estimate of the variance for both classes is

$\Sigma = \frac{(K-1)\Sigma_c + (D-1)\Sigma_d}{K+D-2}$. Let \bar{x} be a new classification vector. The Mahalanobis distance from \bar{x} to \bar{x}_c is:

$$d_1^2 = (\bar{x} - \bar{x}_c)' \left(\frac{K}{K+1} \right) \Sigma (\bar{x} - \bar{x}_c)$$

and from \bar{x} to \bar{x}_d is

$$d_2^2 = (\bar{x} - \bar{x}_d)' \left(\frac{D}{D+1} \right) \Sigma (\bar{x} - \bar{x}_d)$$

If $d_1^2 < d_2^2$, then \bar{x} is more likely to be in the clean space, and if $d_1^2 > d_2^2$, then \bar{x} is more likely to be in the dirty space.

Proof: By the central limit theorem, the means for large number of samples are normally distributed [7]. Therefore, if a vector v represents the three color channels for a pixel of an image, and \bar{v} is the mean of the channels for all the pixels in the image, then, the probability distribution for the means of a large number of images, $P(\bar{v})$, is normal.

Consider, then, the following functions expressing the probability of \bar{x} belonging to the clean and dirty space respectively:

$$P_c(\bar{x}) = \frac{1}{(2\pi)^{\frac{3}{2}} \left(\frac{K+1}{K} \right)^{\frac{1}{2}} |\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2} (\bar{x} - \bar{x}_c)' \frac{K}{K+1} \Sigma^{-1} (\bar{x} - \bar{x}_c)}$$

$$P_d(\bar{x}) = \frac{1}{(2\pi)^{\frac{3}{2}} \left(\frac{D+1}{D} \right)^{\frac{1}{2}} |\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2} (\bar{x} - \bar{x}_d)' \frac{D}{D+1} \Sigma^{-1} (\bar{x} - \bar{x}_d)}$$

If we assume that $d_1^2 < d_2^2$, then, $P_c(\bar{x}) > P_d(\bar{x})$ since $\left(\frac{K+1}{K} \right)^{\frac{1}{2}} \approx \left(\frac{D+1}{D} \right)^{\frac{1}{2}} \approx 1$, which implies that \bar{x} is more likely to belong to the clean class. □

Theorem 2.4.1 is the justification of the classification decision. It is summarized in figure 4. But as it can be observed in figure 4, if the spaces overlap, there is the possibility of misclassification.

The power of this classifier is a function of its ability to correctly classify samples. As we saw, there is a chance of incorrectly classifying some new samples; therefore, the higher the probability of the classifier to correctly classify new samples, the more powerful it is. We computed the misclassification

error using the Jackknife approach in order to obtain a quantitative measure of the power of our classifier.

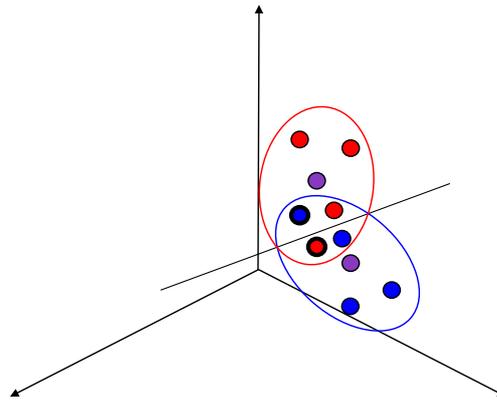


Figure 4. Illustration for theorem 2.4.1. The blue ellipsoid represents the clean class space, and the red ellipsoid represents the dirty class space. The purple points are the centroids of each class. Elements in blue are known to belong to the clean class. Elements in red are known to belong to the dirty class. The black line represents the decision plane created by theorem 2.4.1. The decision plane passes through the adjacency points of the two ellipsoids. All elements (in the case of this representation) above the plane will be classified as clean, while the rest are classified as dirty. All elements will be correctly classified by the theorem, except those with bold outline in the intersection.

2.4.2 Jackknifing

We train the classifier as follows: we use the data from each class to compute the classification vectors for that class, then we use these computed vectors and the parameters obtained by the classifier to define the class. Then we take a vector v that already belongs to a class c , re-compute c without v , and then we use our classification algorithm to classify v . We repeat this process for each vector from each class, and we estimate the probability of correct classification as the ratio of the number of vectors classified correctly over the total number of classification vectors (see equation 3). This is known as the Jackknifing approach.

$$Accuracy = \frac{TN + TP}{TN + TP + FN + FP} \quad (3)$$

where TN (true negative)/ FN (false negative) are the number of samples correctly/incorrectly classified as clean, and TP (true positive)/ FP (false positive) are the number of samples correctly/incorrectly classified as dirty.

This approach allowed us to find outliers; for example, clean samples with a lighter blue color might be considered dirty because their data behaves similarly to a dirty panel, which means we should not include them with the other clean samples. The more the overlap there is between the set of characteristics that distinguish samples from clean and dirty, the larger the probability of a classification error [8-10]. This was another important concept that was taken care of by the testing method, because it allowed us to filter out samples that barely made the cut; for example, a clean sample could have been correctly classified, but it could have easily been misclassified (i.e. became a false positive) if the dirty and clean samples were slightly different, because sometimes taking out, adding in, or substituting an image upsets the balance. However, we also needed to pay attention to samples that were far from the threshold, known as outliers, because they were too good to be a true negative or positive.

Once the classifier is trained, it continues learning from experience with new samples. If \bar{x} represents the classification vector for a new sample and it is classified correctly to class c , then c 's space is recalculated, which means a new centroid and variance–covariance matrix is computed.

3 Experimental Results

We used our sample sets as described in section 3.2 to determine the accuracy of our classifier. Table 1 lists the results of the groups of samples computed by using equation 3, as well as the misclassification error, which is the number of false negatives and false positives divided by the total number of classification vectors. It is no surprise that Group 1 has the best accuracy, because it is the strictest. The other two groups have favorable results, considering the data is not as good as it could be.

Table 1. Results of equation 3 from applying Jackknife test on all three groups of training data.

Group	TN	FN	TP	FP	Accuracy (%)	Misclassification Error
1	12	0	12	0	100	0
2	17	3	19	1	90	0.10
3	17	2	19	0	94.4	0.0526

Figures 5, 6, and 7 illustrate the distribution of samples in Group 1, Group 2, and Group 3, respectively. As expected, the graphs do justice to the results listed in table 1, with Group 1 having no overlap due to its 100% accuracy, and groups 2 and 3 having some overlap. The images also reveal unexpected results. Although Group 3 is the least restrictive, its clean and dirty samples are closest to their centroid than the samples of the other two groups. Although Group 1 is the most restrictive, its dirty samples are furthest from their centroid than the other two groups; however, it is the only group where not all pairs of principal components are parallel, specifically the green and blue graph.

Sample images and data collected are available from the authors upon request.

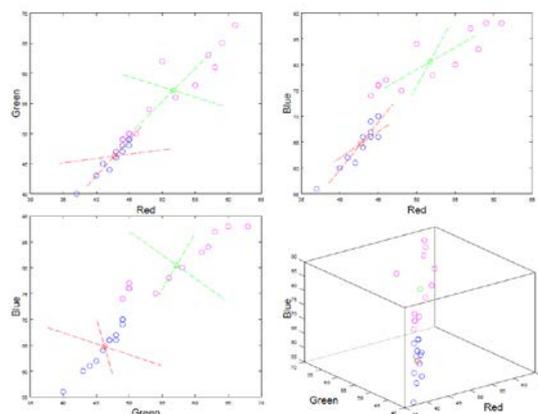


Figure 5. Distribution of Group 1. The blue and magenta circles are clean and dirty samples, respectively. The red and green circle are centroids of the clean and dirty samples, respectively.

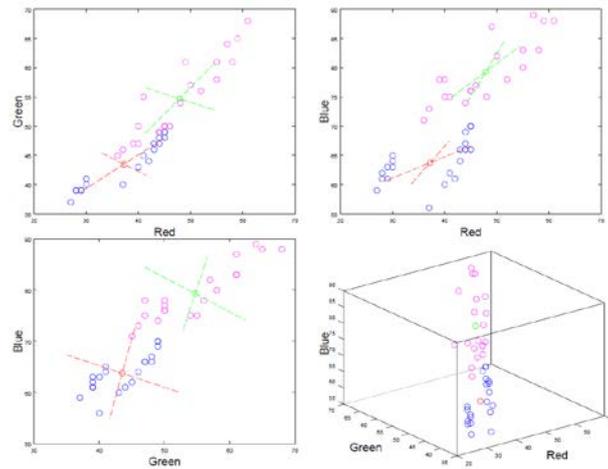


Figure 6. Distribution of Group 2. The blue and magenta circles are clean and dirty samples, respectively. The red and green circle are centroids of the clean and dirty samples, respectively.

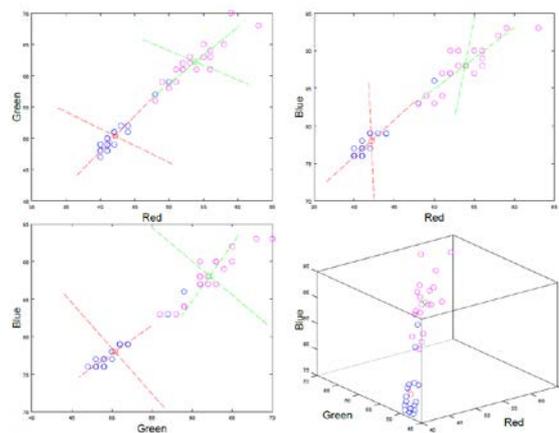


Figure 7. Distribution of Group 3. The blue and magenta circles are clean and dirty samples, respectively. The red and green circle are centroids of the clean and dirty samples, respectively.

4 Conclusions

We presented an algorithm and the supporting mathematical basis to automatically analyze and classify digital images of solar panel surfaces as clean or dirty. The algorithm takes advantage of the observation that photovoltaic cells absorb light, reflecting and scattering very little as a result. Deposits on dirty panels block some light from reaching the photovoltaic cells, reflecting the light back, making the overall image of the surface brighter. Our experimental results show that, if well trained, our algorithm can efficiently and successfully classify dirty solar panels with accuracy above 90%.

Dirty panels are currently detected by inspection or when a power drop occurs. Our algorithm aims to assist solar power operators by warning them about dirty panels before a significant power drop happens, thus minimizing the loss of energy.

5 Acknowledgement

This research is based upon work supported by National Science Foundation under grant number IIA-1301726.

The authors would like to thank Dr. Boehm, professor of Mechanical Engineering at University of Nevada, Las Vegas for allowing access to his solar laboratory and the data obtained from the instruments, and special thanks to Dr. Batista, professor of Civil and Environmental Engineering for her support.

REFERENCES

- [1]. K. W. Böer, *Solar Cells*. (2015) Chemistry Explained [Online]. Available: <http://www.chemistryexplained.com/Ru-Sp/Solar-Cells.html>
- [2]. A. J. McEvoy, T. Markvart and L. Castañer, *Principles of Solar Cell Operation*. Practical handbook of photovoltaics: fundamentals and applications, 2nd ed., Academic Press, Elsevier, 2012, pp. 7–30.
- [3]. C. P. Ryan, F. Vignola, and D. K. McDaniels, *Solar cell arrays: Degradation due to dirt*. Proceedings of 1989 Annual Conference of The American Solar Energy Society, 1989, pp. 234-237.
- [4]. M. K. Mazumder *et al.*, *Solar Panel Obscuration by Dust and its Mitigation in the Martian Atmosphere*. Particles on Surfaces 9: Detection, Adhesion and Removal, 2006, pp. 1-29.
- [5]. G. J. McLachlan, *Mahalanobis Distance*. Resonance, Jun. 1999, pp. 20-26.
- [6]. E. Zamora Ramos, *Using Image Processing Techniques to Estimate the Air Quality*. McNair Scholars Research Journal, UNLV chapter, 6th ed., 2012, pp 189-194.
- [7]. E. W. Weisstein, *Central Limit Theorem*. [Online]. Available: <http://mathworld.wolfram.com/CentralLimitTheorem.html>. [Accessed 15 Nov 2015].
- [8]. R. Maitra, *Discrimination and Classification – Introduction*. [Online]. Available: <http://www.public.iastate.edu/~maitra/stat501/lectures/Classification-I.pdf>, 2012. [Accessed 30 May 2014].
- [9]. E. A. Yfantis *et al.*, *Pollution Detection in Urban Areas Using the Existing Camera Networks*. International Journal of Multimedia Technology, 2013, Vol. 3, No. 3, pp. 98-102.
- [10]. E. A. Yfantis and A. Fayed, *A Camera System for Detecting Dust and Other Deposits on Solar Panels*. Journal Of Advances in Image and Video Processing, 2014, Vol. 2, No. 5, pp. 1-10.