Image-Based State Recognition for Disconnect Switches in Electric Power Distribution Substations

Bogdan T. Nassu
Federal University of Technology — Parana, Brazil
Email: btnassu@utfpr.edu.br

Lourival Lippmann Jr., Bruno Marchesi, Amanda Canestraro and Rafael Wagner
Institutos Lactec, Brazil

Vanderlei Zarnicinski
Companhia Paranaense de Energia, Brazil

***Abstract*—Knowing the state of the disconnect switches in a power distribution substation is important to avoid accidents, damaged equipment, and service interruptions. This information is usually provided by human operators, who can commit errors because of the cluttered environment, bad weather or lighting conditions, or lack of attention. In this paper, we introduce an approach for determining the state of each switch in a substation, based on images captured by regular pan-tilt-zoom surveillance cameras. The proposed approach includes noise reduction, image registration using phase correlation, and classification using a convolutional neural network and a support vector machine fed with gradient-based descriptors. By combining information given in an initial labeling stage with image processing techniques to reduce variations in viewpoint, our approach achieved 100% accuracy on experiments performed at a real substation over multiple days. We also show how modifications to the standard phase correlation image registration algorithm can make it more robust to lighting variations, and how SIFT (Scale-Invariant Feature Transform) descriptors can be made more robust in scenarios where the relevant objects may be brighter or darker than the background.**

I. Introduction

Distribution substations are the part of the electric power delivery system responsible for converting the high transmission voltage to medium voltage, and directing electric power to distribution lines. Disconnect switches (also called disconnecting switches or disconnectors) are used in substations to reconfigure the network and isolate equipment for maintenance. A medium-sized substation can have dozens of disconnect switches, and a single city can have dozens of substations. That makes remotely operated, automated switches not always economically viable — given the conditions inside a substation, they require proper insulation and frequent maintenance, raising costs when compared to manual switches, which are operated by a human using a long pole. Figure 1 shows a disconnect switch.

Knowing the state (open or closed) of each switch in a substation is important to avoid accidents, damaged equipment, and service interruptions. As most switches are manually operated, this information is usually provided in reports by humans, who can commit errors due to lack of attention, bad weather or lighting conditions (rain, fog, nighttime), or simply because of the cluttered environment (see Fig. 2).

In this paper, we propose an approach for identifying the state of each disconnect switch in a substation by analyzing images captured by regular pan-tilt-zoom (PTZ) surveillance cameras. Cameras can be installed and maintained without disrupting power distribution services, and a single camera can monitor multiple switches, besides being used for other surveillance tasks, reducing the cost of the system as a whole. On the other hand, image-based recognition has to deal with some of the same problems that lead to human errors, such as rain, fog, and bad lighting. Disconnect switches are also not designed to be particularly distinctive, having no special color or texture, and being surrounded (or even partially occluded) by other similar structures and equipment. Furthermore, besides variations caused by grime and oxidation, switches with the same function and electrical specification can have very different appearances, as seen in Fig. 3.

The approach described in this paper tackles those problems by working on images obtained from pre-programmed camera framings, with additional information about each switch being given in an initial labeling stage. That way, the system knows beforehand the approximate location of each switch in an image. After noise reduction and image registration using a phase correlation algorithm [1], we employ classifiers to determine the state of each switch. Two techniques were tested: a small convolutional neural network (CNN) [2], [3], and a support vector machine (SVM) [4], [5] fed with descriptors extracted by a modified version of the description stage from the Scale-Invariant Feature Transform (SIFT) [6]. The proposed approach achieved 100% accuracy on experiments performed at a real-world substation on the course of multiple days. Additional contributions of this paper include modifications to the image registration algorithm, which was made more robust to lighting variations; and to the SIFT descriptors, which were made more robust in scenarios where the relevant objects may be brighter or darker than the background.

The rest of this paper is divided as follows. Section II discusses related work. The proposed approach and the experiments performed to evaluate it are described, respectively, in Sections III and IV. Section V concludes the paper and points to future work.

II. Related Work

The idea of using image processing and machine vision for monitoring substation equipment is not new. For example, a patent from 1998 describes that idea in general terms [7]. It

Fig. 1. Close view from a closed disconnect switch.

Fig. 2. Disconnect switches in a substation, among other structures and equipment. At least 27 switches are visible in this image.

mentions disconnect switches as a possible target for monitoring, but algorithms are only briefly mentioned in a high-level manner. Other work address detecting or recognizing the state of various types of substation equipment based on images (including from thermal or infrared cameras) [8]–[12], but research dealing specifically with disconnect switches is not common.

Chen *et al.* [13] describe a system for detecting and determining the state of disconnect switches in transmission substations. For detection, they employ histograms of oriented gradients (HOG) [14] and linear discriminant analysis (LDA) [15]; and for state recognition, projection profile analysis and SVMs [4], [5]. They obtained good results (90% precision and 98.2% recall for closed switches) on a dataset containing a few hundred images. Although they address a similar problem to the one we are dealing with, their technique is not directly applicable in our scenario, as they explore the shape and symmetry of the contact area from a single model of high voltage switch, while our work deals with several models of medium voltage switches, which have different characteristics.

III. Proposed Approach

To be a viable alternative to other sensors, an image-based system for determining the state of disconnect switches must achieve very high accuracy. Given the challenges posed by the setting, that may prove too difficult for an approach that

Fig. 3. Disconnect switches. All these switches have the same function and electrical specification. Each image shows a switch with its base point at the lower right quadrant. In the uppermost 9 images, that switch is open, in the bottommost 9 images, it is closed.

receives an arbitrary image, locates switches, and determines their states. One possible solution would be adding to the switches markers displaying distinctive colors or patterns (e.g. QR codes), which could be used not only for detection, but also for identifying each switch. However, that solution requires power distribution to be halted for installation, and needs frequent maintenance to keep markers clearly visible. Moreover, the patterns must appear with a high enough reso-

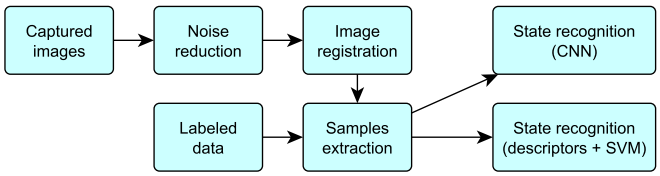


Fig. 4. Overview of the proposed approach.

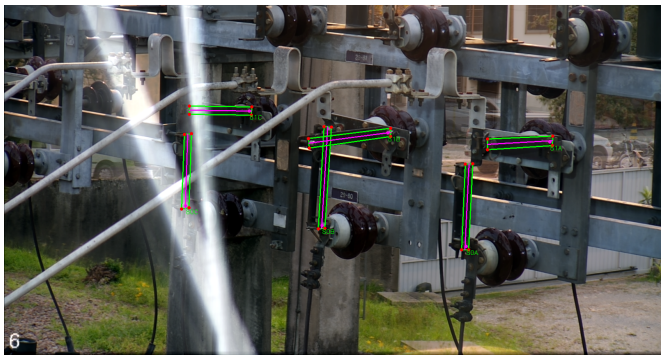


Fig. 5. A camera framing showing multiple switches.

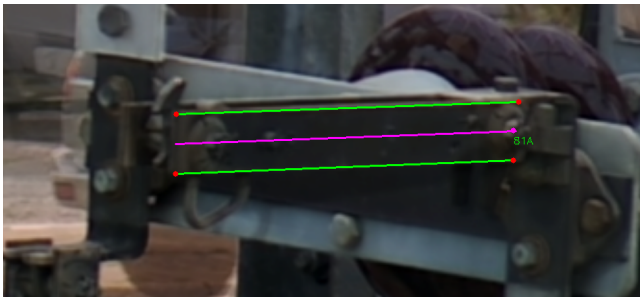
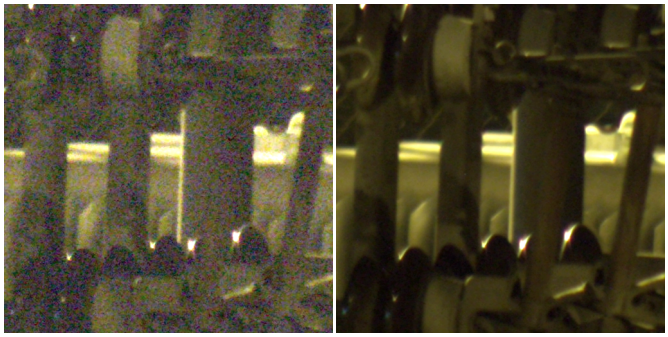


Fig. 6. A labeled switch.



(a) (b)

Fig. 7. Detail from a region from a single capture (a) and the average of 100 captures (b). Contrast was equally enhanced for both images, to emphasize the difference in noise levels.

lution, demanding cameras with highly precise PTZ control.

Our approach is non-intrusive, and combines image processing and machine vision techniques with data input by a human in an initial labeling stage. That data allows the system to know beforehand the approximate location of each switch in an image (note that any monitoring system must contain data about the switches, which must be uniquely identified in the power delivery system). Figure 4 shows an overview of the stages in the proposed approach. The following subsections detail each of those steps.

A. Data Labeling

Data labeling is part of the system setup, and occurs only once for each installation of the system. In this stage, a human operator defines a number of camera framings (PTZ settings) that show the monitored switches approximately (but not strictly) sideways, so that a change in state can be approximately described as a rotation of the conductive part (the switch "arm"). Once the framings are defined, the following data is given about each switch in each framing:

- The switch identifier in the power delivery system.
- A *base point* around which the conductive part rotates.
- Two line segments over the sides of the conductive part. A *centerline* is computed between these segments.
- A *point of view flag* that indicates if the switch opens by rotating to the left or to the right in the image.

The system must be able to cope with small imprecisions on the points given by the operator. Figure 5 shows an example of a framing, with the monitored switches highlighted. Figure 6 shows a zoomed-in view of a labeled switch.

B. Noise Reduction

To reduce noise, especially under low lighting, each input image is generated as the average of several captures (100, in our tests), taken in quick succession. This also makes objects passing quickly in front of the switches (such as rain drops) almost invisible. Figure 7 exemplifies the impact of this simple noise reduction measure.

C. Image Registration

Our approach was designed to work with regular surveillance cameras, positioned at a distance that allows the same camera to monitor several switches using different PTZ settings. This type of setup introduces variations when trying to reproduce the preset camera framings. The impact of these variations may be positive or negative — they can help machine learning algorithms to generalize better and avoid overfitting; or make learning a good model harder. To verify how these variations affect the results, our approach includes an optional image registration step.

Since the framing variations observed in our test setup were small enough to be approximated by translations, we employ a computionally inexpensive phase correlation algorithm [1] to align each input image to a reference. This algorithm is based on the fact that translations have small influence on the magnitude of the Fourier spectrum, while leading to mensurable changes on the phase. Given images f and g, we compute their respective discrete Fourier transforms (DFT) $\mathcal{F}\{f\}$ and $\mathcal{F}\{g\}$, and obtain the normalized cross power spectrum C by:

$$C = \frac{\mathcal{F}\{f\}\mathcal{F}\{g\}^*}{|\mathcal{F}\{f\}\mathcal{F}\{g\}^*|} \quad (1)$$

Here, $\mathcal{F}\{g\}^*$ denotes the complex conjugate. We then convert C back to the spatial domain:

$$c = \mathcal{F}^{-1}\{C\} \quad (2)$$

The final cross-correlogram r is obtained by taking the real and imaginary parts of c (c_r and c_i, respectively), and computing the magnitude in cartesian form:

$$r(x, y) = \sqrt{c_r(x, y)^2 + c_i(x, y)^2} \quad (3)$$

The offset between images f and g is estimated by taking the position of the peak intensity in r in relation to the origin (which is usually shifted to the center of the image). Sub-pixel accuracy can be achieved with interpolation techniques [1].

The shift between the images would be perfectly determined if the images were circularly shifted versions of each other (i.e. if elements that "leave" the image "re-enter" it at the opposite side). That is not the case in a real setting, where framing variations and time differences may result in new objects entering the scene, but as long as the scene remains mostly the same, a good estimation is possible. However, the scenario we consider also has severe lighting differences between images. This proved problematic for the original registration algorithm, so we have added a modification: instead of taking the DFT from the original images, we take it from the magnitude G of their gradient fields, computed by:

$$G(x, y) = \sqrt{G_x(x, y)^2 + G_y(x, y)^2} \quad (4)$$

where G_x and G_y are the partial derivatives from an image f:

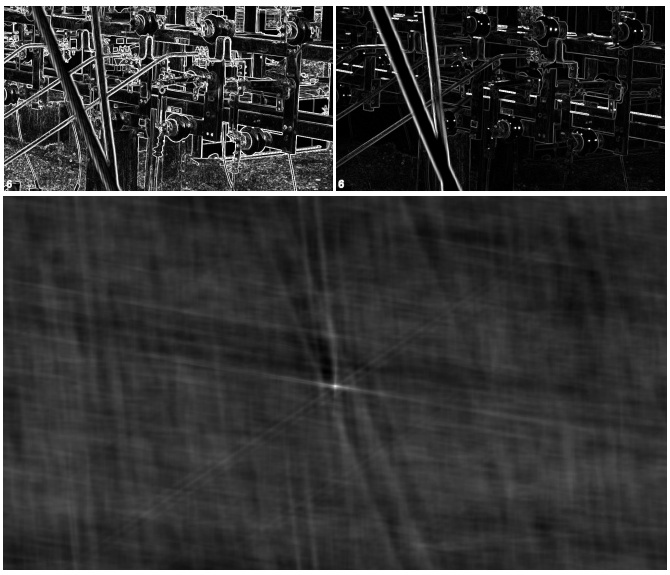
$$G_x(x, y) = f(x + 1, y) - f(x - 1, y) \quad (5)$$

$$G_y(x, y) = f(x, y + 1) - f(x, y - 1) \quad (6)$$

As gradient magnitudes are stronger over edges, registration is then guided by the scene structure and the shape of objects instead of pixel intensities. To exemplify this, Fig. 8 shows the normalized cross-correlograms computed, respectively, from two images and from their gradient magnitudes. It can be seen that the original images produce a large region with high values (in this example, in fact, the peak does not correspond to the correct translation between the images), while the gradient magnitudes produce a better localized peak.

The proposed modification was effective in improving image registration results under varying light. Figure 9 shows the average from 81 captures of the same scene, without image registration, with the original algorithm, and with the modified version. The amount of blur obtained without image registration is determined by the variations that occur when the camera tries to reproduce the preset PTZ settings. The modified version resulted in less variation than the original

(a)



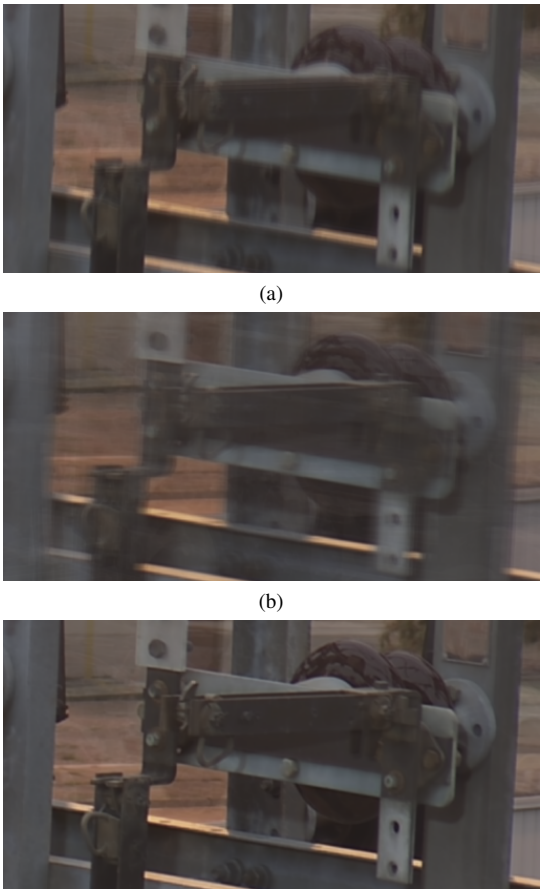
(b)

Fig. 8. Phase correlation between two images (a), and between the magnitudes of their gradient fields (b). Both images were captured using the same camera PTZ settings, at different times, and include a small translation.

algorithm — in fact, the original algorithm produced some incorrect results, in some cases performing worse than not performing image registration at all. In Sec. IV, the performance of the system as a whole with and without image registration will be compared.

D. Samples Extraction

The training samples and input images used by the classification algorithms are not the full images captured by the cameras. Instead, we use the information given during the labeling stage to crop a square region around each switch, with the *base point* (b_x, b_y) positioned at the lower right quadrant. The region size is based on the length c of the *centerline*,



(a) (b) (c)

Fig. 9. Region from the average of 81 captures of a scene, without image registration (a), with the original image registration algorithm (b), and with the modified version (c).

and scaled to a fixed size $w \times w$ (256×256 pixels, in our tests), with scale factor $s = w/(2c)$. The region may also be mirrored, based on the *point of view flag* m, which has a value of 1 or -1. All these transformations are represented by an affine transformation matrix T:

$$T = \begin{vmatrix} s \cdot m & 0 & -s \cdot b_x \cdot m + w \cdot 0.65 \\ 0 & s & -s \cdot b_y + w \cdot 0.65 \end{vmatrix} \quad (7)$$

Each sample will show the conductive part of one monitored switch at the right or the bottom, respectively for closed or open switches. The samples from Fig. 3 were extracted in this manner. Note that other switches may appear inside the same image, but only one switch will be analyzed in each sample.

E. State Recognition (Convolutional Neural Network)

Machine learning techniques are employed to determine the state of the monitored switch in each sample. The approximate position of the monitored switch in an image is known, thanks to the information given in the labeling stage. This has a direct impact on the models that will be learned: while there is still a large variety of possible situations, a model does not have to describe every possible aspect a switch may have, nor distinguish between switches and other structures — it only has to decide if the switch shown in a sample is open or closed.

The first technique we consider is a small convolutional neural network (CNN) [2], [3] with the architecture summarized in Table I. The network is trained from scratch using the Adam optimizer [16], using categorical cross-entropy as the loss function. We tested many other, more complex, architectures, including some that are commonly used in practice [17], but they were prone to overfitting — this is partially due to the relatively small number of training samples we used, as well as the limited number of points of view. By keeping the network small, we were able to avoid overfitting while keeping classification highly successful.

TABLE I
ARCHITECTURE OF THE CONVOLUTIONAL NEURAL NETWORK. DOF: DEGREES OF FREEDOM; CV: CONVOLUTION; MP: MAX. POOLING; LRN: LOCAL RESPONSE NORMALIZATION; FC: FULLY CONNECTED. PROCESSING SEQUENCE GOES FROM THE TOP TO THE BOTTOM.

Input Size	Operation	Activation	DOF
$256 \times 256 \times 3$	32×CV $3 \times 3 \times 3$ MP 2×2 LRN	ReLu	896
$128 \times 128 \times 32$	16×CV $3 \times 3 \times 32$ MP 2×2 LRN	ReLu	4,624
$64 \times 64 \times 16$	flatten	—	—
65536	16× FC dropout 20%	ReLu	1,048,592
16	2× FC	softmax	34

One important factor when training the network is providing images from all the switches, both open and closed, under various lighting and weather conditions. Since we will have multiple samples coming from the same framings, we also add random modifications to the samples, to prevent the model from learning to describe objects in the background. We modify a random rectangular region from each sample, in the upper or left half, for images containing, respectively, open or closed switches. Possible modifications include blur, mirroring, and changing the brightness.

F. State Recognition (Descriptor + SVM Classifier)

A second alternative was tested for determining the state of each switch, using a "descriptor + classifier" approach. For classification, we use a support vector machine (SVM) [4], with a radial basis function kernel and parameters automatically selected by a grid search [5].

Descriptors are extracted by a modified version of the description stage from the scale-invariant feature transform (SIFT) [6] (i.e. without interest point detection and orientation assignment). For each pixel in sample f, we compute the gradient magnitude G (see equations 4, 5 and 6), as well as the orientation θ:

$$\theta(x, y) = tan^{-1}(G_y(x, y)/G_x(x, y)) \quad (8)$$

Fig. 10. Two captures of the same switch at different times. Although the state remained the same, differences in lighting make the conductive part darker than the background in the first capture, but brighter at some points in the second capture.

The region is then split into a grid of 12×12 blocks, and a histogram of gradient orientations is computed for each block, with 8 orientation bins per histogram. The magnitude at each pixel position is divided into up to 8 histogram bins using trilinear interpolation. The final descriptor is obtained by concatenating all the values from the histograms from each block, resulting in a vector with $12 \times 12 \times 8 = 1152$ dimensions. Since we are interested in describing the entire region, we remove the weighting based on the distance to the region center. To make the descriptor more robust to lighting variations, it is normalized to an unit vector, has any values above 0.2 truncated, and is then normalized again.

Besides using different grid sizes and weight parameters, we also modified the SIFT algorithm by limiting gradient orientations to the interval $[0, 180°)$. The modified orientation θ_2 is computed by:

$$\theta_2(x,y) = \begin{cases} \theta(x,y) & \text{if} \quad \theta(x,y) < \pi \\ \theta(x,y) - \pi & \text{otherwise.} \end{cases} \tag{9}$$

As gradient orientations indicate the direction of change in contrast, this modification makes bright-dark and dark-bright transitions the same. This will turn the descriptor less capable of describing textures, but more robust in situations where the relevant objects may be brighter or darker than the background (and potentially more compact, as fewer histogram bins are needed to cover the same orientations). An example is shown in Fig. 10 (the images are in grayscale because SIFT uses only image intensities). Even though the state of the switch remained the same, the orientation of the gradients along the sides of the conductive part will change at some points, due to differences in lighting. The proposed modification attempts to reduce the impact of these differences on the extracted descriptors. The effect of this modification on the system as a whole was tested, as reported in Sec. IV.

Note that descriptors could be extracted from images with different sizes. However, we took the same fixed-size samples used by the CNN, to test both techniques under the same conditions.

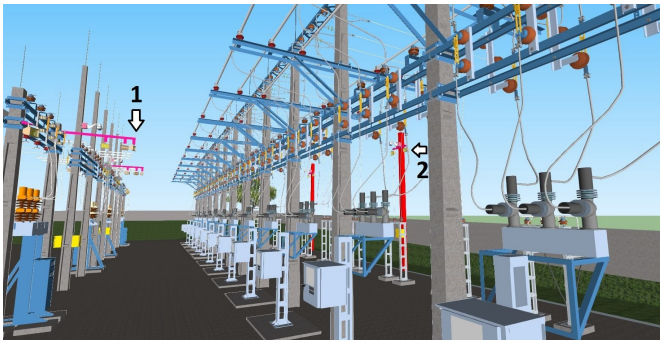


Fig. 11. Example image from the 3D model of the substation. Labels 1 and 2 indicate structures added to support the installation.

IV. Experiments and Results

The proposed approach was tested on images collected from a real distribution substation, over the course of multiple days. The prototype was written in the Python[1] language, using the OpenCV[2], TensorFlow[3] and TFLearn[4] libraries. The following sub-sections detail the experimental setup, how the dataset was created, and the tests performed to evaluate the proposed approach.

A. Experimental Setup

To collect images, we installed 4 regular PTZ surveillance cameras, with resolution 1920×1080, on a distribution substation in the city of Curitiba, Brazil. The substation has two horizontal rows containing disconnect switches, one with 36 switches and the other with 69. The cameras were positioned so that each row is monitored by two cameras, with each switch being visible approximately sideways from at least one camera. These cameras are capable of reproducing framings only in an approximate manner, and given that higher zoom levels lead to more sensitivity to small motor imprecisions, this prevented us from framing individual switches, so at least 3 switches are monitored in each framing. When describing the variations as translations, we observed shifts of up to 40 pixels in any direction. Each camera has 8 or 9 framings, with a total of 34 different framings. Four 100W LED illuminators were installed for nighttime illumination.

Since this setup demanded construction work in a risky environment, a 3D model of the substation was created prior to the system installation, so that the positions of cameras, illuminators and new structures could be planned (see Fig. 11).

B. Dataset Creation

To create the dataset, we collected images over a period of 9 days, between 11/28/2017 and 12/06/2017. During this period, which includes nighttime, the weather was mostly clear, with light fog or light rain on some occasions. Sunlight sometimes produced bright spots and strong shadows with visible edges,

[1] www.python.org
[2] www.opencv.org
[3] www.tensorflow.org
[4] tflearn.org

and gusts of wind made some images blurry. Overall, the dataset covers a wide range of capture conditions, except for snow, heavy rain and heavy fog.

A total of 4,114 images was collected. These images were divided in two sets, with the 1,601 images collected between 12/03 and 12/06 being used for training the models, and the remaining 2,513 images for testing them. The number of training and test samples extracted from these images is shown in Table II.

TABLE II
NUMBER OF TRAINING AND TEST SAMPLES IN THE DATASET.

	Open Switches	Closed Switches	Total
Train	2,095	4,014	6,109
Test	3,205	6,285	9,490

C. Tests

To evaluate the performance of the proposed approach, we trained models using the training samples, and used them to determine the state of the switches in the test samples. To cope with the randomness when modifying the samples, as well as in the initial CNN weights, each test result is obtained by going through 5 train-test runs, discarding the best and worst results, and taking the mean of the 3 remaining results (this mean was also very close to the median in all tested cases). Six variations of the proposed approach were tested, with and without the image registration stage, using a CNN or descriptor+SVM for classification, and with and without limiting the orientations in the descriptors. The same modified samples were used for all variations. Table III shows the number of errors and percentage of correct results (regarding the total number of test samples) with each variation.

TABLE III
MEAN NUMBER AND PERCENTAGE OF ERRORS OBTAINED WITH DIFFERENT VARIATIONS OF THE PROPOSED APPROACH.

Technique	Registration	Limit ori.	Errors	% Correct
CNN	Yes	—	6.33	99.93%
CNN	No	—	47.67	99.50%
Desc.+SVM	Yes	Yes	0	100%
Desc.+SVM	Yes	No	16.67	99.82%
Desc.+SVM	No	Yes	0	100%
Desc.+SVM	No	No	54.33	99.43%

All the variations performed well, with perfect accuracy in some cases. Registering the input images and limiting gradient orientations improved the results. The approach based on a descriptor and an SVM performed better than the CNN, but we note that a different network architecture, possibly trained with a larger variety of examples to avoid overfitting, could possibly attain the same results.

One relevant question when evaluating an approach based on machine learning is: what do the models actually describe? By design, the descriptors are based on the magnitude and orientation of gradients, with stronger contribution from pixels around image edges. Understanding the innards of a CNN is a more complex problem [3], but we have observed that many of the filters in the first layer produce strong responses around image edges with different orientations. Based on this, we raised the hypothesis that both models could be learning to identify the general orientation of the conductive part from the switches, which has two nearly parallel edges, mostly vertical for closed switches, and mostly horizontal for open switches. To test this hypothesis, we have extracted rotated samples, and evaluated them using the learned models. Testing several images, with both models, there seems to be a direct relation between the angle of the conductive part and the detected state (see Fig. 12). This hypothesis will be further investigated in future work.

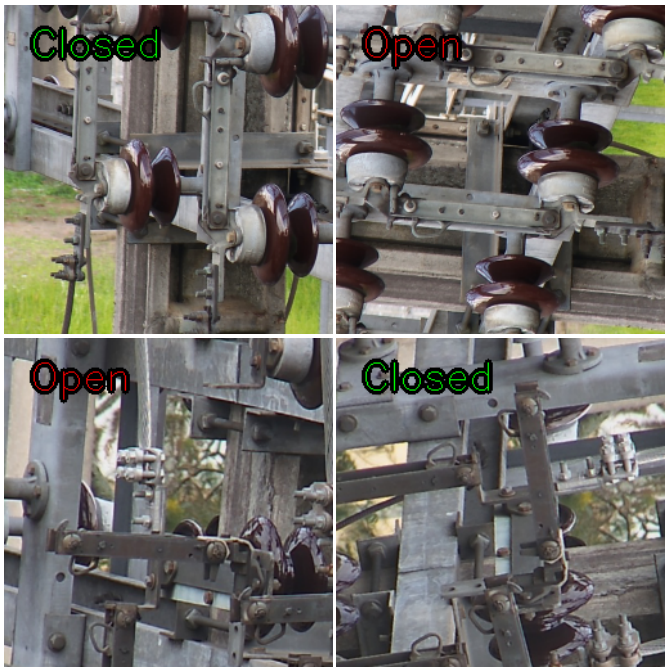


Fig. 12. The output of the system is related with the angle of the conductive part in the image. The same behavior was observed for models learned using CNNs and descriptors+SVMs.

V. CONCLUSIONS AND FUTURE WORK

We introduced an image-based approach for identifying the state of disconnect switches in power distribution substations. By combining knowledge provided by a human in an initial labeling stage with image processing techniques, the proposed approach achieved 100% accuracy in tests performed at a real-world substation. We tested variations considering models learned by CNNs and by a combination of gradient-based descriptors and SVMs. Modifications to the image registration algorithm and to the description stage from SIFT are additional contributions of this paper.

In future work, the proposed approach will be further tested on images captured over a longer period of time, as well as in other substations. Furthermore, although the proposed

approach was highly successful when recognizing situations observed during training, it remains unknown if the learned models can be directly employed for other substations, or when the switches cannot be maneuvered during training. These scenarios may create new challenges, requiring additional training, or different strategies for improving dependability, such as cross-checking results for switches visible by more than one camera, or combining the outputs from different classifiers. A challenge that will be addressed in future work is the case of "almost closed" switches — switches that seem closed but are not fully pushed, due to human errors or rust.

ACKNOWLEDGMENT

The authors would like to thank NVIDIA Corporation for kindly donating the Titan Xp GPU used for this research.

REFERENCES

[1] H. Foroosh, J. B. Zerubia, and M. Berthod, "Extension of phase correlation to subpixel registration," *IEEE Transactions on Image Processing*, vol. 11, no. 3, pp. 188–200, 2002.

[2] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015.

[3] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.

[4] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, September 1995.

[5] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, no. 3, pp. 27:1–27:27, May 2011.

[6] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, November 2004.

[7] E. O. S. III, "System for visual monitoring of operational indicators in an electric power system," u.S. Patent 5 805 813, September 8, 1998.

[8] A. Rahmani, J. Haddadnia, and O. Seryasat, "Intelligent fault detection of electrical equipment in ground substations using thermo vision technique," in *International Conference on Mechanical and Electronics Engineering*, vol. 2, 2010, pp. V2–150–V2–154.

[9] M. J. B. Reddy, B. K. Chandra, and D. K. Mohanta, "A DOST based approach for the condition monitoring of 11 kv distribution line insulators," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 18, no. 2, pp. 588–595, 2011.

[10] W. Cai, J. Le, C. Jin, and K. Liu, "Real-time image-identification-based anti-manmade misoperation system for substations," *IEEE Transactions on Power Delivery*, vol. 27, no. 4, pp. 1748–1754, 2012.

[11] Q. Zhou and Z. Zhao, "Substation equipment image recognition based on SIFT feature matching," in *International Congress on Image and Signal Processing*, 2012, pp. 1344–1347.

[12] X. Changfu, B. Bin, and T. Fengbo, "Research of substation equipment abnormity identification based on image processing," in *International Conference on Smart Grid and Electrical Automation (ICSGEA)*, 2017, pp. 411–415.

[13] H. Chen, X. Zhao, M. Tan, and S. Sun, "Computer vision-based detection and state recognition for disconnecting switch in substation automation," *International Journal of Robotics and Automation*, vol. 32, no. 1, 2017.

[14] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2005, pp. 886–893.

[15] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711–720, 1997.

[16] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proceedings of the International Conference on Learning Representations (ICLR)*, 2015.

[17] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet large scale visual recognition challenge," *International Journal of Computer Vision (IJCV)*, vol. 115, no. 3, pp. 211–252, 2015.