Journal of Imaging Science and Technology® 66(1): 010501-1–010501-9, 2022.

Occluded Image Function: A Novel Measure for Evaluating Machine Learning Classifiers for Biometrics

Adam Morrone
Mechanical Engineering Department, Colorado State University, USA
E-mail: adam.morrone20@alumni.colostate.edu

Wes Anderson
School of Biomedical Engineering, Colorado State University, USA

Steven J. Simske▲
Mechanical Engineering Department, Colorado State University, USA
School of Biomedical Engineering, Colorado State University, USA
Systems Engineering Department, Colorado State University, USA

Abstract. *A novel method is presented for evaluating the efficacy of object recognition algorithms on occluded images, called the occluded image function (OIF). The OIF describes system behavior in occluded environments and thus gives qualitative insight into their mechanisms; derivative metrics from OIF can also be used to quantitatively compare classifiers. To showcase the utility of the OIF, an experiment is performed by obstructing optical gait images from two biped robot models and using four binary machine learning classifiers to distinguish between them. The OIF diagrams are created from each experiment, and the resulting insights about the classifiers are discussed. Using the OIF, it is shown that the primitive classifiers can sometimes perform better under occlusion conditions, possibly due to pre-filtering of gait data by uniform occlusions. This result serves to demonstrate that the OIF is a useful tool for classifier evaluation.* © *2022 Society for Imaging Science and Technology.*
[DOI: 10.2352/J.ImagingSci.Technol.2022.66.1.010501]

1. INTRODUCTION

In this article, a novel method for assessing the efficacy and behavior of classification algorithms for occluded images is presented. Occluded images can be found in many places, such as optical forensic or biometric applications. In some of these applications, such as Iris biometrics and facial recognition, where glasses and reflections are common occlusions, having an incomplete image is expected and accurate classification in occlusion conditions becomes especially important [1–4]. Though there has been much research into object recognition and other imaged-based machine learning application in occlusion conditions, there is no universal measure available that specifically lends insights into the comparative efficacy of these algorithms in a variety of conditions [5–7]. There is a similar measure, however, that is very commonly used in the field of optics, called the optical transfer function (OTF) and the more common subset of the OTF, the modulation transfer function (MTF) [8]. These two measures describe how well an optical system can resolve high contrast sinusoidal frequencies in an image. The OTF and MTF are determined for increasingly thinner color bars, and the results are plotted as OTF/MTF vs bar thickness. The resulting curve's shape is insightful in evaluating the behavior of optical systems and thus, comparing them directly [9]. The method presented here applies the concepts of an OTF curve to the field of forensic cybersecurity and biometrics in the form of the occluded image function (OIF). This function arises from a plot of classification accuracy versus percent image occlusion. Similarly to the OTF or MTF curves, the shape and character of the OIF prove to be insightful into the efficacy of classification algorithms, even in cases where the algorithm is not intended to be used with occluded images. An example of the utility that the OIF provides can be seen in the works of Huang [10], Min [11], and Budzinskiy [12]. These representative studies all compare variations in different elements of an optical system in how they affect either the OTF or MTF. This allows them to isolate and compare the effects of subsystem variation scientifically, without resorting to system-level effectiveness test, which can often be expensive and open to distortion by subsystem interactions. This same form of behavior description and subsystem isolated comparison was not previously feasible for the occluded image recognition problem.

This study aims mainly to develop and describe the optical image function, and to demonstrate its utility. Namely, it provides a much-needed measure for system behavior in occluded conditions. This is important particularly to ensure robust classification solutions in situations where the amount of occlusion may not be known or expected. The OIF is a tool for understanding the efficacy of model behavior itself,

▲IS&T Member.

Received July 10, 2020; accepted for publication Mar. 14, 2021; published online May 20, 2021. Associate Editor: Oh-Seol Kwon.
1062-3701/2022/66(1)/010501/9/$25.00

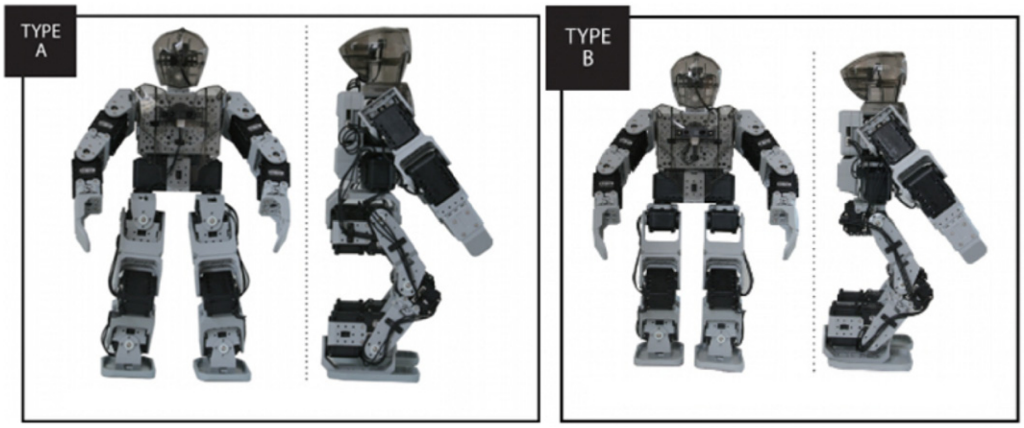

Figure 1. Bipedal robot configurations A (left) and B (right), which are used to simulate two bipedal subjects [24]. Note the difference in servo positions around the hips. Type A used an extra servo and adds the ability for leg adduction and abduction.

as opposed to the effectiveness of a model on a particular data set, as many others have already in the field of iris biometrics [13, 14].

To demonstrate the potential utility of this method, it is applied to an experiment in which two bipedal robot models are biometrically recognized. The robots, which simulate human models, are pinned with various visual markers, and are recorded walking along a stretch. The x and y coordinates of the markers over time act as the raw input data, and several engineered features are derived. Four primitive binary classifiers are trained and tested on this data. The level of visible obstruction is varied both physically and virtually, and the accuracy of these binary classifiers with respect to percent obstruction defines the OIF curve. The resulting behavioral insights are discussed.

In addition to the primitive classifiers, which lend themselves well for demonstrative purposes due to their low computation requirements, a classifier based convolutional neural network (CNN) architecture is compared for context in the physical obstruction case. The fields of deep learning and neural networks have been prolific in image processing [15, 16] and recognition [17, 18], with much advancement and focus on the classification task [19, 20]. As such, this proposed measure will have many applications in comparing these classification engines, though they are not discussed in depth here. The CNN used here is based on AlexNet [21], which won first place in ImageNet Large-Scale Visual Recognition Challenge in 2010, and has since been one of the gold standards for two-dimensional (2D) image recognition [22, 23].

This study first describes the high-level methods of generating the OIF curve, then walks through the details of the experimental set up used to demonstrate and validate this measure. Resulting OIF curves are then presented and their behaviors are discussed, as well as the overall utility of the OIF in describing and comparing these behaviors. Finally, a broad summary, known issues, and potential future work is included in the conclusion.

2. MATERIALS AND METHODS

2.1 *Plotting OIF*

Occluded image function (OIF) is plotted based on data collected from the biped robot using the method described in the following section. This function uses the percent occlusion as the independent variable and the cross-validated F1-score as the dependent variable. These are plotted against each other to form a curve. For all curves except the physical tests, the plots shown are mild moving average the raw data to remove low-level noise. The general shape, trends, and other features of this curve can provide insight and information on the efficacy of these classifiers with respect to occluded images.

2.2 *Biped Robot Subject*

For this experiment, gait analysis will be used to classify two bipedal robot gaits. The robot used in this experiment was the Bioloid Premium, which can be reassembled in three different configurations that each result in a slightly different walking style by rearranging the servo motors that constitute his hips and legs, which results in three distinct gaits. The two configurations that appeared to be most different, A and B, were chosen for comparison (Figure 1). The gait data will be derived from video footage of the robots walking to imitate security camera biometrics.

2.3 *Computer Vision Tracking*

Colored markers were fixed to the head, foot, and upper and lower legs of the robots (Figure 2). Alternating red and blue markers were used in pairs for each location to simplify differentiation and downstream data processing. Finally, green markers are placed in two locations along the walking area for use as coordinate positions and length scaling. To gather the raw data, a digital camera was placed at a fixed location to record footage of the robots walking. The footage was processed using OpenCV [25] by generating an HSV binarization mask and three-stage erosion–dilation algorithm. After filtering, the geometric mean center of a

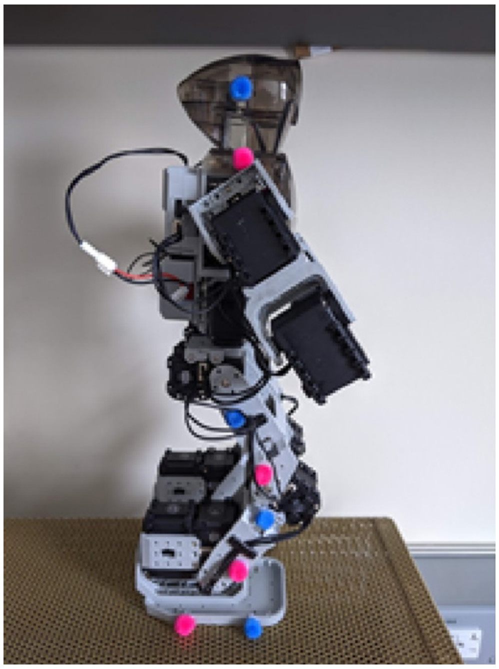

Figure 2. Robot B is shown with red and blue marker pairs attached. The pairs allow the calculation of tilt and joint angle. The four pairs are head, upper leg, lower leg, and foot. The markers are in the same locations in Robot A.

minimum encompassing circle is calculated to avoid partial occlusion effects on the centroid and used for the final Cartesian coordinates for each marker. This method is repeated on each frame for a total of 30 times per second.

All x and y coordinates are first scaled within the boundaries of the green table markers. As is shown in bold red lines in Figure 3(a), a straight line is drawn between green markers 1 and 2, *G1* and *G2*, respectively, then two perpendicular lines are drawn which go through the two markers, thus making a universal bounding box that can adjust for tilt and any irregularities in the camera position between recordings. *G1* is then used as the coordinate axis origin, and all other data points are scaled relative to this. The bounding box also allows for universal starting and ending lines to be established. This can also be seen in Fig. 3(a), where all the tracing lines start together. The actual distance between *G1* and *G2* is 90 cm.

2.4 *Physical and Virtual Occlusions*

Two methods were used and compared to simulate occluded footage: physical and virtual occlusions. The physical occlusions consisted of evenly spaced cardboard slats resembling a picket fence (Figure 4). The physical slats are uniformly 5 cm wide and are spaced evenly to construct four arbitrary occlusion percentages: approx. 25%, 33%, 50%, and 67%.

To perform the virtual occlusions, the data from the non-occluded footage is used, and an algorithm removes data in predefined geometrical pattern, so as to create virtual obstructions. A number of different virtual occlusion patterns are used. It is important to note that the virtual occlusions were applied to the marker signals extracted from the footage and not the footage itself; in other words, the virtual occlusions were applied downstream of the CV processing.

(1) Constant-width slats: this pattern mimics the physical occlusions used in this study. It uses a scaled, constant-width occlusion. No partial occlusions are used. As the percent occlusion is increased, whole slats are added. To compare directly with the physical occlusion footage, slats of five centimeters are simulated. In addition to this, slats of 1.25, 2.5, 5, and 10 cm are simulated for a total of 4 virtual slat widths. To eliminate any phase-shift variables, 10 phase-shifted versions of each percent occlusion are simulated, and the mean values of the classification outcomes are used as the final metrics.
(2) Expanding-width slats: this pattern simulates evenly spaced slats as well, but the slats are not a constant width. Instead, the total number of slats is constant, with constant center points. They expand pixel by pixel from 2% occlusion through 96% occlusion. For this study, 3, 6, and 9 total slats were simulated using this pattern. Phase shift is accounted for in the same way as described for the constant-width slats pattern.
(3) Random: this pattern generated random slat widths and spacings. For each percent occlusion, 1000 trials were performed and the mean scores were used. Since the spacing is entirely random, no accounting for the phase was made. This attempts to show a generalized trend with respect to percent occlusion, regardless of the size or pattern of the actual occlusions.

2.5 *Object Recognition*

The robot was filmed walking across the stage 24 times in both configurations and for each occlusion setting. After all of the physically and virtually occluded images were gathered, video samples were assigned 50/50 to training and test sets, so as to provide a large enough test set for validation. The robots did not always walk exactly straight across the table, frequently drifting to one side. This drift toward or away from the camera resulted in odd 2D features such as fixed markers appearing to grow closer or further apart throughout the video. Another inconsistency in the gathered data is due to the fact that the slats were not always perfectly upright, sometimes leaning slightly one way or another. No upstream effort was made to account for this; instead, the introduced error was incorporated into the learning model.

Four primitive machine learning models—support vector machine (SVM), K-nearest neighbors (KNN), binary decision tree, and an ensemble method—were used for binary classification, where the system tried to correctly sort footage from robots A and B. In addition to the primitive classifiers, a CNN based on AlexNet was used on the experimental data for context. This classifier used the one-dimensional (1D) signals of the marker positions and maps them into 2D images before training on the image with AlexNet. Nakano et al. [26] noted the potential of

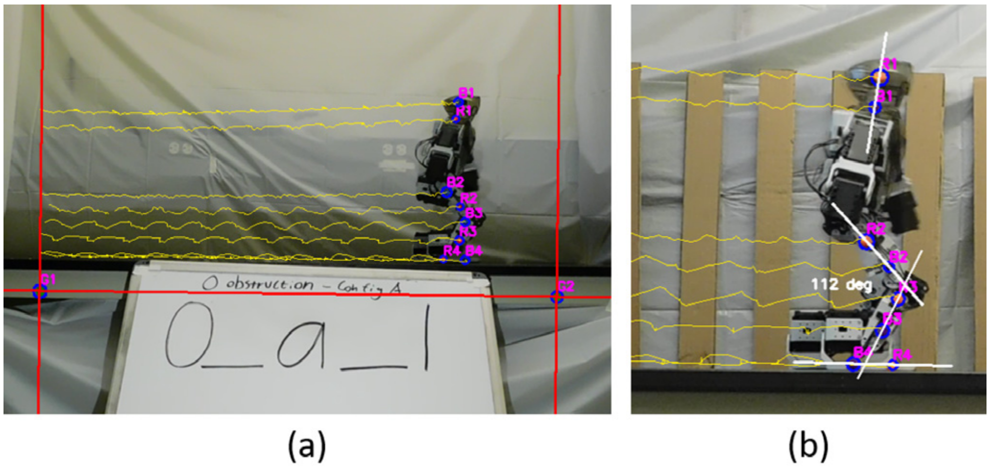

Figure 3. CV-based tracking of bipedal robot gait. (a) shows the green markers creating absolute boundaries and coordinates, and (b) shows the calculation of joint angles using lines drawn through adjacent markers. Blue markers are labeled with B#, red with R#, and green with G#.

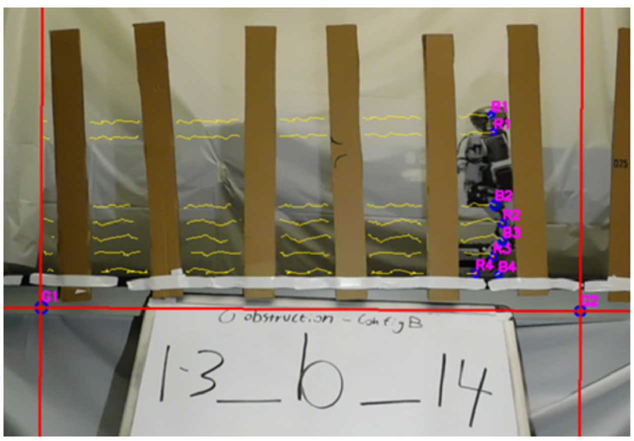

Figure 4. Physical occlusions. Each slat is 5 cm wide. Shown is a 33% occluded footage.

converting 1D signals to other forms for the purpose of improved classification, as techniques for data reduction and ease of visualization are needed for accurate classification. Representing time series data as images, therefore, is an interesting and attractive method for potential value in 1D signal classification.

Cross-validation classification was performed to determine the F1-score of the system (defined as the harmonic mean of precision and recall [27]), and thus how effective the system is as distinguishing between robot A and robot B based on marker-assisted video footage. The classifiers were trained using a targeted set of engineered features as shown in Table I. The classifiers used for SVM, KNN, binary decision tree, and an ensemble method were the MATLAB proprietary fitcsvm [28], fitcknn [29], fitctree [30], and fitcensemble [31] functions, respectively. MATLAB's automatic optimization algorithm was employed, which determines hyperparameters based on fivefold cross-validation loss [32]. The resulting hyperparameters are recorded in Table II. F1-score is used as the primary outcome metric when evaluating these classifiers.

Table I. Engineered feature set used by the classifiers.

Feature name	Description	Number of features
Mean	Mean value for horizontal (x) and vertical (y) coordinates of each marker	16
Median	Median value for horizontal (x) and vertical (y) coordinates of each marker	16
Step count	Total number of local maximums for timeseries horizontal (x) coordinates	8
Step width	Mean horizontal distance between local maxima in timeseries horizontal (x) coordinates	8
Step height	Mean vertical distance between local maxima in timeseries horizontal (x) coordinates	8
Slope	Slope of the best-fit line for timeseries horizontal (x) coordinates	8
	Total number of features:	64

2.6 *Runtime*

The efficiency of generating the OIF for a classifier, expressed as the computation runtime, is a linear function of training runtime, the base classifier being used, and the chosen fidelity of the OIF in question. The time it takes to populate an obstructed image function plot can therefore

Table II. Optimized hyperparameters and efficiency in terms of average training runtimes on a typical engineering laptop for the primitive machine learning classifiers.

Classifier	Parameter 1	*Value 1*	Parameter 2	*Value 2*	Parameter 3	*Value 3*	Training runtime (s)
SVM	Function	*Linear*	Box constraint	*0.031486*	Kernel scale	*0.98824*	107
KNN	Distance	*Seuclidean*	Number of neighbors	*1*	—	—	62
Decision tree	Minimum leaf size	*11*	Minimum parent	*18*	Maximum splits	*23*	36
Ensemble	Method	*Bag*	Number of learning cycles	*498*	Minimum leaf size	*4*	291
CNN*	Number of epochs	*6*	Initial learning rate	*1 e-4 seconds*	—	—	35

*Not trained with a grid search.

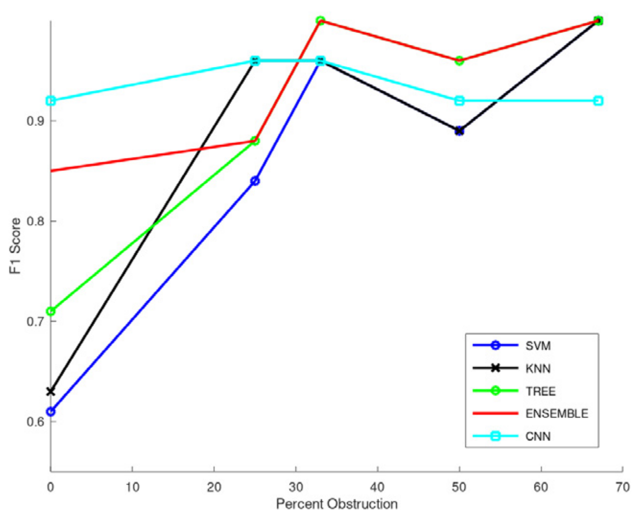

Figure 5. OIF for physical occlusion results. Measurements were taken at 0, 25, 33, 50, and 67% occlusion. This plot is scaled out to 100% occlusion for consistency with the virtual occlusion results. [R^2 values: SVM = 0.776; KNN = 0.597; Tree = 0.754; Ensemble = 0.674; CNN = 0.047.]

be approximated by $\boldsymbol{t} = r * \boldsymbol{f}$, where $\boldsymbol{t}$ is the total time, $\boldsymbol{r}$ is the runtime of the base classifier, and $\boldsymbol{f}$ is the fidelity, in terms of number of iterations, number of tested obstruction layers, or both. So, for instance, the time to generate the curve for SVM in Figure 5, using the values in Table II, would be $\boldsymbol{t} = 107$ s $*$ 5 obstruction ratios $= 535$ s. Likewise, the OIF curve generation time for the KNN method on even spaced virtual obstructions would be $\boldsymbol{t} = 62$ s $*$ 100 obstruction ratios $*$ 10 phase iterations $=$ 62000 s $= 17.2$ h. These runtimes were recording on typical engineering laptop, with an Intel Core i7-8550U CPU running at 1.99 GHz.

3. RESULTS

The results from the physical occlusion tests are shown in Fig. 5. Unexpectedly, the data shows a steep and significant increase in classification accuracy for all primitive classifiers as occlusion increased. At 67% occlusion, all primitive classifiers demonstrated 100% accuracy. In addition, all primitive classifiers showed a local maximum at 33% occlusion. Thus, the two maxima occur at complementary occlusion percentages, with the first maximum occurring when the signal is twice as occluded as the second. The CNN demonstrated distinct behavior compared with the primitive classifiers, as it was nearly straight, sitting between an F-score of 0.9 and 1.0 regardless of the level of obstruction.

The OIF from virtual occlusions with constant-width slats is shown in Figure 6. All four slat widths used showed an overall increase in classification accuracy as occlusions increased, which largely confirms the results found in the physical experiments. A similar pattern is noted, especially with the wider slats like 5 cm, where there is a quick increase in classification accuracy until around 20–30% occlusion, at which point it seems to plateau.

The next virtual experiment was to keep a constant number of evenly spaced slats and increase their thickness, as opposed to increasing the number of identical slats. The OIFs from these tests are shown in Figure 7. The trend from these experiments is, overall, a decrease in classification accuracy as percent occlusion increases.

The OIF results for the random virtual occlusions are shown in Figure 8. This curve is by far the smoothest, which suggests that the large number of random iterations helped capture the overall trend.

It is worth noting that the variance within these data sets is fairly high. The error bars shown in the figures represent ±1 standard deviation within each percent occlusion set. For

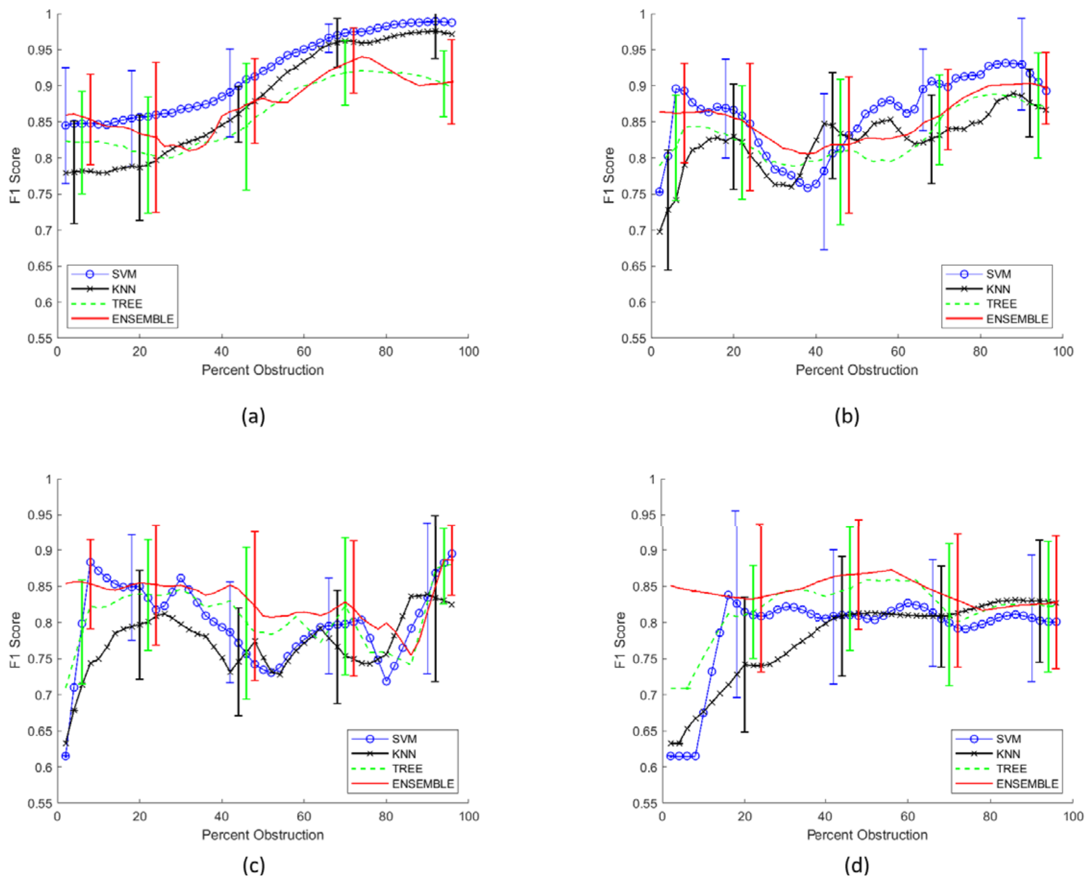

Figure 6. OIF for constant-width slats virtual occlusions. These virtual occlusions most closely match the physical occlusions. The virtual slat widths are (a) 1.25, (b) 2.5, (c) 5, and (d) 10 cm. Each datum shown is a mean value of 100 phase-shifted versions of the same percent occlusion. The error bars represent plus and minus 1 standard deviation within each set. [R^2 values: (a) SVM = 0.712; KNN = 0.734; Tree = 0.564; Ensemble = 0.459 | (b) SVM = 0.289; KNN = 0.441; Tree = 0.309; Ensemble = 0.174 | (c) SVM = 0.002; KNN = 0.146; Tree = 0.009; Ensemble = 0.146 | (d) SVM = 0.179; KNN = 0.702; Tree = 0.195; Ensemble = 0.074.]

most plots this refers to the distribution of F1-scores among the 100 different phased-shifted versions of each percent occlusion. For Fig. 8, however, it represents the distribution among the 1000 different randomly generated occlusion widths and spacing for each percent occlusion. Often the standard deviation was in the range 0.05–0.1. This means that small changes in phase or even in the order of occlusion could result in fairly large changes in classification accuracy with the classifiers and features used here. In the mean, the SVM had the lowest variation, and the binary decision tree had the highest.

4. DISCUSSION

The results of the various experiments in this study serve to illustrate the potential utility of the OIF in comparing biometric or other classification algorithms when occlusions are present. Using this function, an unusual behavior pattern appeared where the classification accuracy increased with occlusion in some circumstances, as seen in Figs. 5 and 6, which show the OIFs for physical and virtual constant-width slats, respectively. This behavior seems counterintuitive since it would be expected that more data (less occlusion) would result in higher classification accuracy, but the reverse appears to be true with both the physical and virtual constant-width slats. This increasing accuracy with occlusion ($R^2 > 0.7$) pattern did not show up in any other occlusion configurations, which indicates that the specific configuration was likely the cause of the unexpected behavior. Since many of the features used in this study reflected the gait cycle, it is possible that the constant-width slats actually performed some level of pre-filtering, removing relatively more noise than data (i.e. improving SNR) for the chosen features and classifiers to analyze. This seems plausible, because the constant widths of the slats consistently obstruct the same percentage of the gait cycle, as opposed to the random or pseudo-random breakup of gait cycles that take place in all the other occlusion scenarios. This behavior, then, may reflect the behavior of a true "picket-fence" scenario, such as trying to biometrically identify footage of a person walking, occluded by a literal fence. However, this pattern would likely not be found with more sophisticated features and/or classifiers and may instead simply indicate that these features and classifiers and particularly non-ideal for this occlusion type. This is an insight that the OIF makes abundantly clear. It is worth noting that the ensemble

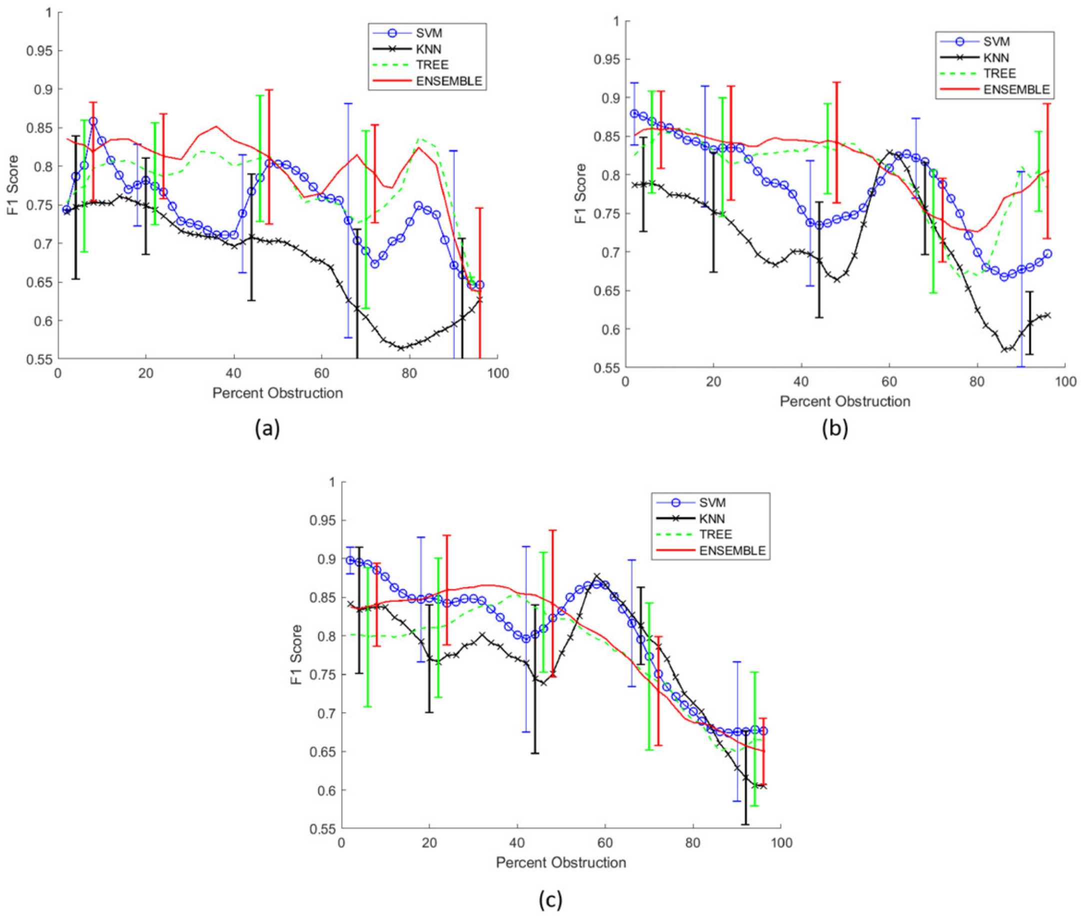

Figure 7. OIF for expanding-width slats virtual occlusions. The number of slats present is (a) 12, (b) 6, and (c) 4. Each datum shown is a mean value of 100 phase-shifted versions of the same percent occlusion. The error bars represent plus and minus 1 standard deviation within each set. [R^2 values: (a) SVM = 0.277; KNN = 0.581; Tree = 0.178; Ensemble = 0.399 | (b) SVM = 0.471; KNN = 0.286; Tree = 0.350; Ensemble = 0.449 | (c) SVM = 0.590; KNN = 0.328; Tree = 0.487; Ensemble = 0.637.]

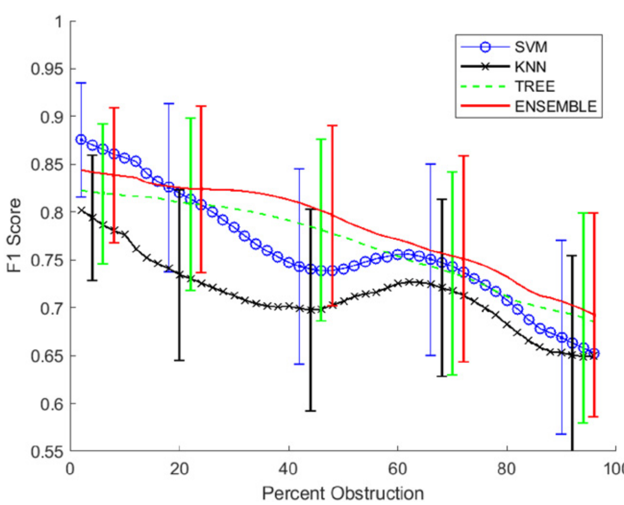

Figure 8. OIF for random virtual occlusions. Each datum shown in the mean of 2400 randomly generated occlusion widths and spacings. The error bars represent plus and minus 1 standard deviation within each percent occlusion set. [R^2 values: SVM = 0.861; KNN = 0.275; Tree = 0.934; Ensemble = 0.939.]

classifier performed markedly better than the other three classifiers with 0% occlusion (Fig. 5). Based on this visual function measure, it would be reasonable to conclude that using the ensemble method would be the best choice among the four primitive classifiers investigated.

The results from the CNN on the experimental data further illustrate the utility of the OIF. Whereas the efficacy of the primitive classifiers all appear to have a direct relationship ($R^2_{\text{max}} = 0.77$) with the amount of obstruction, the CNN seems unrelated to the obstruction levels used ($R^2 = 0.05$). This suggests that the CNN algorithm used here is not affected by cyclical pattern obstruction in nearly the same was that the primitive algorithms were. Interestingly, the best results from the primitive algorithms are better than the best results from the deep learning algorithm, but the mean F1-score from the five obstruction ratios CNN (0.94) was on par with the highest mean of the primitive classifier, the ensemble method (0.94). The CNN's accuracy in distinguishing the two robots is relative unaffected by even large amounts of obstruction.

Another helpful case study to demonstrate the utility of the OIF is found by evaluating the results of the randomized trials, shown in Fig. 8. A few notable observations can be made. First, there appear to be two curve shapes, one which is shared by the ensemble and decision tree methods, and a second which is shared by the SVM and KNN methods. This suggests a similarity in the underlying classification mechanisms. A second notable observation is made by directly comparing the accuracy of each classifier. The SVM is clearly the most accurate classifier in the mean if the percent occlusion is below about 20%, after which the ensemble classifier is consistently the most accurate. In addition, the KNN classifier is the least accurate at all occlusion levels. Based on information provided by analyzing the OIFs, one can make an informed decision about which classifier to use. Namely, if there is little expected occlusion in most cases, the SVM may be the optimal choice, but if there is a considerable amount of occluded footage (or other image types) expected, the ensemble method may be most effective.

Thanks to the OIF, all of these observations were made possible, as there is no similar visual function measure, which so clearly offers insight into occluded image classification behavior. This metric is envisioned to be used in two main ways: as an objective comparison with standardized tests, and as a case-by-case tool for evaluating specific situations. Several potential standardized tests have been shown in this report, any of which could be used as a standard. The randomized test, with a sufficient number of random trials, seemed to give the most consistent data. Secondary metrics could be derived from a standardized test as well, such as the average slope of a best-fit line, or the total area under the curve. It is worth noting that all the experiments in this study, such as OIF can also be used as a case-by-case evaluation tool for applications such as forensic biometrics. If a known occlusion pattern is present, for instance, in a security camera footage of a harbor or across a street with mailboxes and street lamps in the way, the exact pattern could be mimicked virtually, and the best classifier and features could be determined for that specific application as a pre-flight test for occlusion. This kind of test could use standardized images, such as the ones used in this study or any other publicly available optical gait database. As such, the optimal biometric classification algorithm for a specific, occluded, security camera feed could be determined entirely virtually, without the need for any physical experimentation on site. This means that this method would be effective at evaluating systems that are mounted on moving objects, such as drones, satellites, and other UAVs.

5. CONCLUSIONS

This study described the OIF curve metrics and presented various examples that demonstrate its utility. In doing so, an interesting phenomenon was also observed in which evenly spaced, constant-width slat occlusion improved the classification accuracy of the simple classifiers used in this study. This means that the classification algorithm was not taking full advantage of the available data and might be improved by incorporating better image features or some form of pre-filtering. Thus, even if an image processing algorithm is not intended to be used for occluded images, generating as OIF may still prove to be insightful. Here, we proposed the OIF as a method for classifier assessment in occluded conditions. In addition, this OIF study demonstrates the differences in behavior between primitive and deep learning algorithms in how they are affected by obstructions. The CNN's behavior appeared to have no relationship with the amount of visual obstruction in the images.

The experiments here serve to illustrate this measure, how it works, and the value it adds. Much more can be done to showcase this function with various other machine learning, deep learning, and neural network-based classifiers. Though the utility of this method was explored here, a universal standardization of assessment for this metric is still needed and could be further developed in future work. Any or all of the virtual occlusion experiments used in this study could be incorporated into that standard. For such a standard to be widely accepted, an efficacy versus effectiveness study may have to be done so as to link comparisons made with OIF to the actual performance of these classifiers in real world scenarios. In addition, more derived metrics from OIF can be created and standardized to provide a means of quantitative comparison. Though this metric was developed primarily with forensic and biometric applications in mind, it may be useful anywhere occluded signals are found, such as a faulty long-range radio connection.

REFERENCES

1. A. Pandey, *Iris Biometric Analysis under Occlusion*, Ph.D. diss (2016).
2. N. D. Kalka, J. Zuo, N. A. Schmid, and B. Cukic, "Estimating and fusing quality factors for iris biometric images," IEEE Trans. Syst. Man Cybern. A **40**, 509–524 (2010).
3. K. W. Bowyer, K. Hollingsworth, and P. J. Flynn, "Image understanding for iris biometrics: A survey," Comput. Vis. Image Underst. **110**, 281–307 (2008).
4. D. D. Sawat and R.S. Hegadi, "Unconstrained face detection: a deep learning and machine learning combined approach," CSI Trans. ICT **5**, 195–199 (2017).
5. M. Tan, Z. Hu, B. Wang, J. Zhao, and Y. Wang, "Robust object recognition via weakly supervised metric and template learning," Neurocomputing **181**, 96–107 (2016).
6. Y. K. Jung and Y. S. Ho, "Multiple-object tracking under occlusion conditions," Vis. Commun. Image Process. **4067**, 1011–1021 (2000).
7. G. Hetzel, B. Leibe, P. Levi, and B. Schiele, "3D object recognition from range images using local feature histograms", *Proc. 2001 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition*, CVPR 2001 (IEEE, Piscataway, NJ, 2001), Vol. 2, p. II.
8. G. D. Boreman, *Modulation Transfer Function in Optical and Electro-Optical Systems* (SPIE Press, Bellingham, WA, 2001).
9. C. S. Williams and O. A. Becklund, *Introduction to the Optical Transfer Function* (SPIE Press, Bellingham, WA, 2002).
10. J. Huang, Y. Bao, and T. K. Gaylord, *Analytical Phase Optical Transfer Function for Gaussian Illumination and the Optimized Profiles* (Optical Society of America, Washington, DC, 2020).
11. J. W. Min and H. W. Jeong, "Comparison of modulation transfer function in measurements by using edge device angle in indirect digital radiography," J. Radiol. Sci. Technol. **42**, 259–263 (2019).
12. S. Budzinskiy and A. Razgulin, "Defocus optical transfer function: fast evaluation and lightweight storage based on cubic spline interpolation," J. Opt. Soc. Am. A **36**, 436–442 (2019).

[13] H. Proenca and L. A. Alexandre, "A method for the identification of noisy regions in normalized iris images," *18th Int'l. Conf. on Pattern Recognition (ICPR'06)* (IEEE, Piscataway, NJ, 2006), pp. 405–408.

[14] X. D. Wang, H. Q. Wei, S. M. Li, C. Gao, and R. Y. Huang, "Unsupervised facial image occlusion detection with deep autoencoder" *Eleventh Int'l. Conf. on Digital Image Processing (ICDIP 2019)* (SPIE, Bellingham, WA, 2019).

[15] Q. Zheng, M. Yang, J. Yang, Q. Zhang, and X. Zhang, "Improvement of generalization ability of deep CNN via implicit regularization in two-stage training process," IEEE Access **6**, 15844–15869 (2018).

[16] Q. Zheng, X. Tian, M. Yang, Y. Wu, and H. Su, "PAC-Bayesian framework based drop-path method for 2D discriminative convolutional network pruning," Multidimens. Syst. Signal Process. 1–35 (2019).

[17] H. Qin, R. Gong, X. Liu, X. Bai, J. Song, and N. Sebe, "Binary neural networks: a survey," Pattern Recognit. **105**, 107281 (2020).

[18] Q. Zheng, X. Tian, N. Jiang, and M. Yang, "Layer-wise learning based stochastic gradient descent method for the optimization of deep convolutional neural network," J. Intell. Fuzzy Syst. **37**, 5641–5654 (2019).

[19] Q. Zheng, M. Yang, X. Tian, N. Jiang, and D. Wang, "A full stage data augmentation method in deep convolutional neural network for natural image classification," Discrete Dyn. Nat. Soc. **2020** (2020).

[20] Q. Zheng, X. Tian, M. Yang, and H. Su, "CLMIP: cross-layer manifold invariance based pruning method of deep convolutional neural network for real-time road type recognition," Multidimens. Syst. Signal Process. **32**, 1–24 (2020).

[21] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," Adv. Neural Inf. Process. Syst. **25**, 1097–1105 (2012).

[22] S. Lua, Z. Lu, and Y. Zhang, "Pathological brain detection based on AlexNet and transfer," J. Comput. Sci. **30**, 41–47 (2019).

[23] P. Ballester and R. M. Araujo, "On the performance of GoogLeNet and AlexNet applied to sketches," *AAAI'16: Proc. Thirtieth AAI Conf. on Artif. Intell.* (ACM, New York, NY, 2016), pp. 1124–1128.

[24] Introduction. ROBOTIS e-Manual. Retrieved 11 November 2019, from http://emanual.robotis.com/docs/en/edu/bioloid/premium/ (2019).

[25] G. Bradski and A. Kaehler, "Dr. Dobb's Journal of Software Tools," The OpenCV Library **25**, 120 (2000).

[26] K. Nakano and B. Chakraborty, "Effect of data representation for time series classification—a comparative study and a new proposal," Mach. Learn. Knowl. Extr. **1**, 1100–1120 (2019).

[27] N. Chinchor, "The statistical significance of the MUC-4 results", *MUC4'92: Proc. 4th Conf. Message Understanding* (ACM, New York, NY, 1992), pp, 30–50.

[28] fitcsvm. Copyright 2013-2019 The MathWorks, Inc.

[29] fitcknn. Copyright 2013-2019 The MathWorks, Inc.

[30] fitctree. Copyright 2013-2019 The MathWorks, Inc.

[31] fitcensemble. Copyright 2016-2018 The MathWorks, Inc.

[32] Train support vector machine (SVM) classifier for one-class and binary classification - MATLAB fitcsvm. (2021). Retrieved 24 January 2021, from https://www.mathworks.com/help/stats/fitcsvm.html.