optica

Scatterer recognition via analysis of speckle patterns: supplementary material

EADAN VALENT AND YARON SILBERBERG

Department of Physics of Complex Systems, Weizmann Institute of Science, Rehovot 76100, Israel
**Corresponding author: yaron.silberberg@weizmann.ac.il*

Published 14 February 2018

This document provides supplementary information to "Scatterer recognition via analysis of speckle patterns," https://doi.org/10.1364/optica.5.000204.

In this supplementary material we explain in further detail the two supervised algorithms we used to classify speckle patterns created by different scatterers, the manner in which we used these algorithms, how we utilized multiple images for material classification, and the use of measurements in transmission to determine sample width.

CNN Algorithm:
The convolutional neural network (CNN) [1] is a type of deep learning algorithm that receives many inputs (image pixels in this case), organizes them in a vector, and multiplies that vector by a number of matrices (known as layers). The dimensions of the final matrix are such that the resulting vector will contain M values. M is the number of options (types of scatterers in our case) the algorithm can classify each input. Each of the M values of the output represents the algorithm's predicted probability of the input belonging to that class. Initially, the algorithm must be trained in order to learn its predictive traits. The training is done by first inputting many labeled examples (90% of the data in our case) into the algorithm. Then, the algorithm's predicted outputs (initially random) are compared to the true class of each input. The difference is then minimized using one of several possible optimization techniques, such that the values predicted by the algorithm are as close to the correctly labeled classes as possible. Then, the algorithm's predictive rate is tested by using new unlabeled inputs and checking whether the algorithm predicted their class correctly. The algorithm applied in this work (described in Code 1, Ref. [2]) ran on python using a code based on an example from the Keras [3] website with a Tenserflow backend. It uses cross-entropy as its defined loss (what it`s trying to minimize), the Adadelta optimizer [4], and 10 epochs (iterations on the complete data) per material set.

In our experiment, we first inserted N speckle images into our CNN algorithm. Then, each image was split into 16 subsections for the network to analyze, and each image section underwent the following layer sequence: 2D convolution (kernel size of ~8), 2D convolution (kernel size of ~2), max pooling, dropout, flatten, dense, dropout, dense. Several kernel sizes were tested for each material set, and the size giving the best result was used; this was done to the total data consisting of M data sets. We then split the data into training and testing sets with a 9:1 ratio. The total number of image sections in the training data is: $\frac{9}{10} \times 16 \times N \times M$, with the rest used to test the algorithm's predictive rate.

LDA Algorithm:
The linear discriminant analysis algorithm [5] is less computationally expensive than deep learning algorithms, and allows the user to define the parameters it uses for its classification purpose. It is first given some set of vectors of dimension k, each pre-labeled as one of M possible classes, and attempts to cluster them into M subspaces in a k dimensional space. This can be visualized as plotting each vector (k statistical parameters of each speckles image in our case) as a point in a k-dimension volume, and finding M-1 surfaces of dimension k-1 that separate between the M clusters of data. For example: in 2D (vectors containing only two parameters) belonging to M=3 classes, the algorithm would find the two lines that best separate the 3 clusters (scatterer types in our case) plotted in a 2D space.

In this work, for each of the N speckle pattern images collected, we calculated k=11 statistical parameters (described in Code 2, Ref. [6]). These include the speckles' contrast C and skewness S, and the following newly defined parameters extracted from $\log(G_I)$, where $G_I(u_x, u_y)$ is the power spectral density, as defined in Eq. (4) of the main text:

$$V = \frac{var\left(\log(G_I)\right)}{\log(G_I)^2} \tag{S1}$$

Five moments $\langle u_R^t \rangle$ for t = -4, -2, -1, 0, 1; defined as:

$$< u_R{}^t > \equiv \int_0^{L_x}\int_0^{L_y} \left[\delta + \sqrt{u_x^2 + u_y^2}\right]^t \times \log\left[G_I\left(u_x, u_y\right)\right] du_x du_y \tag{S2}$$

where δ is set to one for negative values of t and to zero otherwise (to avoid divergence), and three constants C_{max}, C_{width}, C_0, that define the function:

$$f \equiv C_{max} \times \exp\left(-\frac{u_x}{C_{width}}\right) + C_0 \tag{S3}$$

which are determined by the best fit of $f(u_x)$ to:

$$C(u_x) \equiv \frac{\log\left[G_I\left(u_x, u_y; u_y = 0\right)\right]}{var\left(G_I\right)} \tag{S4}$$

No Pol.	0.48	0.50	0.48	0.63	0.74	0.78	0.80	0.47	0.50	0.50	0.66	0.79
Perp. Pol.	0.51	0.48	0.44	0.61	0.73	0.73	0.80	0.54	0.43	0.57	0.72	0.74
Parallel Pol.	0.54	0.51	0.46	0.45	0.73	0.74	0.72	0.50	0.41	0.47	0.59	0.75
	C	S	$\langle u_R^{-4} \rangle$	$\langle u_R^{-2} \rangle$	$\langle u_R^{-1} \rangle$	$\langle u_R^{0} \rangle$	$\langle u_R^{1} \rangle$	C_{max}	C_{width}	C_0	V	Total Accuracy

Table S1. The LDA predictive rate when using only a single parameter to differentiate between white paint samples with different surface roughnesses using data from three sets: no polarizer and a polarizer perpendicular/parallel to the illumination's polarization.

These newly defined parameters are used to describe G_I (a 2D function containing $\sim 10^6$ data points) using only 9 scalars. Matlab was used to fit the function f (Eqn. S2) to the measured $C(u_x)$ (Eqn. S3) in order to extract: C_{max}, C_{width}, and C_0. All these parameters are then saved as a "parameter vector" representing the speckle pattern created from a specific point of a given material. Finally, all vectors are labeled according to the sample/polarization they represent. The importance and physical meaning of the contrast and the autocorrelation (and thus the G_I function) are described in the main text, but the rest of the measured parameters were chosen arbitrarily, and are shown to be of less value (Table S1).

Once the parameter vectors are extracted, we use them to test the classification algorithm's (described in Code 3, Ref. [7]) ability to differentiate between M data sets (M samples in a specific polarization). The algorithm starts by splitting the $N \cdot M$ parameter vectors into training and testing sets. 2/3 of the parameter vectors are used to train a Linear Discriminant Analysis (LDA) tool built into Matlab [8] that dissects the parameter space into subsections, each representing one of the samples it examined. The remaining 1/3 are then used (without their labels) to test the algorithm's predictive accuracy. This is done by comparing each unlabeled vector to the LDA function independently (single-shot) and the algorithm predicts which material it originated from. Then the accuracy is defined as the algorithm's percentage of correct predictions.

In order to check the consistency of the measured predictive accuracy and to reduce its statistical variance, we repeat the train and test process 100 times. This is done using the same data, only split into different randomly chosen training and testing sets, and then averaging the predictive accuracy measured for each realization (we refer to this average accuracy – the predictive rate). This leads to the error bars of less than 1% for the results presented in Fig. 2 of the main article.

An indicative, yet approximate way to check the value of each parameter towards the LDA algorithm's success is by measuring the algorithm's predictive rate when using only a single parameter (Table S1). The most useful parameters for most measurements turned out to be the contrast C, $\langle u_R^{-1} \rangle$, $\langle u_R^{0} \rangle$, and $\langle u_R^{1} \rangle$. The contrast seemed extremely useful for samples made of different materials (unlike the data shown in Table S1), due to the contrast relation to the polarization mixing, which in turn depends on the material's scattering parameters.

Using multiple speckle images to identify a sample:
The LDA algorithm predicts the probabilities of a given parameter vector originating from a list of M known samples previously measured. In the single-shot mode, simply the sample with highest probability is predicted. In the multiple images case, the algorithm is fed with n unlabeled parameter vectors corresponding to n spots of the same sample. It separately predicts the probabilities of each of the n parameter vectors belonging to each of the M samples. Then, the probabilities predicted for each sample are summed up for all n parameter vectors, and the maximal sum corresponds to the predicted sample. A sum is a naive measure that proved effective, but probably can be improved.

Transmission Measurements:
In addition to the measurements in reflection reported in the main text, we have performed also experiments in transmission through scattering media. For a highly controlled and known medium thickness, the substance could be identifiable in transmission as well. This would be restricted to using samples thin enough for light to transmit through and is therefore a less convenient method to identify and study scatterers. Through observing the width of the autocorrelation of speckles created via transmission; we could see a clear inverse correlation between the material thickness and the speckle size. Also, the speckles' contrast seemed to decrease with material thickness, implying that the polarization ratio increased (Figure S1). This verified the assumption that the polarization ratio should rise with the material thickness for the relevant thickness scales we worked with. These effects are due to the amount of light beam expansion and loss of initial linear polarization, both caused by multiple scattering which increases with the diffuser thickness. By recording the speckle patterns created via transmission through a known substance, we were able to remotely predict its thickness (Figure S1(a)) with high accuracy.

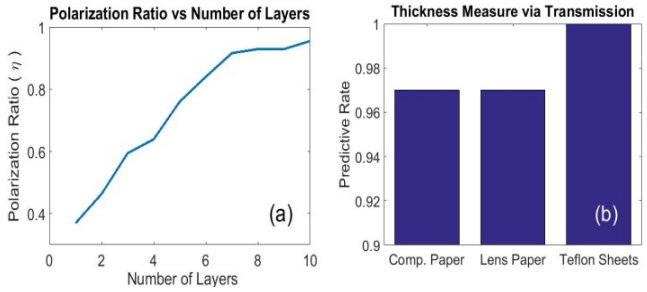

Fig. S1. (a) The polarization ratio (I_y/I_x) as a function of layers of lens papers. (b) Comparing between 4 data sets for each substance: {1,2,3,4} layers of computer paper and Teflon sheets, and {2,4,6,8} layers of lens paper.

References

1. M. Egmont-Petersena, D. Ridderb, H. Handelsc, Pattern Recognition 35, 2279 (2002).
2. E. Valent, "CNN to predict materials from their corresponding speckles", figshare (2018), https://doi .org/10.6084/m9.figshare.5777598.
3. https://keras.io/getting-started/sequential-model-guide/#examples
4. M. D. Zeiler, arXiv:1212.5701v1M (2012).
5. G. McLachlan, Discriminant Analysis and Statistical Pattern Recognition (John Wiley & Sons, 2004).
6. E. Valent, "Parameter vectors from speckle images", figshare (2018), https://doi.org/10.6084/m9 .figshare.5777595.
7. E. Valent, "LDA to predict materials from their corresponding parameter vectors", figshare (2018), https://doi.org/10.6084/m9.figshare.5777601.
8. https://www.mathworks.com/help/stats/classificationdiscriminant-class.html.