Sensing Indoor Lighting Environments and Analysing Dimension Reduction for Identification

Tushar Routh, Nurani Saoda, Md Fazlay Rabbi Masum Billah, Nabeel Nasir, Bradford Campbell
University of Virginia, Charlottesville, Virginia, USA
{tkr2w,saoda,masum,nabeeln,bradjc}@virginia.edu

Abstract

A generalized indoor light sensor can provide information to build and monitor indoor lighting arrangement that is aesthetically pleasing and conforming to the requirements set forth by the inhabitants. However, the identification of the surrounding lighting environment from the sensed parameters has some limitations and challenges. Till-to-date, classifiers are designed to identify only a single source, even in a multi-source environment. Classification based only on sensed values can be imperfect, as multi-type sources can share common parameters, or readings from a single source can fluctuate over time. The classification performances are mostly evaluated in controlled environments. In this work, we use a customised *Bluetooth Low Energy (BLE)* based light sensor that can sense and advertise major lighting parameters as instructed. Based on sensed parameters and adopting several Machine Learning (ML) and Neural Network (NN) based models off-board, we try to identify the singular and mixed presence of the four dissimilar types of sources: Incandescent, LED, CFL, and Sunlight in indoor surroundings. Off-board identification can get challenging where packet loss scenario is common. For that, we study how IoT devices with superior computational capability can utilise dimensional reduction techniques to minimize the required on-air traffic for classification. We then test classifiers with all those approaches both in controlled environments and real-world testbeds. The result shows that our best model can detect lighting environments with an accuracy of up to ***98.22%*** in the controlled scenario and ***83.33%*** in real-world testbeds.

CCS Concepts

• Computer systems organization → Embedded systems, Sensor networks.

Keywords

Light source classification, BLE with Machine Learning

ACM Reference Format:
Tushar Routh, Nurani Saoda, Md Fazlay Rabbi Masum Billah, Nabeel Nasir, Bradford Campbell, University of Virginia, Charlottesville, Virginia, USA, {tkr2w,saoda,masum,nabeeln,bradjc}@virginia.edu. 2023. Sensing Indoor Lighting Environments and Analysing Dimension Reduction for Identification. In *1st International Workshop on Advances in Environmental Sensing Systems for Smart Cities (EnvSys '23), June 18, 2023, Helsinki, Finland*. ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3597064.3597341

EnvSys '23, June 18, 2023, Helsinki, Finland

ACM ISBN 979-8-4007-0214-3/23/06…$15.00
https://doi.org/10.1145/3597064.3597341

1 Introduction

Indoor lighting has a substantial influence on determining people's frame of mind, alertness at work, and finally, circadian rhythm [9]. Daylong indoor mobility exposes inhabitants to natural and various types of artificial lights, which operate at specific wavelengths and generate illumination of contrasting features. A key requirement there is to first sense and then process the lighting parameters for the identification of surrounding lighting environments. Based on that, inhabitants can calculate the number of daily hours in blue-enriched sources (*like LED*) for correlating nighttime sleep quality with blue light hours or for documenting natural light disclosure period to meet the minimum daily recommended vitamin D generation [1, 8].

Apart from meeting the inhabitants' visual necessity, multi-source lighting environments are encouraged by experts for highlighting consumer products/artworks or controlling agro-environments [6, 7]. However, illumination from natural light can vary at different hours of the day or the quality of illumination from artificial sources can degrade over time. Even after arranging proper lighting setup, disruption can take place due to renovation, or repositioning of fitments/appliances due to various reasons. A generalized light sensor can help us not only monitor and maintain optimal conditions but also identify which combination of sources is responsible for bringing out a particular luminous environment.

For classification, offloading sensed parameters is always encouraged, as it allows deployment of light sensors that are low-powered, transfers major computational and memory-intensive classification tasks to the receiving side, and facilitates long term data storage and retrieval. One of the key challenges is to establish a communication strategy that can perform classification with minimal information. This is particularly essential for resource-bound sensing at enclosures that are overcrowded with multi-type sensors, where the packet loss scenario is common due to network congestion and communication bottleneck.

For minimizing traffic by air, we need to first figure out how to condense information for classification and what specific operation needs to get performed onboard. For memory/power constrained scenarios, sensors may have the capacity to operate for a certain period but can fail to perform all day long. Prior knowledge regarding methods to classify indoor lighting environments with minimal information and on-device resource requirements for such method adaptation, is therefore crucial.

By addressing these aforementioned concerns, we make the following key contributions to this paper:

- We conduct experiment to identify both single and multi-source environments, for which we *collect* standalone light parameter

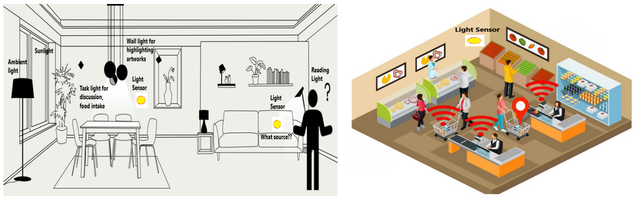


Figure 1: Illustration of our work: Identification of single/multi source environments (left), Efficient off-board source detection through dimension reduction in a crowded sensor surrounding (right).

data from the most common indoor light sources, (*Incandescent (Inc.), CFL, LED, and Sun*). We *utilise a customised Bluetooth Low Energy (BLE) enabled color sensing board* that can sense basic lighting parameters in the background [11]. For multi-source environments, we first *observe* multi-source effect by placing a sensor in between multi-type sources and *recreate* similar environments through blending sources of different types at various ratios. We then *divide* dataset into 10 classes (*Inc, CFL, Led, Sun, Inc-CFL, Inc-LED, Inc-Sun, CFL-Led, CFL-Sun, Led-Sun*). Later, we *implement* different Machine Learning (ML) and Neural Network (NN) based techniques for identifying sensed time varying RGB patterns. After *analyzing* performance under ideal conditions, we *select* the best-performing classifier, *observe* relationship between accuracy and blending ratios, and *record* classifiers' identifying performances under unfamiliar sources and smart on/off scenarios.

- We *investigate* different dimension reduction methods with the necessary number of principal components to minimize on-air traffic. Dimension reductions were accomplished using with and without original data reconstruction approaches. We *compare* them based on time, power, and memory requirements and *analyse* the performances of all the different techniques both at ideal and real-world scenarios (Figure 1).

2 Related Works

C12666MA mini-spectrometer from Hamamatsu electronics was used to study melanopic influence from different sources [2]. Necklace shape wearable spectrometer titled *Spectrace* was developed including motion elements for personal light monitoring [12]. RGB-based machine learning approach was also implemented. [3] utilized low cost *TCS3414CS color sensor*. Later they classify various artificial sources (34 LED, 16 incandescent and 6 fluorescent sources).Alejandro et. al Ma et. al [5] did similar kind of research with *TSL2561, ISL29125* color sensors, *AM1815CA, POW11D2P* solar cells and *USB2000+* spectrometer, where sensor data for Halogen, Fluorescent, LED and Incandescent bulbs were collected via USB interface and I-V tracers and KNN, SVM and Decision tree algorithms were utilised.

3 Limitations of earlier approaches

Information from spectrum analyzers can be used for source classification. However, they are expensive, high power consuming and rely on memory-intensive high-resolution spectral information. Classification based on a specific spectral window is incorrect, as the same source can exhibit different spectra throughout the day/different phases after lightening up and different classes can

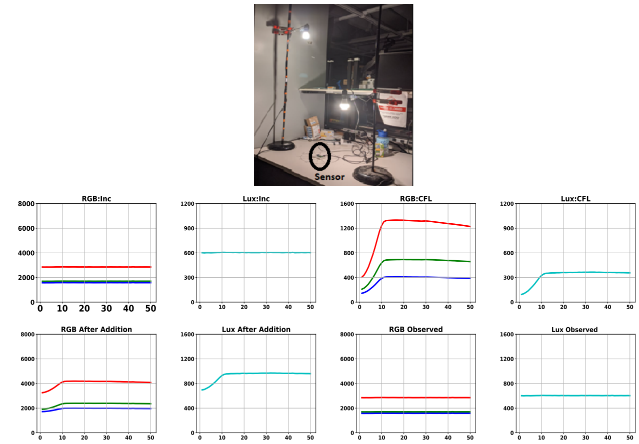


Figure 2: Observing 50 recorded RGB and Lux values in a multi-source environment, where Inc was set as primary (60cm from sensor) and CFL as secondary source (90 cm from sensor). Result deviates from simple arithmetic addition.(*x-axis*: Number of Samples, *y-axis*: Recorded RGB and Lux intensity in Lux/m^2 values in decimal).

share a common spectral band. With RGB color-sensitive sensors, sources were differentiated based on color component magnitudes. Nonetheless, different sources can share common magnitudes based on the distance between sensor/source placement. Sources were thought to exhibit specific RGB patterns. However, such patterns are only observed during switching.

Machine Learning (ML) based classifiers were utilised, but the effectiveness of such classifiers was barely examined outside the training set or smart on/off environment for universal deployment, where RGB signals can consist of transient shapes from missing samples and can significantly differ from regular patterns.

With the superior memory and calculative power of modern IoT edge devices, dimension reduction techniques can be adapted to facilitate the minimization of heavy on-air workload. However, such techniques may require a trade-off with accuracy, where a systematic investigation is missing.

4 Methodology

Different distances bring various structures and elements into the scene as reflecting elements, which create a complex signal, where readings are not inversely proportional to distance. We vary distances from 50cm to 150cm and set the advertising interval to 25 ms for rapid acquisition. The customised board advertises *BLE* data packet containing device ID, the number of the latest packet, and raw data (clear, red, green, and blue) split into two bytes per color [11]. A *Bluegiga BLED112 Bluetooth Smart Module* was set as a receiver with a distant computer through which we collect 500 samples for each observation. **Based on [10], we consider 25 successive samples containing only RGB information to capture varying features of different source types for classification.**

4.1 Realistic data generation and classification

To capture wide variation for each class of source, we have included 27 different bulbs (9 from each of LED, Incandescent, and CFL) on dark environments and in sunlight at different weather conditions/times of the day, with/without window blinds and various indoor corners at different buildings. For multi-class scenarios, we

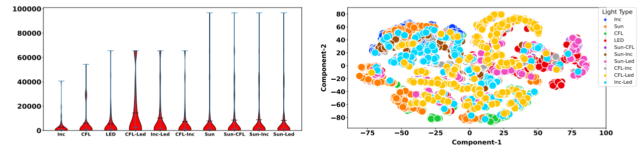


Figure 3: Violin plot showing red color component distribution of training data [top] (x-axis:source type, y-axis: recorded values). As observed, different classes generate common readings and cannot be segregated solely based on threshold setting. First two principal components of 2D tSNE plot exhibits linear inseparability of single and multi-type source environments [bottom].

first place two sources of different types and observe their resultant on the sensor point (Figure 2).

In a multi-source environment, when nearby sources interact, the result depends on different measurements: (a) the intensity of each source, (b) the relative distance between two sources, and (c) the positioning of the sensor. As a result, a simple addition of RGB values to mimic a multi-source environment is inaccurate, as shown in figure 2. We first fix the position of the sensor and the sources and mix signals from different light sources at various ratios where the secondary signal intensity went down to 20% of the primary. To encompass a variation of phase, we consider the extremes of constructive and destructive interference in our dataset, as these cases contain the highest deformations where the possibilities of misclassifications are the greatest.

Figure 3 (left) shows the red color component distribution of all samples. As depicted, different sources/scenarios share common values, which were also observed for Green and Blue readings. As a result, classification based on magnitudes is inaccurate. Two-dimensional tSNE visualization of RGB values depict linear inseparability of classes (Figure 3) (right).

For classification, we investigate ML-based classifiers which are simple, high-performing, memory friendly, and easily transferable on embedded devices. From ML based classifiers, we pick Decision Tree (DT), Random Forest (RF), Gradient Boost (GB), Gaussian Naive Bias (NB), K-Nearest Neighbour (KNN), Logistic Regression (LR), and Support Vector Machines with linear (SVM-Lin), radial (SVM-Rad) and polynomial (SVM-poly) kernels. For reputation in identifying unseen pattern-based examples, we also study NN-based Feedforward Multilayer Perceptron Model (MLP), Convolutional Neural Networks (with 1-D and 2-D filters) and Long Short Term Memory (LSTM) for classification.

After collecting sensor data, we have **scaled, normalized, and divided the balanced dataset** of all 25 samples' windows into training, test and validation sets (80%, 10% and 10% respectively). For better evaluation of the built model and to ensure representation from each group, we have implied stratified 10-fold cross-validation by tuning aforementioned models to their best hyperparameters using *Gridsearch*.

4.2 Dimension reduction

Classification with 25 consecutive samples, each with 3-channel information can be alternatively considered as 75-dimensional data. This high-dimensional data can be reconstructed using fewer dimensions, with necessary information packed inside. With the modified dimensions, compressed RGB data has elevated robustness, and

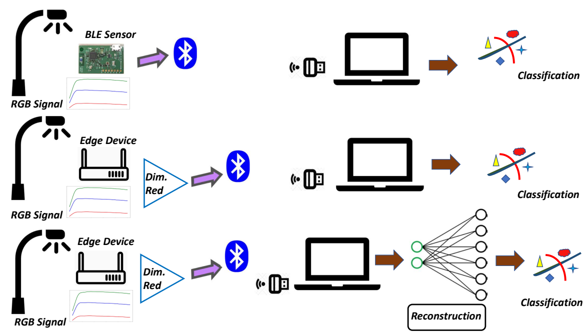


Figure 4: Different classification approaches: Without dimension reduction with our BLE sensor (top), with reduction by method 1 (middle) and method 2 (bottom)

can be advertised even in a single packet with repetition if necessary. This process can significantly truncate the operational power budget. In this work, we analyze classification by compressing (dimension reduction) sensed RGB data using linear and non-linear compression techniques, record their accuracy and compare the result with the uncompressed method (Figure 4). Two different approaches were analysed for classification with compressed information, as shown in Table 1. To the best of our knowledge, ***dimension reduction to minimize on-air traffic for indoor light sensing*** has been implemented for the first time.

For dimension reduction, we choose six different methods: Principal Component Analysis (PCA), Truncated Singular Value Decomposition (Trun-SVD), Linear Discriminant Analysis (LDA), Kernel-PCA with cosine function (KPCA), ISOMAP, and tSNE. Our goal is to select the best classifier among all scenarios. We start by taking only two principal components with method 1, apply our best-performing classifier KNN for classification and compare accuracy with familiar sources, smart on-off scenarios ,and with unfamiliar sources. We select PCA from linear and Kernel PCA from non-linear technique and consider dimensions up to 10 principal components (based on Figure 5). We then classify with reduced dimension utilizing *method 1* (PCA/KPCA) and *method 2* [PCA-Reconstructing from PCA (RPCA)/ KPCA-Reconstructing from Kernel-PCA (RKPCA)].

Finally, we deploy our sensor in multiple testbeds. Scenarios include multi-type unfamiliar bulbs with random switch-overs/window

Table 1: Approaches for Dimension Reduction

Method 1	***Method 2***
Step 1: Perform dimension reduction on training RGB data, tuning ML models and storing classifier model at receiver	*Step 1*: Storing classifier model at receiver after tuning with original training data (25 samples consisting RGB info)
Step 2: Store specific number of RGB samples, perform dimension reduction on stored data and advertising	*Step 2*: Store specific number of RGB samples, perform dimension reduction on stored data and advertising
Step 3: After receiving advertised packets, we feed the input values as input to the classifier	*Step 3*: After receiving advertised packets, first we reconstruct data to get back the original dimension, then we feed values as input to the classifier

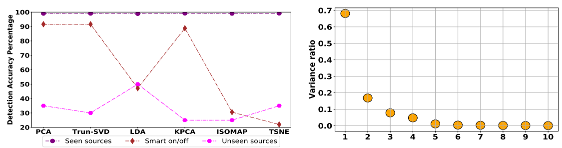


Figure 5: Comparing reduction techniques performances reveals superiority of PCA for linear and KPCA for non linear classification (left), PCA over collected data reveals overall variance lies within first few principal components (x axis: no of principal components, y-axis: variation ratio) (right)

blinds on/off, the presence of other indoor smart sensors and arbitrary switching. The workflow of the whole process is shown in figure 6.

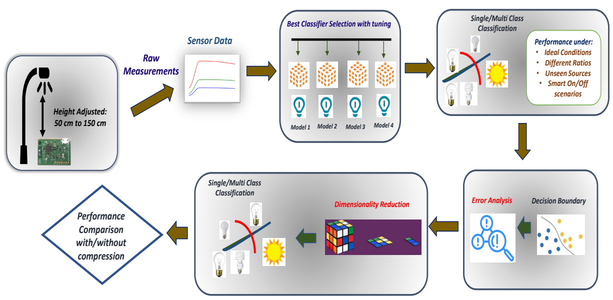
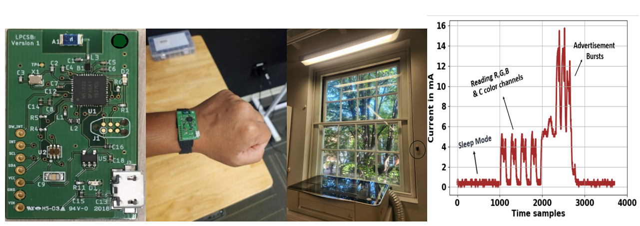


Figure 6: Workflow of this paper.

5 Data Collection Hardware

Our utilised BLE sensor is a printed circuit board (PCB) that interfaces *TCS3475* sensor and is regulated with *nRF51822* micro controller. It communicates over 2.4GHz Bluetooth Low Energy (BLE) and has a dimension of roughly $24mm \times 39.5mm$. *nRF51822* micro controller has been reduced to only contain clock circuits, 3.3 V regulatory circuitry and a power supply connector for minimum energy consumption. This device senses the surrounding light and stores red-filtered, green-filtered, blue-filtered, and clear (unfiltered) diodes data of *TCS34725* as a 16-bit value, split between two registers. *TCS34725* was chosen as it offers low operating voltage *(2.7-3.6 V)* and higher sensitivity *(dynamic range: 3.8M:1)*. Figure 7 depicts a BLE advertisement event. The per unit cost of the whole package is nearly \$10-\$15, compared to several hundred dollars for wearable spectrometers. With the setup of recording one observation per second, it can run up to 45 days with 3.3V conventional lithium coin battery [10].



(a) (b) (c) (d)

Figure 7: Fully assembled *BLE* enabled customised Low Power Color Sensing Board (a), Board Dimension is suitable to deploy as a handheld (b) or fixed point device (encircled) (c). Color Sensing followed with an advertisement event (d).

6 Evaluation

We analyse prediction accuracy in different backgrounds and record mean values of classifying accuracy (with standard deviations).

6.1 Performance of classifiers

Figure 8 (top left) illustrates the variability of accuracy for different classification techniques with deviations. As observed, overall performance of ML algorithms is better than NNs in the known scenarios. The best result with maximum average accuracy (up to ***98.2%***, F1 score:***0.97***) with minimum deviation was recorded with KNN.

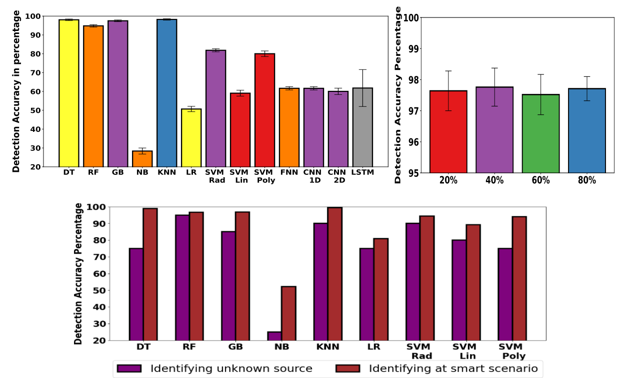


Figure 8: Comparing accuracy among different ML and NN architectures(top left), Accuracy with KNN after mixing two sources at various proportion reveal disassociation of accuracy with mixing ratios (top right, Variation of accuracy for tuned ML classifiers at smart env. and with unseen sources (bottom).

Now we observe whether the blending ratio has any impact on the accuracy. As observed (Figure 8 (top right)), accuracy is not proportional to the mixture ratio of multiple sources. With tuned ML classifiers, we then examine how each classifier performs at smart on/off environments and at identifying sources outside training set. As seen, KNN performs the best at classifying in smart scenarios whereas Random Forest (RF) is the highest performer at identifying unknown sources (Figure 8 (bottom)) . We continue with KNN for being the overall best classifier.

Our tests with different methods and various scenarios exhibit that classification accuracy varies with the varying principal components (Figure 9). Finally, we deploy our sensor at two different real-world testbeds that include single/multi-source scenarios with completely unfamiliar indoor bulbs, random switching from one type to another and human movements. Based on the results in all scenarios in Figure 9 and classification with minimal information, we consider 2 principal components with PCA, KPCA, PCA-RPCA, and KPCA- RKPCA and apply KNN algorithm for classification. We record accuracy with/without reduction methods (Figure 10). As seen, after real-world deployment, the classification accuracy degraded significantly from ***98.22%*** down up to ***77.5***%. Again, incorrect predictions could not be specified for any single/multi-source type and noticeably, some of them occurred during switch-overs/movements, as the signal patterns were significantly different

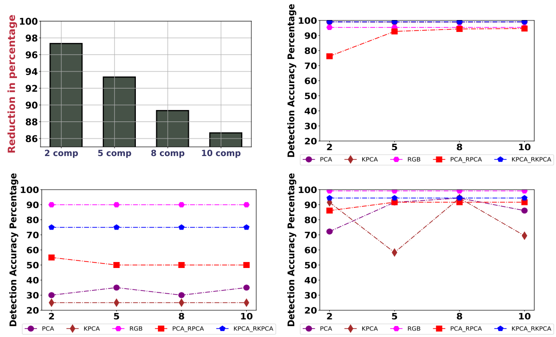
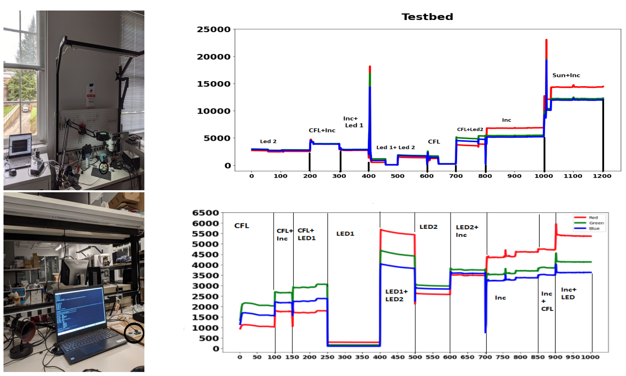


Figure 9: Reduction of on-air traffic for advertising only principal components compare to advertising 25 RGB samples (x-axis:no. of principal components, y-axis: Reduction Percentage), Comparing accuracy: familiar (top right), unfamiliar (bottom left) and smart on/off scenario (bottom right)(x-axis:no. of principal components, y-axis: accuracy).



RGB	***PCA***	***KPCA***	***PCA-RPCA***	***KPCA-RKPCA***
83.33%	41.67%	66.67%	58.83%	83.33%
77.5%	40.0%	60.0%	52.5%	77.5%

Figure 10: RGB readings at two different testbeds (circle point placement of sensor, (x-axis: sample no., y-axis= recorded values). Performance of KPCA/RKPCA was found comparable to RGB value based classification.

at those instances. However, ***KPCA-RKPCA*** accuracy in both cases was comparable to accuracy with 25 RGB samples. For general comparison, we run dimension reduction on two different edge devices. Along with recording the highest accuracy, KPCA-RKPCA method demands the highest time, memory, and power to operate (Figure 11). Compare to PCA, KPCA constructs a full $n \times n$ kernel matrix over n data points to allow for nonlinear relations, which demands extra resources for execution [4].

6.2 Discussion

Data acquisition: Integration time and ADC gain of BLE sensor settings can be modified for superior sensitivity at low light levels. Sampling and advertising rates can also be readjusted for better power management on the transmitting side. Our analyses were sensor specific, which has limitations regarding the highest value

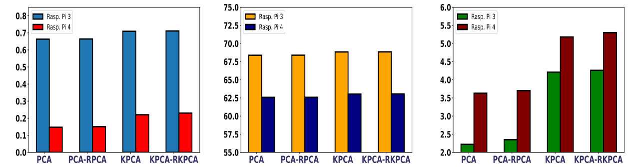


Figure 11: Overall Comparison exhibits KPCA-RKPCA method requires the highest time (left) [y-axis= seconds], memory (middle) [y-axis= MB] and power(right)[y-axis=Watt] for execution.

it can record or the functional spectra, especially for natural light. Similar accuracy is expected from light sensors that have different sensitivity and operating spectra *(like AMS AS726x sensors)*, as values considered for analysis were normalised and algorithms for source recognition were pattern based.

Elevating performance: Synthetic time series examples can be generated from the current data distribution to improve the identification of unknown sources. To mimic random changing of bulbs or smart on/off cases, we can pass regular RGB signals through different filters and then we can include them in the training set to familiarize our classifiers with such instances. Statistical features of time series observations may be used for classification, but that will require additional power, memory, and time. While running the control tests, we have not considered multi-source environments that contain more than two sources, colored glass sources, or searchlights in our training set.

7 Conclusion and Future work

Our work demonstrates BLE based RGB sensor, coupled with Machine Learning can be used for identifying single/multi-source surroundings. Analyses also reveal where transmitters can be operated with additional capacity, low dimensional data can be used alternatively for classifications, where reduction/reconstruction (*method 2*) was found better performing than only reduction (*method 1*). However, identification accuracy in familiar environments with non-ideal scenarios and in a completely unknown environments with real-world interruptions can degrade.

Our future endeavors include modification of sensor parameters to gather knowledge regarding indoor illumination more quickly, power efficiently and with better accuracy after real-world deployments. We plan to further investigate how we can utilize the computational capability of existing devices for dimension reduction and build an intelligent system that can execute classification based on surroundings. Finally, as time series signals, feature-based classification can be considered but will require extra memory and processing power.

8 Acknowledgements

We thank the anonymous reviewers and our shepherd for their valuable insights and feedback on improving this paper. Our special thanks goes to *NobleReach Foundation* for supporting this work.

References

[1] Timothy M Brown, George C Brainard, Christian Cajochen, Charles A Czeisler, John P Hanifin, Steven W Lockley, Robert J Lucas, Mirjam Münch, John B O'Hagan, Stuart N Peirson, et al. 2022. Recommendations for daytime, evening, and nighttime indoor light exposure to best support physiology, sleep, and wakefulness in healthy adults. *PLoS biology* 20, 3 (2022), e3001571.

[2] Sean W Cain, Elise M McGlashan, Parisa Vidafar, Jona Mustafovska, Simon PN Curran, Xirun Wang, Anas Mohamed, Vineetha Kalavally, and Andrew JK Phillips. 2020. Evening home lighting adversely impacts the circadian system and sleep. *Scientific reports* 10, 1 (2020), 1–10.

[3] Alejandro Fernández-Montes, Luis Gonzalez-Abril, Juan A Ortega, and Francisco Velasco Morente. 2009. A study on saving energy in artificial lighting by making smart use of wireless sensor networks and actuators. *IEEE network* 23, 6 (2009), 16–20.

[4] Mina Ghashami, Daniel J Perry, and Jeff Phillips. 2016. Streaming kernel principal component analysis. In *Artificial intelligence and statistics*. PMLR, 1365–1374.

[5] Xinyu Ma, Sebastian Bader, and Bengt Oelmann. 2017. Characterization of indoor light conditions by light source classification. *IEEE Sensors Journal* 17, 12 (2017), 3884–3891.

[6] Francesco Murano et al. 2016. LIGHTING ARTWORKS IN ART EXIBITIONS. (2016).

[7] Darren Chi Jin Neo, Maxine Min Xin Ong, Yeong Yuh Lee, Ee Jin Teo, Qunya Ong, Hendrix Tanoto, Jianwei Xu, Kian Soo Ong, and Vignesh Suresh. 2022. Shaping and tuning lighting conditions in controlled environment agriculture: A review. *ACS Agricultural Science & Technology* 2, 1 (2022), 3–16.

[8] NIH. 2021. *Vitamin D Fact Sheet for Health Professionals*. Technical Report. National Institute of Health.

[9] Swathi Matta Reddy, Debkumar Chakrabarti, and Sougata Karmakar. 2012. Emotion and interior space design: an ergonomic perspective. *Work* 41, Supplement 1 (2012), 1072–1078.

[10] Tushar K Routh, Nurani Saoda, Md Fazlay Rabbi Masum Billah, and Brad Campbell. 2022. Low Cost Light Source Identification in Real World Settings. In *2022 19th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON) (SECON 2022)*. virtual, USA.

[11] Alexander Sarris. 2020. *LPCSB: A Device to Distinguish Between Natural and Artificial Light*. Master's thesis. University of Virginia, Charlottesville, Virginia.

[12] Forrest Simon Webler, Giorgia Chinazzo, and Marilyne Andersen. 2021. Towards a wearable sensor for spectrally-resolved personal light monitoring. In *Journal of Physics: Conference Series*, Vol. 2042. IOP Publishing, 012120.