A Programmable Optical Phantom to Train Unsupervised Machine Learning Algorithms for Kinetic Imaging

Helena ARIAS[1], Chao CHENG[2], Stanley MARKMAN[2], Marvin RAISER[3]

Spain[1], United States[2], Germany[3]

Dr. Vyacheslav Kalchenko

In Vivo Optical Unit, Department of Veterinary Resources, Faculty of Biology, Weizmann Institute of Science, 76100 Rehovot, Israel

Abstract

This paper presents a novel optical phantom for testing unsupervised machine learning algorithms utilized in kinetic fluorescent imaging analysis. We propose a new technique, as an alternative to injecting mice with fluorescent material, in which fluorescence-simulating LEDs shine through an artificial optical phantom based on an anatomically accurate mouse model. The data collected from this technique is analyzed by optical flow segmentation algorithms, which are then evaluated to determine the best-performing algorithm.

Introduction

Biomedical research is a broad area of science that studies the behavior of chemical substances in animals and humans. Preclinical trials are the first stage in evaluating the effectiveness of a specific drug in both healthy and sick subjects. Studies on mice are widespread due to the similarity of their genetic composition to humans and other mammals, which results in sacrificing a certain number of these animals.

One useful tool for preclinical studies is optical modality. For example, fluorescence optical imaging is an analytical method based on the study of the behaviour of a fluorescent material when it is injected into the body. The fluorescent dye flows via the blood vessels throughout the entire body, showing anomalies such as tumors. This method presents several advantages over the widely-used PET and CT scan methods. While PET and CAT scans are time-consuming processes that result in a still image, fluorescence imaging can be used to produce kinetic images which can provide researchers and medical professionals with a greater understanding of possible anomalies within the body [1]. Furthermore, while PET, CT, and X-ray scans produce a quantity of radiation that can be harmful to a patient or lab animal, optical

fluorescence imaging produces no radiation and can therefore be used frequently without the possibility of inducing cancer [2, 3]. However, issues such as light scattering, optical noise, and other distortions have long prevented visible light spectrum and near-infrared fluorescence imaging from being widely used as a diagnostic tool, and have impaired its usability in preclinical studies. One recent approach towards improving optical imaging is machine learning. Both supervised and unsupervised algorithms have strong potential applications in improving fluorescence imaging accuracy. Due to the fact that supervised machine learning algorithms require massive amounts of data to train themselves to a particular problem-set, and both supervised and unsupervised algorithms require diverse datasets for testing and development, it is unfeasible to use live lab animals in the development of these algorithms.

Despite the significant amount of animals needed for preclinical trials there are both ethical and practical reasons to reduce or eliminate the usage of live animals in these studies.

In this project, we aimed to entirely replace the usage of lab animals in fluorescence kinetic imaging studies, while creating an efficient and accurate method for testing machine learning algorithms which are used to analyze fluorescence imaging data. Therefore, in order to simulate the properties of fluorescent dye and tissue scattering, phantom mice of varying compositions were produced, containing programmable LEDs to model optical flow. The low-cost materials such as silicone, glycerin, and inexpensive dye used in the fabrication of the phantom allowed for a wide variety of phantoms simulating different optical properties to be rapidly produced.

This adaptable optical phantom which can closely model the optical properties of animal tissue is therefore an essential tool for the further development of fluorescence imaging techniques.

Materials and methods

Fabrication of the Optical Phantom

The phantom was built using a molding and casting process. An anatomically accurate plastic 3-D printed mouse model from Prof. Igor Meglinski of Oulu University (Oulu, Finland) was used to produce a silicone mold. Then, the mold was used to create optical phantom models with a wide variety of compositions. A phantom composed of silicone and potato flour was first created (Phantom 1). Next, a phantom composed of a gelatin-water mixture in a 50/50 ratio (Phantom 2), and variations of the gelatin-water mouse containing black ink, red ink, and zinc oxide were produced (*Figure 1*).

Figure 1. Examples of different gelatin based optical phantoms

Imaging Setup

The electronic system consisted of four resistors and two LEDs. The circuit shown in *Figure 2* was connected to the 5V analog outputs of a NI USB-6009 device.

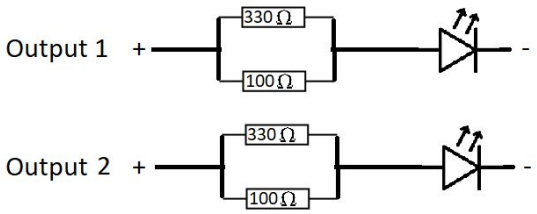

Figure 2. An electronic circuit of two LEDs light source that were connected to the analog outputs of a NI USB-6009 device.

In order to control the outputs, we designed two sinusoidal functions with LabView 2018 that can be easily corrected by defining the time gap between their relative maximums (*Figure 3*). On the first stage, we used a phase difference of 2,56s and a distance of 5 cm between the LEDs; the first pictures were taken with an Ueye IDS 4103105867 camera at 30 frames per second.

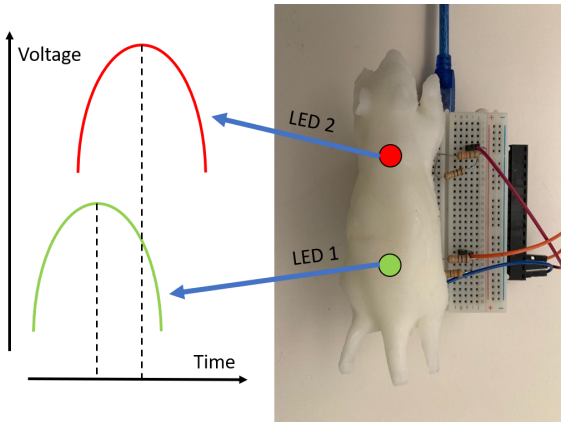

Figure 3. Schematic representation of computer-controlled (silicone) optical phantom setup simulating optical properties of mouse tissue.

To improve the quality of the images, a black layer was added between the mouse and the circuit. In addition, the camera was substituted by the model IDS UI-3240LE-NIR-GL with a 25mm 1:1,4 2/3' Computar lens. *Figure 4* shows the images taken at this stage.

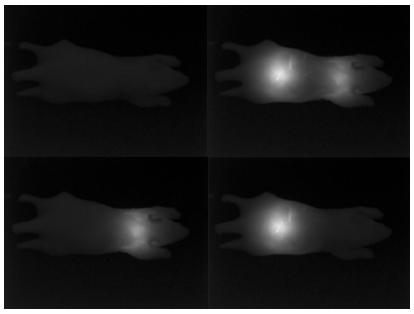

Figure 4. Monochrome image of Phantom 1 (Phantom 1.1, four frames from the original video): the dark background improves the quality of the pictures whilst the diaphragm of the camera helps focus on the bright parts of the images.

For the last data taken from Phantom 1 (Phantom 1.3) conditions of Phantom 1.2 were maintained with the exception of the phase difference, which was shortened to 0,9s.
Finally, the gelatine mouse (Phantom 2) was tested. Using the same settings as in Phantom 1.3 in exception of the LEDs distance (2cm), the goal of this video is to test the algorithm with more light scattering. The variance of the images can be seen in *Figure 5*.

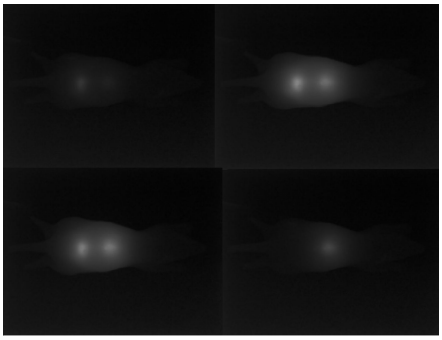

Figure 5. Phantom 2 (four frames from the video): the gelatine reduces the scattering, simulating more accurately the light behaviour inside a mouse.

The collected images were preprocessed in ImageJ to 8-bit depth and cropped and resized to 192 x 153 pixels.

Software

The segmentation of the recorded images was done using unsupervised machine learning algorithms from the scikit-learn library (v.0.21.2) in Python 3.7.4. Unlike supervised learning, which requires an enormous amount of data, unsupervised learning does not require training data and also has the added benefit of being adaptable to various scenes; in other words, each data point does not have to be recorded under the same conditions. The tested algorithms were: Principal Component Analysis (PCA) [4], Sparse Principal Component Analysis (SPCA) [5], Truncated Singular Value Decomposition (TSVD) [6], Factor Analysis (FA) [7], Independent Component Analysis (ICA) [8], Non-negative Matrix Factorization (NMF) [9], Latent Dirichlet Allocation (LDA) [10].

Every algorithm gets transposed and then scaled image data from 0 to 1 to fit them in order to segmentate each component. After the fitting of the algorithms the segmented components will be returned as a graph with their specific appearance on a scale of frames (relative magnitude in y-axis, number of frame in x-axis).

Results

The data collected from each phantom was analyzed using the unsupervised algorithms, which were programmed to find three components. Each generated component highlights the optical flow of an LED and produces both a still image and a light curve. The white region within the still image represents the isolated component. The horizontal axis of the light curve represents the flow of time in frames, and the vertical axis represents the relative magnitude of a component, which in turn reflects the brightness of the component in relation to an arbitrary reference point; in other words, a higher y-value indicates a stronger luminescence, while a lower y-value indicates a weaker luminescence.

One of the most remarkable results is the result of running the Sparse Principal Component Analysis algorithm on Phantom 1.2, as shown in *Figure 6*. Component A corresponds to the left LED, while component C represents the right LED. The algorithm also recognises the shape of the phantom as Component B.

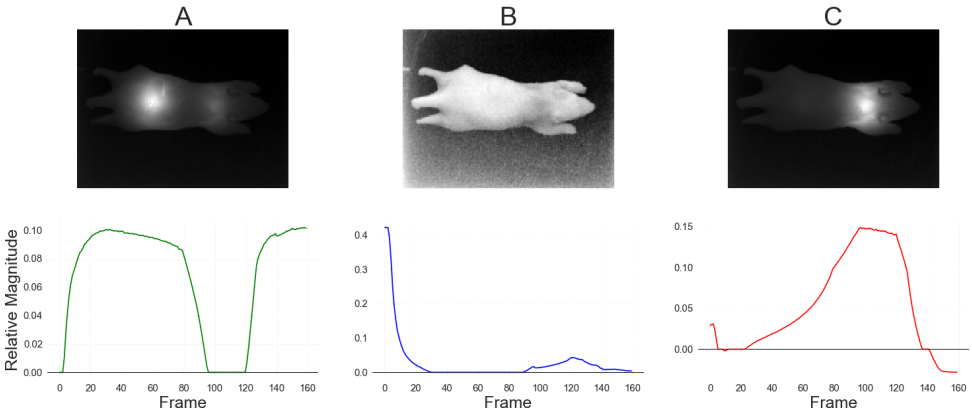

Figure 6: SPCA in Phantom 1.2; A is the left LED; B is background noise; C is the right LED

Another discovery is the observation of a correlation between the original function and the final result, as seen in *Figure 7*. The graph of the combined light curves closely resembles the simulated optical flow emitted by the LEDs, due to the algorithm's ability to precisely differentiate between the individual optical signals from each component.

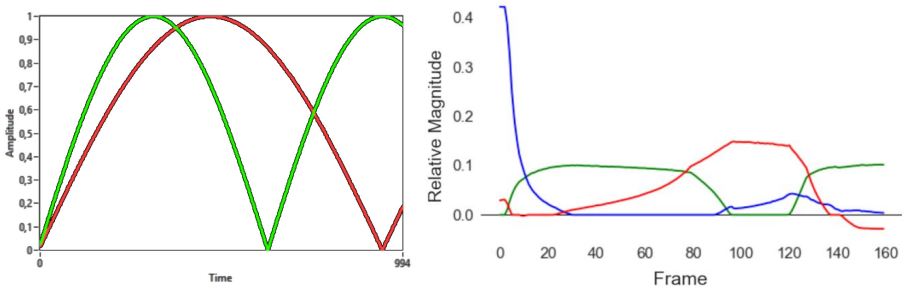

Figure 7. left: LED curve; right: Composition of the component presence curve over the frames.

Despite the small distance between the LEDs in Phantom 2 and the different optical conditions of the material, the Independent Component Analysis (ICA) algorithm reliably generated accurately separated component images and graphs — see *Figure 8* and *Figure 9*.

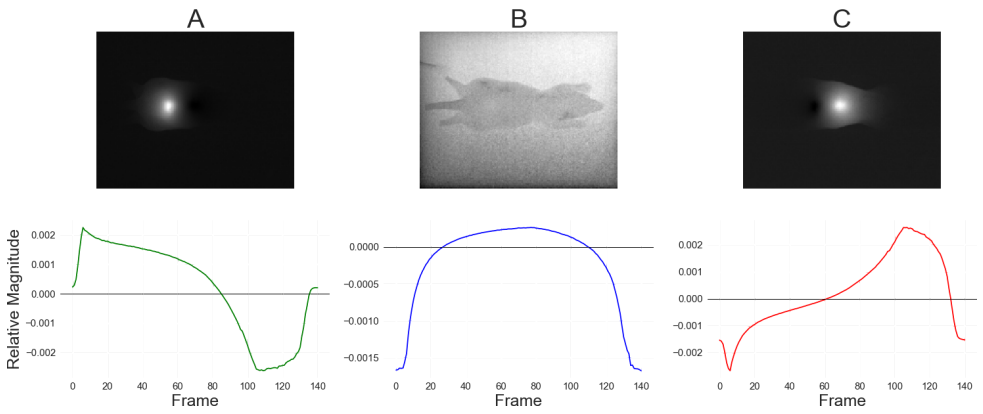

Figure 8: ICA in Phantom 2; A is the left LED; B is background noise; C is the right LED.

Figure 9. Composite color coded image obtained after the processing of algorithm.

Discussion

In *Table 1*, the algorithms were evaluated on a 0-5 point scale according to their ability to differentiate the LED outputs, the accuracy of the component light curve, and the presence of noise in the components.

	Phantom 1.1	Phantom 1.2	Phantom 1.3	Phantom 2	TOTAL
ICA	4	3	5	5	17
SPCA	4	5	4	2	15
NMF	5	3	4	2	14
FA	4	2	3	2	11
PCA	4	2	2	2	10
TSVD	4	2	2	2	10
LDA	0	3	3	0	6

Table 1. Comparison of Algorithms. 0) Fail; 1) Differentiate one LED with vague light curve and noise; 2) Differentiate one LED with accurate light curve and no noise; 3) Differentiate both LEDs with vague curves and noise; 4) Differentiate both LEDs with accurate light curves and little noise; 5) Differentiate both LEDs with accurate light curves and no noise.

ICA consistently produces accurate results with indiscernible noise while differentiating the two LEDs. SPCA and NMF both produce accurate results for each version of Phantom 1, while failing to do so for Phantom 2, the gelatin mouse. When the data from Phantom 1.1 is fed to NMF, the algorithm is able to separate the left and right LED signals into components B and C, respectively, and produces light curves nearly identical to the programmed intensity functions for the LEDs. Likewise, running SPCA on the data from Phantom 1.2 results in the separation of the left LED and right LED into distinct components C and B, respectively, with a third component A capturing the vast majority of noise from the video. However, neither NMF nor SPCA are able to separate the two LED signals shining through the gelatin mouse upon being given the data from Phantom 2, whereas ICA successfully and accurately differentiate the signals into components A and C with negligible noise. Many algorithms, such as PCA and TSVD, produce lower-quality results. For instance, the output of running PCA on the data from Phantom 1.3 results in partial differentiation of the LED signals; while the right LED signal is isolated to Component C, the algorithm is unable to isolate the left LED signal, which results in Component B containing signals from both the left and right LEDs. In other scenarios, many algorithms produce results with a large amount of noise in the components. Regarding the accuracy and quality of the images, it can be concluded that gelatin, which is more similar to the mice tissue, reduces the light scattering. However, the variation in distance of the different LEDs affects the effectivity of the algorithms, being the images with the lowest distance between LEDs, the most difficult to be correctly analyzed by the algorithms.

Conclusion

We conclude that a kinetic programmable optical phantom can be used as a replacement for living mice in the development of optical analysis technologies. This method has the potential to be a cost-effective and humane approach to collecting data for preclinical studies, reducing or even eliminating the need for sacrificing large numbers of mice in the name of scientific research. Furthermore, we conclude that Independent Component Analysis is the best-performing algorithm due to its ability to successfully and accurately differentiate the left and right LED signals in each version of the phantom into two components with negligible noise.

Acknowledgement

We would like to thank our mentor, Dr. Vyacheslav Kalchenko, for his time and patience, as well as Dorit, Nirit, the International Summer Science Institute (ISSI) counselors and our ISSI colleagues for their support. Additionally, we would like to thank Fundació Catalunya - La Pedrera, the American Committee for the Weizmann Institute of Science, the Deutschen Gesellschaft der Freunde des Weizmann Instituts, and all the supporters of these institutions for giving us the opportunity to be a part of this project.

References

1. Price, Julie C. "Principles of tracer kinetic analysis." *Neuroimaging Clinics* 13.4 (2003): 689-704.

2. M. Yang, E. Baranov, P. Jiang, F. X. Sun, X. M. Li, L. Li, S. Hasegawa, M. Bouvet, M. Al-Tuwaijri, T. Chishima, H. Shimada, A. R. Moossa, S. Penman, and R. M. Hoffman, "Whole-body optical imaging of green fluorescent protein-expressing tumors and metastases", *Proceedings of the National Academy of Sciences* Feb 2000, 97 (3) 1206-1211.

3. Lakowicz, Joseph R., ed. "Principles of fluorescence spectroscopy". *Springer Science & Business Media*, 2013.

4.Abdi, Hervé, and Lynne J. Williams. "Principal component analysis." *Wiley interdisciplinary reviews: computational statistics* 2.4 (2010): 433-459.

5. Zou, Hui, Trevor Hastie, and Robert Tibshirani. "Sparse principal component analysis." *Journal of computational and graphical statistics* 15.2 (2006): 265-286.

6. Maciejewski, Anthony A., and Charles A. Klein. "The singular value decomposition: Computation and applications to robotics." *The International journal of robotics research* 8.6 (1989): 63-79.

7. Harman, Harry H. "Modern factor analysis." (1960).

8. Hyvärinen, Aapo, and Erkki Oja. "Independent component analysis: algorithms and applications." *Neural networks* 13.4-5 (2000): 411-430.

9. Lee, Daniel D., and H. Sebastian Seung. "Algorithms for non-negative matrix factorization." *Advances in neural information processing systems*. 2001.

10. Blei, David M., Andrew Y. Ng, and Michael I. Jordan. "Latent dirichlet allocation." *Journal of Machine Learning research,* 3.Jan (2003): 993-1022.