Principal component analysis of dynamic fluorescence images for diagnosis of diabetic vasculopathy

Jihye Seo
Yuri An
Jungsul Lee
Taeyun Ku
Yujung Kang
Chulwoo Ahn
Chulhee Choi
Principal component analysis of dynamic fluorescence images for diagnosis of diabetic vasculopathy

Jihye Seo, Yuri An, Jungsul Lee, Taeyun Ku, Yujung Kang, Chulwoo Ahn, and Chulhee Choi

Abstract. Indocyanine green (ICG) fluorescence imaging has been clinically used for noninvasive visualizations of vascular structures. We have previously developed a diagnostic system based on dynamic ICG fluorescence imaging for sensitive detection of vascular disorders. However, because high-dimensional raw data were used, the analysis of the ICG dynamics proved difficult. We used principal component analysis (PCA) in this study to extract important elements without significant loss of information. We examined ICG spatiotemporal profiles and identified critical features related to vascular disorders. PCA time courses of the first three components showed a distinct pattern in diabetic patients. Among the major components, the second principal component (PC2) represented arterial-like features. The explained variance of PC2 in diabetic patients was significantly lower than in normal controls. To visualize the spatial pattern of PCs, pixels were mapped with red, green, and blue channels. The PC2 score showed an inverse pattern between normal controls and diabetic patients. We propose that PC2 can be used as a representative bioimaging marker for the screening of vascular diseases. It may also be useful in simple extractions of arterial-like features.

Keywords: biophotonics; fluorescence imaging; indocyanine green; principal component analysis; vasculopathy; pharmacokinetics.

1 Introduction

A number of diagnostic tools have been developed for structural and functional assessment of the vascular system. Our research focuses on identifying functional vascular parameters, because vascular disease prognosis correlates better with functional perfusion level than with vascular structure. However, traditional diagnostic tools, such as the ankle-brachial index test, laser Doppler imaging, and computed tomography angiography, are limited by their low sensitivity, high cost, and poor accessibility, and newer technologies are required to provide more accurate functional information.

For obtaining functional parameters, we previously proposed an optical imaging method, dynamic fluorescence imaging (DyFI), based on the time-series analysis of indocyanine green (ICG) pharmacodynamics. ICG, an FDA-approved, non-specific near-infrared fluorophore, has been widely used for detection of synovitis, sentinel lymph node, rheumatoid arthritis, breast cancer imaging, and studies for vascular events. With DyFI, we can measure perfusion rate with higher accuracy and sensitivity compared to other conventional methods. Previous studies have shown that analysis of time-series ICG images can predict the prognosis of murine hindlimb ischemia. Additionally, by analyzing ICG serial images of dorsal feet, we were able to identify clinical features associated with peripheral vascular insufficiency; for example, a reliable characteristic feature of Raynaud phenomenon (RP) was identified as modified $T_{\text{max}}$, calculated as the length of time between peak onset and maximum peak fluorescence. Perfusion rate has also been used as a quantitative measure of tissue perfusion, with sufficient sensitivity to diagnose mild peripheral arterial occlusive disease. Recently, symmetry analysis of the left and right extremities has been used to diagnose microvascular abnormalities in feet. Our ICG fluorescence images have 768×512 pixels and 120 time frames. The large amount of data contained in these images may cause dimensionality issues that can severely restrict its practical application. The use of raw high-dimensional data makes it difficult to extract the important elements that form the representative pattern of ICG fluorescence dynamics. Previously suggested features, such as the modified $T_{\text{max}}$ and the perfusion rate, are susceptible to noise signals and movement artifacts. Furthermore, in cases of increased vascular permeability, functional parameters may be underestimated. For these reasons, the present study was conducted to assess the entire set of spatiotemporal data using only a few components, rather than thousands of variables.

To project high-dimensional data into a low-dimensional space, a dimensional reduction algorithm, such as principal component analysis (PCA), can be applied to extract desirable features. The new variables detected by PCA, called principal components (PCs), are linear combinations of the original ICG dynamics. PCA with ICG fluorescence dynamics has been used to generate anatomical maps of the internal organs of a mouse. Additionally, spatiotemporal analyses of rheumatoid arthritis have been conducted based on PCA, with good...
The congruence observed between PCA features and synovitis in hand joints. The application of PCA with ICG fluorescence dynamics to the assessment of vasculopathy in diabetic patients has not yet been conducted. In this study, we apply the PCA algorithm to ICG images of diabetic patients and normal controls, mathematically extract the PCs that differentiate the two groups, and attempt to interpret the biological meanings of the PCs. The overall aim is to identify new imaging-based biomarkers that indicate the presence of underlying vasculopathy.

2 Subjects and Methods

2.1 Subject Profiles

The study protocol was approved by the institutional review board of Gangnam Severance Hospital, Seoul, Korea. Written informed consent was obtained from each subject. A total of 19 subjects (9 diabetic patients, 10 normal controls) underwent near-infrared ICG fluorescence imaging. The age of the subjects ranged from 50 to 65 years (mean: 58.78 ± 5.52 years) for diabetic patients and 51 to 60 years (mean: 55.90 ± 3.14 years) for controls. For diabetic patients, exclusion criteria included the presence of acute coronary syndrome (unstable angina, acute myocardial infarction), history of heart failure, and pregnancy/lactation. Exclusion criteria for controls included Raynaud’s syndrome and pregnancy/lactation. Demographic data of the subjects are summarized in Table 1.

Table 1 Demographics of subjects enrolled in the study.

<table>
<thead>
<tr>
<th>Group</th>
<th>Control</th>
<th>Diabetes</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>10 (20 ft)</td>
<td>9 (18 ft)</td>
</tr>
<tr>
<td>Age (years, mean ± SD)</td>
<td>55.90 ± 3.14</td>
<td>58.78 ± 5.52</td>
</tr>
<tr>
<td>Median age (years)</td>
<td>56</td>
<td>57</td>
</tr>
<tr>
<td>Gender</td>
<td>Female</td>
<td>Female</td>
</tr>
<tr>
<td>HTN (n)</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>DM duration (months, mean ± SD)</td>
<td>—</td>
<td>131.89 ± 99.48</td>
</tr>
<tr>
<td>Smoking (n)</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Dyslipidemia (n)</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Note: Data represent the mean ± SD; HTN, hypertension; diagnosed by physicians.

2.2 Dynamic Fluorescence Imaging

DyFl with ICG was used to measure various vascular parameters, as previously reported. The ICG fluorescence imaging system for clinical applications was manufactured by Vieworks Corporation (Anyang, Gyeonggi-do, Korea) and consisted of a charge-coupled device digital camera (RXD-500, Vieworks Co.) with an 830-nm bandpass filter (RG830; Edmund Optics, Barrington, New Jersey) and 760-nm light-emitting diode (LED) arrays (SMC760; Marubeni America Co., Sunnyvale, California). The sum of the current of the three LED arrays was kept under 1.3 A. Time-series images (768 x 512 pixels) of both dorsal feet were obtained while subjects were in a supine position. Each image was taken at 5-s intervals for 600 s, immediately following an intravenous bolus injection of ICG (0.16 mg/kg; Dongindang Pharm. Co., Gyeonggi-do, Korea). Characteristic features of the ICG fluorescence images are presented in Fig. 1 such as the temporal sequence [Fig. 1(a)]. Figure 1(b) shows that each pixel has different ICG pharmacokinetic dynamics. A three-dimensional (3-D) plot of ICG fluorescence dynamics is presented in Fig. 1(c). Data with movement artifacts were excluded from the analyses.
2.3 Principal Component Analysis

A schematic diagram of the ICG dynamics analysis using PCA is shown in Fig. 2. The first step involves loading ICG sequential images, which consist of ICG fluorescence intensities. The region of interest was both feet except the background. The fluorescence intensity was converted to a matrix. The rows of the input matrix correspond to spatially successive pixels, while columns correspond to temporally sequential frames. The covariance matrix was calculated from the input matrix. The components were then extracted as its eigenvector. The eigenvector with the highest eigenvalue was taken as the first PC. The feature vector consisted of the first three PCs. By calculating the inner product of the input matrix with the PCs, high-dimensional raw data could be converted into low-dimensional data without significant loss of information.

The region of interest was both feet except the background and more dispersed than in normal controls [Fig. 3(d)]. In diabetic patients, PC2 curves were smoother with arterial input function (AIF), as an early bolus arrival, a coefficient value was faster in normal controls than in diabetic patients, corresponding to general ICG pharmacokinetic pattern [Fig. 3(c)].

PC2 curves in normal controls is similar to that observed with arterial input function (AIF), as an early bolus arrival, a coefficient value was faster in normal controls than in diabetic patients, corresponding to general ICG pharmacokinetic pattern [Fig. 3(c)].

Statistical differences were analyzed by two-tailed Student’s t test, or Spearman’s correlation where indicated. Data are expressed as the mean ± standard deviation, and a p value of <0.05 is considered statistically significant.

3 Results

3.1 Distinctive Patterns of Principal Components in Diabetic Patients and Normal Controls

PCA was applied to extract the important elements from raw ICG fluorescence dynamics. From the input matrix, which contains the sequential ICG fluorescence intensities of all pixels on the foot, PCs were selected based on their eigenvalue. The eigenvector with the highest eigenvalue was defined as the first PC. The second PC was defined as the second-highest eigenvalue with a direction uncorrelated to the first PC. Each eigenvector has 120 dimensions because the raw data consisted of 120-frame sequential images. Time to the PC1 maximum exponent is the normalized fluorescence intensity at the "th pixel.

\[ I_p = [i_{p1}, i_{p2}, \ldots, i_{p120}] \]

\[ S = \begin{bmatrix} I_1 \\ \vdots \\ I_N \end{bmatrix} \]

The covariance matrix \( C \) is calculated using Eq. (1), while the eigenvector \( U \) and eigenvalue \( \lambda \) are calculated from \( C \) [Eq. (3)].

\[ C = S^T \cdot S, \]

\[ C = U^T \cdot \lambda \cdot U. \]

PCs were extracted according to their eigenvalue. The eigenvector with the highest eigenvalue was considered the first PC (PC1). The PC2 vector was the second highest eigenvalue and orthogonal to PC1. By calculating the inner product between the input matrix \( S \) and eigenvector \( U \) sorted by decreasing order, we could project the entire fluorescence dynamics onto PC space \( (x) \) [Eq. (4)].

\[ e(x; U) = S \cdot U = X \]

We wrote a C++ (Visual Studio 2010, Microsoft) program that can extract ICG fluorescence dynamics from each pixel. PCA was performed using the princomp function included in MATLAB® software (MATLAB® 2014b, Mathworks).

3.2 Statistical Analysis

Statistical differences were analyzed by two-tailed Student’s t test, or Spearman’s correlation where indicated. Data are expressed as the mean ± standard deviation, and a p value of <0.05 is considered statistically significant.
The meaning of PCs can be revealed by investigating the relationship between pixel values projected onto PCs and other known dynamics features. The dynamics features compared with PCs included blood flow index (BFI), mean transit time (MTT), and time to max \( (T_{\text{max}}) \). It has been reported that these parameters are related to vascular conditions. BFI and MTT showed significant differences between normal and middle cerebral artery occlusion in mice. Modified \( T_{\text{max}} \), which means the time from onset of ICG fluorescence to time for maximum intensity (MI), was reported as a diagnosis parameter of RP. We calculated Spearman’s correlation coefficients between pairs of features. The maximum fluorescence intensity and area under curve (AUC) showed positive correlation with the PC1 coefficient \( (r = 0.74 \text{ and } r = 0.73, \text{ respectively, } p < 0.0001, \text{ Spearman’s correlation}) \) [Fig. 4(a)]. MTT, calculated as the center of gravity of the dynamics, had a strong positive correlation with PC2 \( (r = 0.90, p < 0.0001, \text{ Spearman’s correlation}) \). \( T_{\text{max}} \) and percentile \( T_{\text{max}} \) also have strong correlation with PC2 \( (r = 0.87 \text{ and } r = 0.86, \text{ respectively, } p < 0.0001, \text{ Spearman’s correlation}) \). BFI, based on the slope of the rising peak, had a negative correlation with PC2 \( (r = -0.80, p < 0.0001, \text{ Spearman’s correlation}) \). We plotted scatter graphs to visualize the correlation between dynamics features and PCs [Figs. 4(c) and 4(d)]. Our results suggest that PC1 is closely related to the original pharmacokinetic curve (MI and AUC parameters), while PC2 seems to contain functional information related to the BFI and MTT parameters.

### 3.2 Explained Variance of Principal Components

Proportion of explained variance for PCs was identified as a quantitative value that could differentiate between normal controls and diabetic patients. Each PC contributes to a proportion of the total ICG fluorescence dynamic, based on its eigenvalue. The percentage of explained variance for each PC can be calculated using Eq. (6), which is the corresponding eigenvalue divided by the total variance \( (k \text{ is the component number; } \lambda \text{ is the eigenvalue}) \).

\[
\text{PC } k(\%) = \frac{\lambda_k}{\sum_{i=1}^{N} \lambda_i} \times 100.
\]

Table shows the average variance explained by corresponding PCs. For normal controls, the first, second, and third PCs...
account for 77.71, 17.82, and 2.65% of the average variance, respectively. For diabetic patients, these same PCs account for 92.36, 5.19, and 1.32% of the average variance, respectively. The total variance explained by the first three PCs in controls (98.18%) and diabetic patients (98.87%) was sufficient to represent the high-dimensional ICG fluorescence data. The noticeable difference between diabetic patients and controls was observed with the PC1 and PC2 pattern (Fig. 5), where the proportion of explained variance was significantly lower in diabetic patients than in controls ($p < 0.0001$, two-tailed Student’s t test). Figure 5(c) shows the gap between control and diabetes was widened by calculating the ratio of major two variances, PC1 (%) divided by PC2 (%). This finding suggests that explained variance of the first two PCs may be considered a potential imaging-based biomarker that can differentiate between a normal foot and a diabetic foot.

3.3 Multichromatically Visualized Principal Components Map

To represent the spatial distribution of PCs, PCA results were mapped against projection values of the dynamics of each pixel. A schematic representation of the mapping method is provided in Fig. 6(a). By calculating the inner product of normalized ICG fluorescence dynamics and PCs, the high-dimensional raw data were converted to low-dimensional data that conserve the important elements. The first three PCs became the new basis after the projection. Each pixel has three projection values on PC1, PC2, and PC3 space. Pseudocolored images were created

<table>
<thead>
<tr>
<th>Table 2 Explained variance of PCs (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Control</td>
</tr>
<tr>
<td>PC1</td>
</tr>
<tr>
<td>PC2</td>
</tr>
<tr>
<td>PC3</td>
</tr>
<tr>
<td>SUM</td>
</tr>
</tbody>
</table>
by merging the data with RGB channels (red, green, and blue channels represented PC1, PC2, and PC3, respectively). In PC1–PC2 spaces visualizing all the dynamics on the hallux and dorsal feet, only PC2 was capable of differentiating the hallux and dorsal feet regions [Fig. 6(b)]. The PC2 score on the dorsal feet was higher than the hallux in normal controls, while the reverse pattern was observed in diabetic patients. The distribution of PC2 scores between normal controls and diabetic patients from Fig. 6(b) is shown in histogram form in Fig. 6(c). Figure 6(d) compares the RGB maps of a normal control and a diabetic patient. The projection values of PC2 were mapped to veins on the dorsal feet in normal controls and the hallux in diabetic patients.

![Fig. 5](https://www.spiedigitallibrary.org/journals/Journal-of-Biomedical-Optics)

**Fig. 5** A dot plot of variance (%). Data are shown as means (red line) with standard error of the mean (light red area) and standard deviation (light blue area). (a) Variance explained of PC1 (%), (b) variance explained of PC2 (%), and (c) the ratio of two major variances, PC1 (%) divided by PC2 (%).

![Fig. 6](https://www.spiedigitallibrary.org/journals/Journal-of-Biomedical-Optics)

**Fig. 6** Projection of data onto PCs. (a) A map with PC-matching RGB channels. RGB mapping creates a visual representation of the spatial distribution of the PCs. Fluorescence intensity was projected onto the first three PCs. Each projected value has PC1, PC2, and PC3 scores. The spatial pixels corresponding to the first, second, and third orthogonal temporal components of the image were visualized as a red-green-blue map. (b) Comparison of scatter plots. By projecting raw ICG fluorescence dynamics onto PC1 and PC2, the data could be visualized as a scatter plot. (c) Histograms of PC2 and (d) a case study of PC distribution. Red, green, and blue channels represent PC1, PC2, and PC3, respectively.
patients (Fig. 7). These representative PC distribution color-maps show characteristic patterns for normal controls and diabetic patients, and could potentially be used to detect and evaluate vasculopathy in diabetic feet.

4 Discussion

The aim of the present study was to investigate the ICG spatio-temporal profile of vasculopathy in diabetic patients and identify distinct features. Previously, we have defined vascular parameters by selecting features from original raw dynamics. However, these methods are limited in their ability to identify underlying properties of ICG fluorescence dynamics. The methods are not robust enough to signal noise and motion artifacts, as only a single parameter from the original dynamics is selected, and actual vascular disorders are often underestimated. In the current study, we have shown that clinically relevant features of ICG fluorescence dynamics can be successfully identified by extracting PCs. PCA can extract robust-to-noise and scale-invariant features by excluding redundancies in raw data while also retaining hidden information, such as the AIF. Our study demonstrated that the PCs of ICG dynamics differ significantly between normal controls and diabetic patients. Thus, this method has a potential application as a diagnostic system for early detection of vascular complications in diabetic patients.

The eigenvectors extracted by PCA produced different characteristic curves. PC1 represents the original pharmacokinetic curve, while PC2 seems to contain functional information related to vasculopathy. PC2 was considered the most remarkable, as the PC2 curve in normal controls was found to be analogous to the known AIF pattern. AIF is typically estimated by measuring signal changes in major arteries. The AIF and PC2 vectors had a similar shape: a sharp rise to maximum intensity in the early phase, followed by a slower decrease after the peak. AIF has a fundamental role in the quantification of blood flow, as AIF describes the tracer concentration in vessels entering a tissue as a function of time. Despite its importance, there is no standard model for measuring AIF, and various approaches have been reported. Most studies utilize a dynamic susceptibility contrast-magnetic resonance imaging (MRI) technique for quantifying blood flow. AIF is typically modeled as a gamma-variate function, which is calculated using a scaling factor and parameters that determine the shape of the peak. For ICG imaging, the lagged normal density distribution function and recursive convolution algorithm have been demonstrated as modeling methods. Deconvolution analysis is normally required for the quantification process, as true ICG concentration is the convolution of AIF and the tissue residue function. However, in the present study, the characteristic behavior of AIF was detected as the PC2 basis vector time-course, without the use of complex deconvolution methods. AIF is dispersed during its passage from the point of measurement to the tissue. With abnormal vasculature, AIF may lose its distinct curve with mixed dynamics. For example, using perfusion MRI, Calamante et al. observed bolus dispersion in the arteries of subjects with steno-occlusive disease.

In accordance with AIF dispersion, PC2 appears to be dispersed in diabetic patients with vascular complications. While PC2 had a sharp peak in normal controls, a wider peak was observed in diabetic patients. Furthermore, the projected values on the PC2 basis showed an inverse pattern between normal controls and diabetic patients. PC2 scores were mapped to the hallux in diabetic patients and to the vein on the top of the foot in normal controls. PC2 scores were high on the hallux, even in the diabetic patients, because AIF should be observed in every artery and first capillary. However, PC2 scores for the vein and for the ends of capillaries were lower than scores for arteries in diabetic patients. We suggest that this hidden variable can be used to indicate how much of the AIF is conserved.

Fig. 7 The pseudo color-map of PC2 score. (a) and (b) A comparison of PC2 distribution with green color-map. The projection values of PC2 were mapped to veins on the dorsal foot in a normal subject [green dashed box in (a)] and to the hallux in a diabetic patient [green dashed box in (b)]. (c) and (d) A comparison of PC2 distribution with jet color-map. The projection values of PC2 are higher on veins than hallux in a normal subject. Inversely, the score is higher on hallux than veins in diabetic patients.
In conclusion, we have identified a hidden feature extracted from the PCA of ICG dynamics and investigated its biological meaning by known characteristics of AIF. We propose that PC2 could be useful in the diagnosis of vascular disorders. PC2 also represents a new index of AIF conservation, which is fundamental for quantifying pharmacokinetic parameters.

Acknowledgments

This research was supported by the Bio & Medical Technology Development Program of the NRF funded by the Korean government, MSIP (No. 2011-0019697).

References


Jihye Seo received her BS degree in biomedical engineering from Kyunghee University in 2014. After graduation, she joined the Cell Signaling and Bio-Imaging laboratory at Korea Advanced Institute of Science and Technology (KAIST), where she focused mainly on computational analysis of ICG fluorescence dynamics for extracting hidden features related to vascular disorders. Her current research interests include biomedical optics, signal processing, and machine learning. She is a student member of SPIE and IEEE.

Yuri An received her master’s degree in bio and brain engineering from KAIST in 2012. She is now taking the PhD course at the Cell Signaling and Bio-Imaging laboratory at KAIST. Her work is focused on ICG imaging analysis, especially clinical data. She is the coauthor of several articles published in scientific journals indexed in the ISI web of science and has presented at international scientific conferences.

Jungsul Lee is a postdoctoral fellow at the Cell Signaling and Bio-Imaging laboratory in KAIST. He received his BS degree in biology and minor in mathematics from Korea University in 2005. He received his MS and PhD in computational biology from KAIST in 2009 and 2013, respectively. His primary research interests are analysis of any type of large-scale biological data, mathematical modeling, and algorithm development for biological problems.

Taeyun Ku is a postdoctoral associate at the Institute for Medical Engineering and Science at the Massachusetts Institute of Technology (MIT). He received his MD degree from Yonsei University College of Medicine in 2006 and his PhD from the Graduate School of Medical Science and Engineering of KAIST in 2012. His previous and current research fields include biomedical imaging, theoretical and computational biology, neurophysiology, and tissue engineering.

Yujung Kang is currently a chief researcher at the Bio team, R&D center, Vieworks Co., Ltd. Her previous appointments include postdoctoral fellow at Yale Cardiovascular Research Center and KAIST College of Engineering. She graduated with a PhD from KAIST in 2008. One part of his job involves development of new molecular imaging systems and imaging analysis algorithms.

Chulwoo Ahn is a professor, medical doctor and chief of endocrinology and diabetes center in Gangnam severance hospital. He received
his MD degree, Masters degree and PhD degree from Yonsei University College of Medicine in 1991, 1998 and 2002, respectively. His primary research interests are diabetes mellitus and obesity.

Chulhee Choi is a professor at Cell Signaling and Bioimaging Laboratory, department of Bio and Brain Engineering in KAIST. He is also a chair of Optical Bioimaging Center in KAIST. He received his MD Degree, Masters Degree and PhD degree in 1991, 1997, and 1999, respectively. He is a member of several international journals editorial board. His major research interests are cell signaling, biomedical imaging, and biophotonics.