

Tissue Characterization Using Dimensionality Reduction and Fluorescence Imaging

Karim Lekadir¹, Daniel S. Elson², Jose Requejo-Isidro², Christopher Dunsby², James McGinty², Neil Galletly³, Gordon Stamp³, Paul M. W. French², and Guang-Zhong Yang¹

¹ Visual Information Processing Group, Department of Computing

² Department of Physics

³ Division of Investigative Sciences
Imperial College London, United Kingdom

Abstract. Multidimensional fluorescence imaging is a powerful molecular imaging modality that is emerging as an important tool in the study of biological tissues. Due to the large volume of multi-spectral data associated with the technique, it is often difficult to find the best combination of parameters to maximize the contrast between different tissue types. This paper presents a novel framework for the characterization of tissue compositions based on the use of time resolved fluorescence imaging without the explicit modeling of the decays. The composition is characterized through soft clustering based on manifold embedding for reducing the dimensionality of the datasets and obtaining a consistent differentiation scheme for determining intrinsic constituents of the tissue. The proposed technique has the benefit of being fully automatic, which could have significant advantages for automated histopathology and increasing the speed of intraoperative decisions. Validation of the technique is carried out with both phantom data and tissue samples of the human pancreas.

1 Introduction

Fluorescence is an effective means of achieving optical molecular contrast in a wide range of instruments including cuvette-based systems, microscopes, endoscopes and multi-well plate readers. Fluorescent molecules (fluorophores) can be used as “labels” to tag specific molecules of interest. Alternatively, the fluorescence properties of the target molecules themselves may be exploited to provide *label-free contrast*. In addition to providing information about the properties of the fluorophores, the fluorescence process can be sensitive to the local environment surrounding the fluorophore, thus providing a *sensing* function. In principle, different species of fluorophores may be characterized by their excitation and emission spectra, quantum efficiency, polarization response, and fluorescence lifetime. These parameters can change as a function of the local viscosity, temperature, refractive index, pH, calcium, oxygen concentration, and electric field. The practical use of these techniques for tissue characterization is therefore challenging due to the need for an automatic method for extracting

the intrinsic fluorescence properties of the tissue from what is often high dimensional data (combinations of x , y , t , λ_{ex} , λ_{em}).

Fluorescence lifetime imaging (FLIM) is a technique that allows the fluorescence decay profile to be determined for each pixel of an image. After optical excitation, the fluorescence emitted by a fluorophore typically decays with a molecular- and/or environment-dependent average time, called the fluorescence lifetime, (typically ~ 100 ps – few ns). FLIM is becoming increasingly popular as a method of measuring FRET [1] or autofluorescence contrast in tissues [2,3]. Typically, the computation of fluorescent lifetimes or the analysis of raw FLIM data requires a model of the expected decay function to be chosen based on a priori assumptions. For instance, a double exponential decay model may be chosen in the case of a molecule with two distinct decay pathways, or a stretched exponential model may be chosen where a distribution of lifetimes is expected [4]. The results obtained depend on the choice of the model, and there are many situations where either the correct model is unknown or involves more fitting parameters than are meaningful for the limited signal-to-noise ratio. In practice, even if the correct model is used to obtain the fluorescence lifetime at each pixel, the analysis and presentation of the data are complicated for systems designed to resolve additional fluorescence properties such as the excitation or emission profiles.

Thus far, some alternative model-free methods have been proposed for time resolved data only, such as the expansion of fluorescence decays in a discrete Laguerre basis [5]. The challenge, however, is to take advantage of the model-free nature of the algorithms and treat the multi-spectral raw time-gated images as the input dimensions whilst presenting the data in an intuitive image format. The purpose of this paper is to introduce a new framework that is model free and makes no assumptions about the distribution of the data for automated tissue characterization. By the use of manifold embedding, the method reduces the entire time-resolved image stack to a consistent representation using a color map that reflects the intrinsic fluorescence properties of the tissue sample. This technique has the additional benefit of being entirely automated, which could have significant advantages for automated histopathology and increasing the speed of intraoperative decisions. Validation of the technique is carried out with both phantom data and tissue samples of the human pancreas.

2 Methods

2.1 Fixed Reference IsoMap (FR-IsoMap)

Theoretically, the main contribution of the work is the introduction of FR-IsoMap which allows consistent dimensionality reduction across samples. IsoMap [6] is a nonlinear dimensionality reduction technique that is able to describe the global structure of non-linear manifolds by detecting the main meaningful underlying dimensions. The method preserves the interpoint distances in a way that is similar to classical multidimensional scaling (MDS) [7] but caters for complex nonlinear manifolds by the use of geodesic distances for representing the dissimilarities.

Given a set of N pixels $\{P_i\}$ in the original space, a neighborhood graph G is first constructed and used to initialize the geodesic distance matrix D_G as follows:

$$d_G(i, j) = \begin{cases} d(P_i, P_j) & \text{if } P_i \text{ and } P_j \text{ are neighbors} \\ +\infty & \text{otherwise} \end{cases} \tag{1}$$

where d is a dissimilarity distance between two pixels. The final geodesic distances are determined by calculating the shortest paths between all pairs of points in the graph G , by replacing all entries d_G for each value of k from 1 to N by:

$$d_G(i, j) = \min(d_G(i, j), d_G(i, k) + d_G(k, j)) \tag{2}$$

The matrix D_G of geodesic distances is then fed to the classical MDS procedure [7]. The derived N eigenvectors v_p are sorted in the decreasing order of the corresponding eigenvalues λ_p . The new coordinates $\{y_{ip}\}$ of a pixel P_i in the embedded space are then calculated as follows:

$$y_{ip} = \sqrt{\lambda_p} v_{pi} \tag{3}$$

The three main coordinates are used as color channels to construct a color map representation for tissue characterization. For every new image, however, reapplying IsoMap can be time consuming and may not guarantee a consistent embedding as it can change dramatically depending on the data distribution of the manifold. This is due to the lack of a fixed coordinate system, which prohibits the comparison of the embedded results across different tissue samples. To circumvent this problem, FR-IsoMap is developed, which involves applying IsoMap only on a training data set that represents well the variability within the tissue. In order to obtain an optimal reference coordinate system whilst maintaining the topology of the manifold, input vectors that are evenly distributed on the manifold are selected. This is achieved by first selecting one or some seed pixels that correspond to the main constituents in the image and then eliminating points on the manifold that are within a predefined distance of the selected pixel. This is then continued until all pixels are either selected or rejected.

To ensure consistent embedding, the training sample generated at the first stage and the corresponding original and embedding coordinates are used as a model to predict the embedded coordinates of a pixel in a new image. Given the fact that the training sample represents the variability within the data, the position of a new pixel in the embedded space can be predicted using the coordinates of the most similar pixels in the training set. The k nearest neighbors of the training samples are located and the corresponding embedded coordinates are then calculated by minimizing the Sammon’s nonlinear mapping criteria shown in Eq. (4) such that the remapped distances approximates well the original distances [8] *i.e.*,

$$E_s = \frac{1}{\sum_{i < j} [d_{ij}^*]} \sum_{i < j} \frac{(d_{ij}^* - d_{ij})^2}{d_{ij}^*} \tag{4}$$

where d_{ij} and d_{ij}^* are pairwise distances in the embedded and original spaces, respectively. By defining $E_s(m)$ and $d_{ij}(m)$ as the mapping error and the embedded

distance after the m^{th} -iteration, respectively, the above problem can be solved iteratively and the newly estimated coordinates of the sample at iteration $m + 1$ is given by:

$$x_p(m+1) = x_p(m) - \alpha \left[\frac{\partial E(m)}{\partial x_p(m)} \right] \left[\frac{\partial^2 E(m)}{\partial x_p(m)^2} \right]^{-1} \quad (5)$$

where x_p is the p^{th} coordinate component of the new pixel in the fixed coordinate system, and α is the step size. The initial values of the coordinates are given by the nearest neighbor, which in practice gives a good indication of the final result.

2.2 Phantom and Tissue Sample Validation

In order to validate the proposed FR-IsoMap scheme for tissue characterization, both phantom data and tissue samples of the unstained human pancreas were used. Time-resolved fluorescence images were recorded by exciting the sample using a pulsed excitation source (a broadband supercontinuum laser source for the phantom experiment [9] and an amplified and frequency-doubled Ti:Sapphire laser for the pancreas tissue data). The resulting fluorescence was then imaged onto a microchannel plate gated optical intensifier (GOI, gate width ~ 400 ps) and temporal profiles of the fluorescence decay were recorded by acquiring gated images of the fluorescence at different delays after excitation (for details of a similar system see [10]).

For the phantom data set, a total of 17 gated images were recorded over a total delay range of 11 ns of two identical rows in a 384 multi-well plate containing five different dye solutions (DASPI in ethanol (DASPI-I), Coumarin, Rhodamine 700, Eosin, and DASPI in 40% ethanol, 60% glycerol (DASPI-II)). The large area imaged (1.5 x 2 cm) allowed a standard camera objective to image the fluorescence onto the MCP.

For pancreas tissue imaging, the MCP was coupled to an Olympus IX71 inverted fluorescence microscope, and 14 gated images were recorded over a total delay time of 15 ns. All human tissues used in these experiments, including fresh resection specimens and biopsies, were obtained in accordance with local ethics committee approval (2004/6742). All tissue had been ethically consented for use in medical research. Blocks of formalin-fixed paraffin-embedded tissue were cut to 10 μm sections and mounted unstained on uncoated glass slides. Glass coverslips with 1.5 thickness were mounted on the slides using either an aqueous-based coverslip mountant (50% PBS, 50% glycerol) or a commercially available mountant (DPX).

To assess the quality of the mapping, the distances between every new pixel in an image and the entire training sample are calculated in both the original space and the common coordinate system. The preservation of the distances is then measured by calculating the corresponding correlation coefficient:

$$\rho_{D^*D} = \frac{\sum (d_{ij}^* - \bar{d}^*) (d_{ij} - \bar{d})}{\sqrt{\sum (d_{ij}^* - \bar{d}^*)^2 \sum (d_{ij} - \bar{d})^2}} \quad (6)$$

3 Results

Fig. 1 shows the conventional lifetime map (FLIM map) and the proposed FR-IsoMap for the multi-well phantom data with five different dye solutions. It can be seen that the FLIM map shows strong lifetime contrast for the DASPI-I and the Coumarin but is not able to distinguish the remaining dyes well. As highlighted in the histogram in Fig. 2(a), five peaks representing the different dyes are generated but with narrow separation and overlapped with the background, suggesting that ambiguous segmentation may be generated in the FLIM map. Also, the areas immediately surrounding the wells are low intensity background, but lifetimes for these regions are calculated that are similar to the well in the immediate vicinity due to the influence of scattered light or image blurring. Although this background could be thresholded out, this requires user input and is an interruption of the automated process. In addition, signals of low intensity may be lost along with the background. In contrast, the FR-IsoMap shows distinctive sample separation with clearly delineated background. The mapping accuracy calculated using Eq. (6) was equal to 0.989 and 0.993 for the two samples shown in Fig. 1, respectively.

In Table 1, the average interpixel distances in the embedded space were calculated between all combination of dyes and within the five dyes. The results show a low average interpixel distance within a dye while this distance increases significantly across the different dyes, suggesting that the constituents are well separated. For a

Table 1. Average interpixel distances within and between the different dyes calculated in the embedded space

| | DASPI-I | Coumarin | Rhodamine | Eosin | DASPI-II | Background |
|------------|---------|----------|-----------|-------|----------|------------|
| DASPI-I | 11.6 | 293.2 | 146.0 | 293.3 | 174.3 | 191.2 |
| Coumarin | 293.2 | 57.7 | 193.1 | 260.1 | 186.8 | 233.2 |
| Rhodamine | 146.0 | 193.1 | 13.6 | 212.5 | 72.0 | 70.4 |
| Eosin | 292.3 | 260.1 | 212.5 | 41.3 | 145.7 | 229.2 |
| DASPI-II | 174.3 | 186.8 | 72.0 | 145.7 | 17.5 | 113.6 |
| Background | 191.2 | 233.2 | 70.4 | 229.2 | 113.6 | 4.6 |

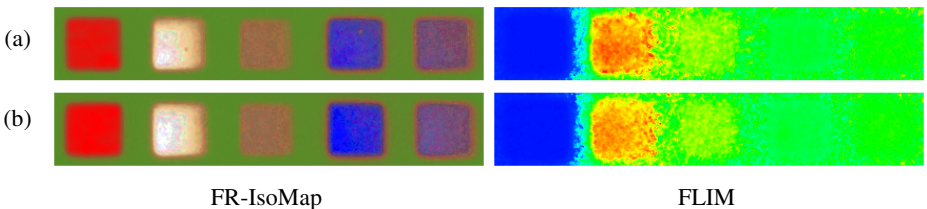


Fig. 1. Results of FR-IsoMap and FLIM as applied to the multi-well phantom data with five different dye solutions. FLIM with single exponential decay is not able to distinguish completely between the different dyes or the background while the FR-IsoMap clearly distinguishes all the dyes in the images.

visual illustration, Fig. 2(b) shows the histogram of the interpixel distances within the Rhodamine dye and between the Rhodamine and the DASPI-II dyes. Although the average interpixel distance between the two dyes correspond to the lowest amongst all dyes combinations in Table 1, the graph shows that the two constituents are clearly separated by using the proposed FR-IsoMap method.

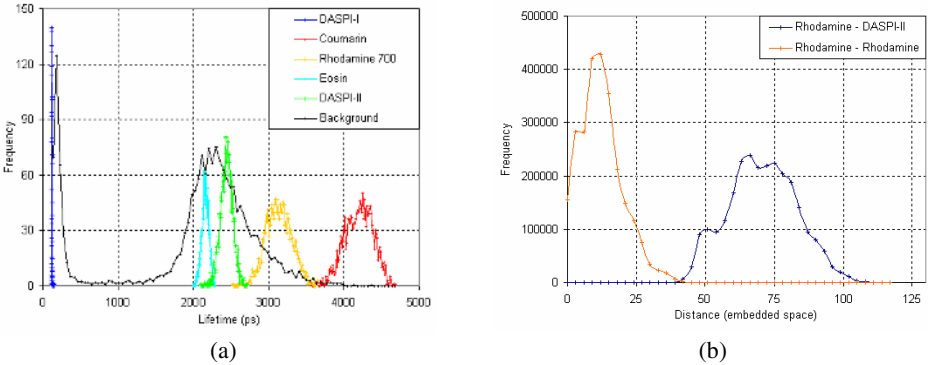


Fig. 2. (a) FLIM histogram showing the peaks corresponding to the different dyes overlapping especially with the background. In (b), the histogram describing the interpixel distances in the FR-IsoMap embedded space within the Rhodamine and between DASPI-II and the Rhodamine clearly separates the two dyes.

For the characterization of the pancreas tissue compositions, the proposed method was applied to the 30 datasets collected. The training sample was constructed by selecting 567 pixels from dataset 28. It is evident from the accuracy results plotted in Fig. 3 that high mapping accuracy is maintained throughout the entire dataset, with an average correlation coefficient equal to 0.993 ± 0.005 .

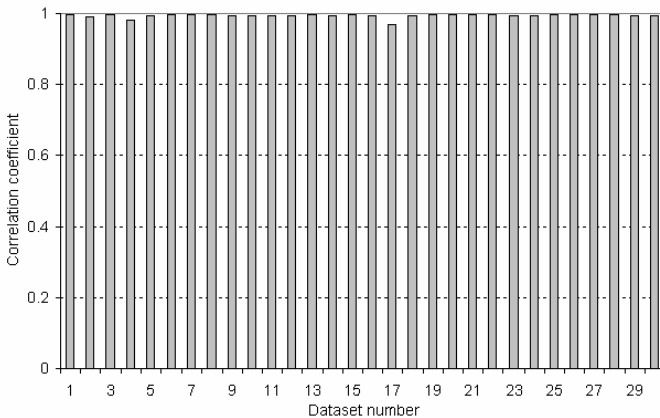


Fig. 3. Reconstruction accuracy plotted for the 30 human tissue pancreas samples studied, demonstrating a high accuracy mapping for the entire tissue datasets

For visual illustrations of the color maps generated for the pancreas images, Fig. 4 (a) and (b) illustrate two examples corresponding to the datasets 1 and 25 involved in this study. It can be seen that the different constituents are well distinguished in the color maps and mapped in a consistent manner across the samples when compared with the FLIM images. In particular, there appears to be more fine structure and improved spatio-contrast visible in FR-IsoMap. The oval structure in the centre of Fig. 4(a) is a duct, and the area just to the right of it (around the centre of the image) is collagen and elastin rich tissue. Fig. 4(b) has two oval islets of Langerhans in the centre, a duct on the right hand side, and the edge of the pancreas towards the left. It is worth noting that Fig. 4(c) corresponds to the same sample than in (b) but with a particularly high level of noise, in this case the obtained FR-IsoMap still displays a good contrast between the different constituents in the image while the FLIM map is affected by the noise, due to the least squares approach to fit the exponential decay.

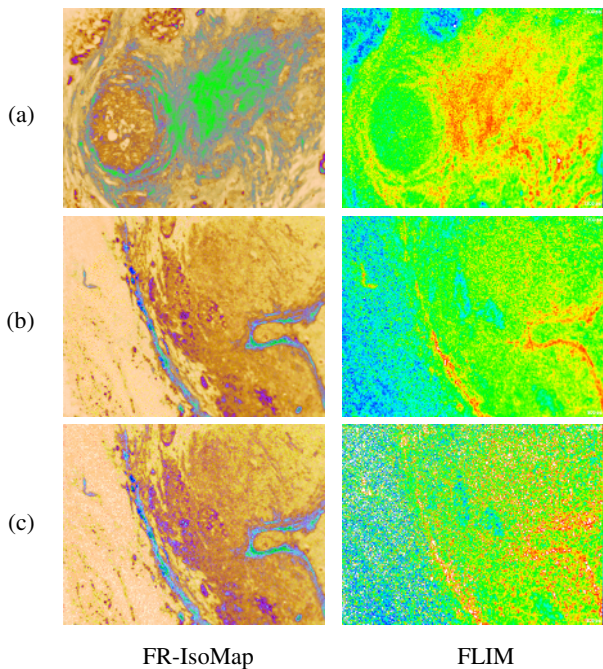


Fig. 4. Example human pancreas tissue sections mapped with FLIM and FR-IsoMap respectively, showing the improved tissue class separation and improved SNR and spatio-contrast introduced by the proposed tissue characterization framework

4 Conclusions

We have presented in this paper a framework for tissue characterization with time resolved fluorescence images. The method is based on soft clustering achieved by using dimensionality reduction. The proposed FR-IsoMap constructs a fixed

dimensional coordinate system by selecting a training set that best represents the variability within the tissue. New samples are then embedded into the reference coordinate system such that local similarities are preserved. Consistent color maps are generated from the embedded data which are capable of enhancing dissimilarities between tissue compositions. In comparison with the conventional FLIM maps, the proposed methods does not require *a priori* knowledge about the sample under investigation and all of the analysis may be presented on just one image in a fully automated manner. Furthermore, this approach may be extended to higher-dimensionality data sets, for instance where combinations of the lifetime, excitation and emission spectral profiles are recorded, and validation work is currently in progress.

References

1. Bastiaens PIH and Squire A, "Fluorescence lifetime imaging microscopy: spatial resolution of biochemical processes in the cell," *Trends in Cell Biology*, vol. 9, pp. 48-52, 1999.
2. Mizeret J, Wagnieres G, Stepinac T, and Bergh HVD, "Endoscopic tissue characterization by frequency-domain fluorescence lifetime imaging (FD-FLIM)," *Lasers in Medical Science*, vol. 12, pp. 209-217, 1997.
3. Cubeddu R, Comelli D, D'Andrea C, Taroni P, and Valentini G, "Time-resolved fluorescence imaging in biology and medicine," *Journal of Physics D: Applied Physics*, vol. 35, pp. R61-R76, 2002.
4. Lee KCB, Siegel J, Webb SED, Leveque-Fort S, Cole MJ, Jones R, Dowling K, Lever MJ, and French PMW, "Application of the stretched exponential function to fluorescence lifetime imaging," *Biophysical Journal*, vol. 81, pp. 1265-1274, 2001.
5. Jo JA, Fang QY, Papaioannou T, and Marcu L, "Fast model-free deconvolution of fluorescence decay for analysis of biological systems," *Journal of Biomedical Optics*, vol. 9, pp. 743-752, 2004.
6. Tenenbaum JB, Silva Vd, and Langford JC, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, pp. 2319-2323, 2000.
7. Cox TF and Cox MAA, *Multidimensional scaling*. London: Chapman & Hall, 1994.
8. Sammon JW, "A nonlinear mapping algorithm for data structure analysis," *IEEE Transactions Computers*, vol. 18, pp. 401-409, 1969.
9. McConnell G, "Confocal laser scanning fluorescence microscopy with a visible continuum source," *Optics Express*, vol. 12, pp. 2844-2850, 2004.
10. Siegel J, Elson DS, Webb SED, Lee KCB, Vlanclas A, Gambaruto GL, Leveque-Fort S, Lever MJ, Tadrous PJ, Stamp GWH, Wallace AL, Sandison A, Watson TF, Alvarez F, and French PMW, "Studying biological tissue with fluorescence lifetime imaging: microscopy, endoscopy, and complex decay profiles," *Journal/Applied Optics*, vol. 42, pp. 2995-3004, 2003.