

# Broadband hyperspectral coherent anti-Stokes Raman scattering microscopy for stain-free histological imaging with principal component analysis

Jingjiang Xu<sup>a</sup>, Baoshan Guo<sup>a</sup>, Kenneth K.Y. Wong<sup>a</sup>, Kevin K. Tsia\*<sup>a</sup>

<sup>a</sup>Applied Life Photonics Lab, Department of Electrical and Electronic Engineering  
The University of Hong Kong, Pokfulam Road, Hong Kong

\*Email: tsia@hku.hk

## ABSTRACT

Routine procedures in standard histopathology involve laborious steps of tissue processing and staining for final examination. New techniques which can bypass these procedures and thus minimize the tissue handling error would be of great clinical value. Coherent anti-Stokes Raman scattering (CARS) microscopy is an attractive tool for label-free biochemical-specific characterization of biological specimen. However, a vast majority of prior works on CARS (or stimulated Raman scattering (SRS)) bioimaging restricted analyses on a narrowband or well-distinctive Raman spectral signatures. Although hyperspectral SRS/CARS imaging has recently emerged as a better solution to access wider-band spectral information in the image, studies mostly focused on a limited spectral range, e.g. CH-stretching vibration of lipids, or non-biological samples. Hyperspectral image information in the congested fingerprint spectrum generally remains untapped for biological samples.

In this regard, we further explore ultrabroadband hyperspectral multiplex (HM-CARS) to perform chemoselective histological imaging with the goal of exploring its utility in stain-free clinical histopathology. Using the supercontinuum Stokes, our system can access the CARS spectral window as wide as  $>2000\text{cm}^{-1}$ . In order to unravel the congested CARS spectra particularly in the fingerprint region, we first employ a spectral phase-retrieval algorithm based on Kramers–Kronig (KK) transform to minimize the non-resonant background in the CARS spectrum. We then apply principal component analysis (PCA) to identify and map the spatial distribution of different biochemical components in the tissues. We demonstrate chemoselective HM-CARS imaging of a colon tissue section which displays the key cellular structures that correspond well with standard stained-tissue observation.

**Keywords:** Hyperspectral coherent anti-Stokes Raman scattering, stain-free, histological imaging, principle component analysis

## 1. INTRODUCTION

Tissue staining/labeling, e.g. using hematoxylin and eosin (H&E) stain, has long been a routine procedure involved in histological study. However, conventional staining method requires laborious work such as fixing, thin tissue sectioning, staining with dyes and packaging. Before it is examined under a microscope, the tissue is prone to physical distortion and chemical information loss. Thus it is necessary to develop alternative techniques for histological imaging with minimal sample preparation steps and the associated complications. Coherent anti-Stokes Raman scattering (CARS) is a powerful stain-free visualization method based on intrinsic chemical vibrational frequency [1, 2]. Compared to spontaneous Raman scattering, CARS offers orders-of-magnitude stronger signals which enable sensitive chemoselective imaging at the video rate [3]. Conventional instrumentation of CARS utilized two synchronized lasers or optical parametric oscillators with a frequency difference to perform single-color or narrow-band CARS image [4]. However, the richness of broadband spectral Raman information in the sample is not efficiently utilized in this narrowband CARS excitation method.

In this paper we present a hyperspectral multiplex CARS (HM-CARS) microscopy of biological tissue sections based on supercontinuum (SC) generation which enables us to obtain broadband CARS spectra in the individual image pixel - which indicate the intrinsic chemical information in the sample [5]. This ultrabroadband Raman spectral range includes not only the CH vibrational signal ( $2800\text{--}3000\text{cm}^{-1}$ ) which mostly generated by lipids [6] but also other Raman signatures in the fingerprint region which is not commonly revealed by CARS bioimaging. As CARS is a third-order optical nonlinear effect in which the anti-Stokes signal is proportional to the squared modulus of complex third-order susceptibility ( $|\chi^{(3)}|^2$ ) which includes both the resonant and non-resonant terms [2]. To minimize the non-resonant background, a phase retrieval

algorithm based on Kramers-Kronig transform is utilized [7]. The HM-CARS system generates a large dataset of rich Raman information which demands for multivariate statistical techniques to unravel the congested Raman spectral signatures. Principal component analysis (PCA) is a powerful algorithm to express all spectral data set in the terms of several significant principal components [8]. We further utilize PCA to create high contrast HM-CARS images showing the spatial distribution of different components in the tissue sections. It could potentially be utilized in and complementary to the existing histological applications, with minimal preparation procedures.

## 2. EXPERIMENTAL SETUP AND METHODS

Figure 1 shows the experimental setup of our HM-CARS system. A high power Ti:Sapphire mode-locked laser (Mai Tai BB, Newport) is utilized to provide the pump and Stokes for CARS. The laser output pulse has the center wavelength of 800 nm with 80 fs pulse width (80 MHz repetition rate). After passing through a Faraday isolator, the pulse width is largely broadened which needs to be compressed back to the transform limited pulse by a prism-pair compressor. By tuning the polarization of the laser with a half wave plate, we can optimize the power ratio after the polarization beam splitter. We split ~100 mW of the laser power for coupling into a photonic-crystal fiber (PCF) (femtoWHITE-800, NKT) as the pump for SC generation, pumped at the anomalous dispersion regime. The output power of PCF is ~22mW which generates broadband SC covering a Raman shift range larger than  $3000\text{ cm}^{-1}$ . Then we select the output signal in the longer wavelengths beyond 805 nm as a Stokes beam by a long pass filter. As the bandwidth of Ti:Sapphire laser is ~10nm, we use a 4-f grating pairs with a slit to narrow down the bandwidth of pump which can improve the spectral resolution for HM-CARS spectra. After the generation of Stokes and spectral narrowing of the pump, we combine the two beams by a dichroic filter. By tuning the delay line in the Stokes path, the temporal overlap between the two beams can be adjusted. An objective lens with high numerical aperture ( $\text{NA} = 0.85$ ) is used to focus the two beams on the sample. This tight focusing condition relaxes the phase matching condition for the generation of the anti-Stokes signal. It can also reduce the background signal from the substrate and provides high diffraction-limited resolution. After the collimating objective, the CARS signal is separated by a short pass filter which blocks the pump and Stokes beam. Consequently, the filtered anti-Stokes has a spectral range as wide as  $\sim 2000\text{ cm}^{-1}$  (ranging from  $\sim 600\text{ nm}$  to  $780\text{ nm}$ ). The HM-CARS signal thus covers most of characteristic vibrational frequencies for biological samples. We use a spectrometer (Acton SP2300, Princeton Instruments) with a sensitive cooled charge-coupled device (CCD) (PIXIS 100, Princeton Instruments) to detect the CARS signal. We construct the HM-CARS image by a motorized raster-scanning stage which can provide  $61 \times 61$  pixel image with step size of  $1\text{ }\mu\text{m}$ .

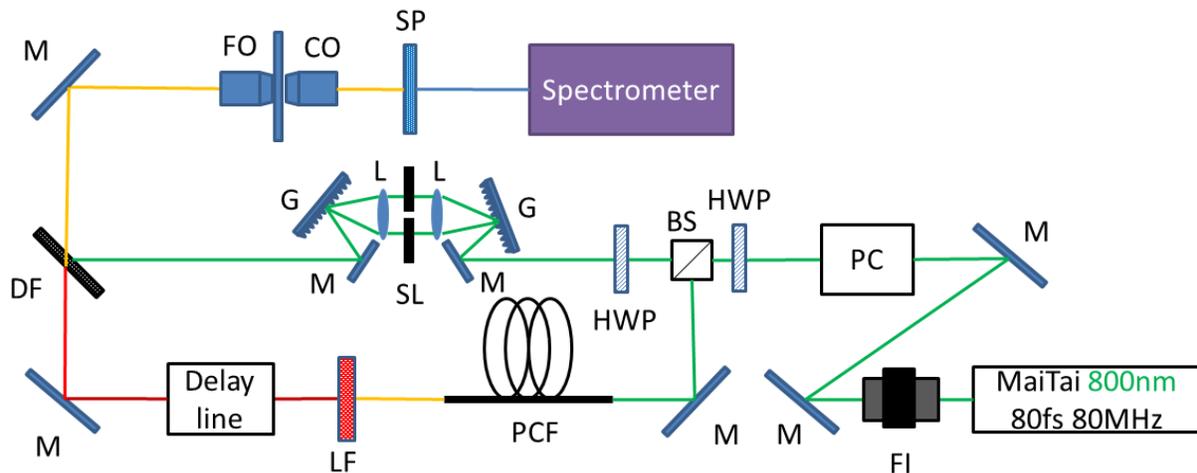


Figure 1. The schematic of HM-CARS imaging setup. FI: Faraday isolator, M: mirror, PC: prism compressor, HWP: half wave plate, BS: beam splitter, PCF: photonic crystal fiber, LF: long pass filter, G: grating, L: lens, SL: slit, DF: dichroic filter, FO: focus objective, CO: collimating objective, SP: short pass filter.

Here we briefly describe the key procedures involved in the reconstruction of HM-CARS images in our system. The intensity of CARS signal is proportional to the squared modulus of complex third-order susceptibility ( $|\chi^{(3)}(\omega)|^2$ ) which has the following relationship [2]:

$$|\chi^{(3)}(\omega)|^2 = |\chi_{NR}^{(3)}|^2 + 2\chi_{NR}^{(3)} \operatorname{Re}[\chi_R^{(3)}(\omega)] + |\chi_R^{(3)}(\omega)|^2 \quad (1)$$

The first term is non-resonant background whereas the third term is resonant part. Because of the mixing between the resonant and non-resonant signal as shown in the second term, the non-resonant background is not additive and is often strong enough to overwhelm or distort the desired resonant part. Among different techniques, we here apply a phase retrieval algorithm based on Kramers–Kronig transform, which is proven to be effective to suppress the non-resonant background [7]:

$$\varphi(\omega) = -\frac{P}{\pi} \int_{-\infty}^{+\infty} \frac{\ln|\chi^{(3)}(\omega'')|}{\omega'' - \omega} d\omega'' \quad (2)$$

Based on the convolution theorem and the Fourier transform of the Heaviside function, we can express the phase by the following equation:

$$\varphi(\omega) = 2 \operatorname{Im} \left\{ \psi(\ln|\chi^{(3)}(\omega)|) - \frac{\ln|\chi^{(3)}(\omega)|}{2} \right\} \quad (3)$$

where  $\psi$  is an operator defined as  $\psi(f(\omega)) = \Gamma[\eta(t : f(\omega))]$ . Consider that the non-resonant background is not flat,  $\eta(t : f(\omega))$  is defined as:

$$\eta(t : f(\omega)) = \begin{cases} \Gamma^{-1}[f(\omega)], t \geq 0 \\ \Gamma^{-1}[f_{NR}(\omega)], t < 0 \end{cases} \quad (4)$$

where  $\Gamma$  and  $\Gamma^{-1}$  are the operator of Fourier and inverse Fourier transform, respectively.  $f_{NR}(\omega)$  is the non-resonant background of CARS signal. Thus the imaginary part of the resonant signal  $\operatorname{Im}\{\chi^{(3)}(\omega)\} = |\chi^{(3)}(\omega)| \sin[\varphi(\omega)]$  is extracted by this phase retrieval algorithm.

The HM-CARS system typically provides a spectral window larger than  $2000 \text{ cm}^{-1}$ . Such a wide spectrum often contains highly congested Raman signatures and is not straightforward to interpret unless the pure Raman spectra of the sample composition are known in advance. Here, we apply another algorithm in our image processing steps, called principle component analysis (PCA), which can efficiently express the main unknown Raman information from the sample, without *a priori* knowledge. Figure 2 shows the schematic of PCA data processing for a large data set of CARS spectra. The accent of PCA here is to represent the raw CARS spectrum at each pixel by a linear combination of spectral vectors ( $\vec{P}_1, \vec{P}_2, \vec{P}_3 \dots$ ) with the linear coefficients (or scores) ( $S_1, S_2, S_3 \dots$ ). In our HM-CARS system, we collect  $3721 \times 1340$  data set for  $61 \times 61$  pixels per image (each spectrum has 1340 data points). Then we generate a covariance matrix from this two-dimensional data set by the following equation [8].

$$C^{m \times m} = \frac{1}{(n-1)} \sum_{k=0}^n (I_k - \bar{I})(J_k - \bar{J}), \quad (5)$$

where  $m$  is the number of spectral data points per spectrum and  $n$  is the total number of the image pixels.  $\bar{I}$  and  $\bar{J}$  are the mean values of row vector  $I$  and column vector  $J$ , respectively. The eigenvectors and eigenvalues of the covariance matrix can then be easily calculated, which is performed by the custom code written in MATLAB. Then each CARS spectrum can be expressed by the linear combination of these eigenvectors based on linear algebra theory. By sorting the eigenvectors in descending order of the corresponding eigenvalues, the most important component vectors are selected out. Thus we can express the main features of the sample by the corresponding scores of these vectors.

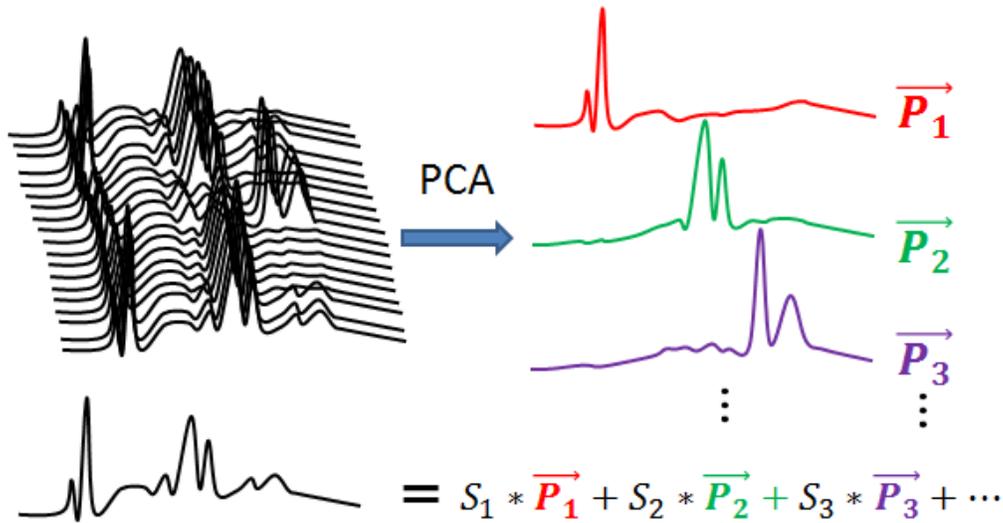


Figure 2. Illustration of the major procedures of PCA for HM-CARS spectral data processing.

### 3. RESULTS AND DISCUSSION

As a proof-of-principle demonstration, here we use a colon tissue section of mice as the biological sample for histological study by using our HM-CARS system. We scan a colon tissue in the area of  $60\mu\text{m}\times 60\mu\text{m}$  in step size of  $1\mu\text{m}$  by our HM-CARS system. Figure 3 (a) shows the CARS image at the Raman shift of  $2862\text{ cm}^{-1}$  which reveals the outline of intestinal crypts. Based on the raw spectral data set, we perform PCA only and the first principal component gives rise to an image (Fig. 3(b)) resembling the single-Raman-shift image (Fig. 3(a)). To further improve the contrast by minimizing the non-resonant background, we perform the phase retrieval algorithm on the raw data before the PCA procedure. The resultant image based on first principal component shows enhanced image contrast (Figure 3 (c)). The raw CARS spectrum at one pixel with a “x” mark in Figure 3(a) is plotted in Figure 3 (d) (see the blue curve). The black curve in this figure is the non-resonant background originated from the glass substrate. By performing the phase retrieval algorithm, we can extract the imaginary part of CARS susceptibility, as shown by the red curve in Figure 3 (d) –As a result, this shows the importance of implementing the phase retrieval algorithm together with the PCA in the image reconstruction procedures in order to ensure high-contrast hyperspectral Raman image information. As a comparison, figure 3(e) shows the H&E stained image of the tissue. The red bracket area in the microscopic image represents the imaging area of the HM-CARS system.

Figure 4 (a-c) shows the HM-CARS images of colon tissue based on the first three principle components using the algorithm of PCA after phase retrieval. The images of the first, second and third components are false-colored by red (R), green (G) and blue (B), respectively. The corresponding principle component vectors are also plotted in red, green and blue curves in Figure 4(d) – each of which represents different chemical components in colon tissue. The merged image is shown in Figure 4(e). According to histological study of colon tissue image, the bright red color area shows the extracellular matrix of colon tissue which mainly contains proteoglycans, sulfate and hyaluronan in the stroma. The nuclei which are composed of nucleic acid (DNA and RNA) are mostly distributed in bright green area. The blue parts in the image represent cytoplasm in colon tissue which typically consists of cytosol, organelles and inclusions. Hence, the

complex chemical composition in the biological tissue can be read out by different principle components in the PCA process – enabled only the HM-CARS technique providing rich broadband spectra across the tissue.

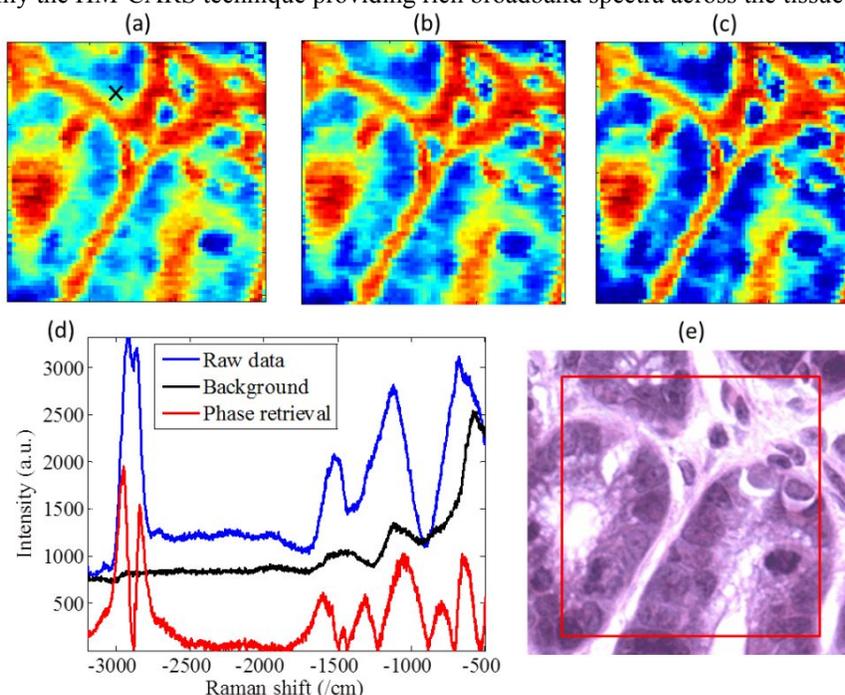


Figure 3 (a) CARS image at  $2862\text{ cm}^{-1}$  Raman shift. (b) CARS image of first principal component after PCA. (c) CARS image of first principal component after phase retrieval and PCA. (d) CARS spectrum of raw data (blue), non-resonant background (black) and after phase retrieval (red) at the position with a “x” mark in (a). (e) Microscope image of colon cells in the same region.

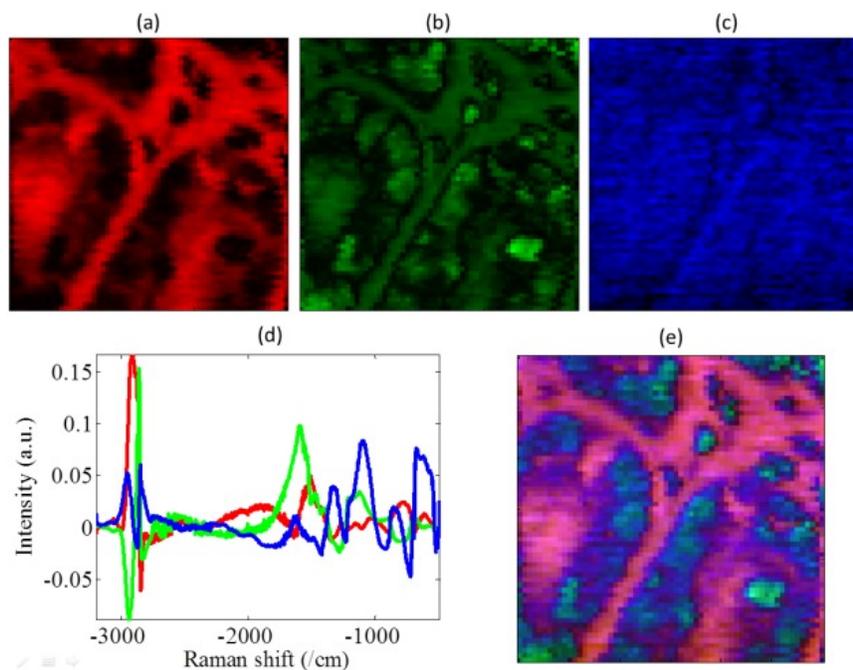


Figure 4 (a-c) The first, second, third principle component images shown in false-color representation: red (R), green (G), blue (B), respectively. (d) The corresponding first, second, third component vectors. (e) The merged RGB image of (a-c).

#### 4. SUMMARY

We demonstrate HM-CARS microscopy for stain-free histological imaging, which achieves a broad Raman spectral coverage  $> \sim 2000 \text{ cm}^{-1}$ . Employing spectral phase retrieval and principle component analysis, our system is capable of accessing spectral signatures in both the fingerprint ( $500\text{--}1800 \text{ cm}^{-1}$ ) and the CH/OH regions ( $2800\text{--}3000 \text{ cm}^{-1}$ ). We show that unraveling the wideband signature enables truly chemoselective coherent Raman imaging of a colon tissue section, displaying the key cellular structures that correspond well with standard stained-tissue observation. HM-CARS would thus be a viable approach for realizing stain-free histopathology, minimizing routine and laborious tissue processing.

#### 5. ACKNOWLEDGEMENT

The work in this Letter is partially supported by grants from the Research Grants Council of Hong Kong Special Administrative Region, China (HKU 717212E, HKU 717510E, HKU 717911E, and HKU 720112E), and the University Development Fund of HKU.

#### REFERENCES

- [1] Zumbusch A., Holtom G. R. and Xie X. S., "Three-dimensional vibrational imaging by coherent anti-Stokes Raman scattering," *Phys. Rev. Lett.* 82(20), 4142-4145 (1999).
- [2] Evans C. L., and Xie X. S., "Coherent Anti-Stokes Raman Scattering Microscopy: Chemical Imaging for Biology and Medicine," *Annu. Rev. Anal. Chem.* 1, 883-909 (2008).
- [3] Evans C. L., Potma E. O., Puoris M., Côté D., Lin C. P. and Xie X. S., "Chemical imaging of tissue in vivo with video-rate coherent anti-Stokes Raman scattering microscopy " *Proc. Natl. Acad. Sci. U.S.A.* 102(46), 16807-16812 (2005).
- [4] Potma E. O., Jones D. J., Cheng J. X., Xie X. S. and Ye J., "High-sensitivity coherent anti-Stokes Raman scattering microscopy with two tightly synchronized picosecond lasers," *Opt. Lett.* 27(13), 1168-1170 (2002).
- [5] Paulsen H. N., Hilligsoe K. M., Thogersen J., Keiding S. R. and Larsen J. J., "Coherent anti-Stokes Raman scattering microscopy with a photonic crystal fiber based light source," *Opt. Lett.* 28(13), 1123-1125 (2003).
- [6] Nan X. L., Cheng J. X. and Xie X. S., "Vibrational imaging of lipid droplets in live fibroblast cells with coherent anti-Stokes Raman scattering microscopy," *J. Lipid Res.* 44(11), 2202-2208 (2003).
- [7] Liu Y., Lee Y. J. and Cicerone M. T., "Broadband CARS spectral phase retrieval using a time-domain Kramers-Kronig transform," *Opt. Lett.* 34(9), 1363-1365 (2009).
- [8] Lin C. Y., Suhalim J. L., Nien C. L., Miljkovic M. D., Diem M., Jester J. V. and Potma E. O., "Picosecond spectral coherent anti-Stokes Raman scattering imaging with principal component analysis of meibomian glands," *J. Biomed. Opt.* 16(2), 021104 (2011).