Research Article | Vol. 25, No. 14 | 10 Jul 2017 | OPTICS EXPRESS 16534
Optics EXPRESS

Temperature extraction in Brillouin optical time-domain analysis sensors using principal component analysis based pattern recognition

ABUL KALAM AZAD,[1,2,5] FAISAL NADEEM KHAN,[3,*] WALED HUSSEIN ALARASHI,[4] NAN GUO,[2] ALAN PAK TAO LAU,[3] AND CHAO LU[2]

[1]*Department of Electrical and Electronic Engineering, University of Dhaka, Dhaka-1000, Bangladesh*
[2]*Department of Electronic and Information Engineering, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong*
[3]*Department of Electrical Engineering, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong*
[4]*Department of Electronic Engineering, University of Science and Technology, Sana'a, Yemen*
[5]*azad.abulkalam@connect.polyu.hk*
**fnadeem.khan@yahoo.com*

Abstract: We propose and experimentally demonstrate the use of principal component analysis (PCA) based pattern recognition to extract temperature distribution from the measured Brillouin gain spectra (BGSs) along the fiber under test (FUT) obtained by Brillouin optical time domain analysis (BOTDA) system. The proposed scheme employs a reference database consisting of relevant ideal BGSs with known temperature attributes. PCA is then applied to the BGSs in the reference database as well as to the measured BGSs so as to reduce their size by extracting their most significant features. Now, for each feature vector of the measured BGS, we determine its best match in the reference database comprised of numerous reduced-size feature vectors of the ideal BGSs. The known temperature attribute corresponding to the best-matched BGS in the reference database is then taken as the extracted temperature of the measured BGS. We analyzed the performance of PCA-based pattern recognition algorithm in detail and compared it with that of curve fitting method. The experimental results validate that the proposed technique can provide better accuracy, faster processing speed and larger noise tolerance for the measured BGSs. Therefore, the proposed PCA-based pattern recognition algorithm can be considered as an attractive method for extracting temperature distributions along the fiber in BOTDA sensors.

OCIS codes: (060.2370) Fiber optics sensors; (060.4370) Nonlinear optics, fibers; (290.5900) Scattering, stimulated Brillouin; (070.5010) Pattern recognition.

References and links

1. X. Bao and L. Chen, "Recent Progress in Distributed Fiber Optic Sensors," Sensors (Basel) **12**(7), 8601–8639 (2012).
2. C. A. Galindez-Jamioy and J. M. Lopez-Higuera, "Brillouin distributed fiber sensors: an overview and applications," J. Sens. **2012**, 204121 (2012).
3. M. A. Soto, G. Bolognini, F. Di Pasquale, and L. Thévenaz, "Simplex-coded BOTDA fiber sensor with 1 m spatial resolution over a 50 km range," Opt. Lett. **35**(2), 259–261 (2010).
4. Y. Mao, N. Guo, K. L. Yu, H. Y. Tam, and C. Lu, "1-cm-Spatial-Resolution Brillouin Optical Time-Domain Analysis Based on Bright Pulse Brillouin Gain and Complementary Code," IEEE Photonics J. **4**(6), 2242–2248 (2012).
5. Y. Muanenda, M. Taki, and F. D. Pasquale, "Long-range accelerated BOTDA sensor using adaptive linear prediction and cyclic coding," Opt. Lett. **39**(18), 5411–5414 (2014).
6. M. Niklès, L. Thévenaz, and P. A. Robert, "Brillouin gain spectrum characterization in single-mode optical fibers," J. Lightwave Technol. **15**(10), 1842–1851 (1997).
7. A. Minardo, E. Catalano, and L. Zeni, "Cost-effective method for fast Brillouin optical time-domain analysis," Opt. Express **24**(22), 25424–25431 (2016).
8. C. Li and Y. Li, "Fitting of Brillouin spectrum based on LabVIEW," in *Proceedings of 5th International Conference on Wireless Communications, Networking and Mobile Computing* (IEEE, 2009), pp. 3495–3498.

#292313 https://doi.org/10.1364/OE.25.016534
Journal © 2017 Received 6 Apr 2017; revised 18 May 2017; accepted 19 May 2017; published 3 Jul 2017

Optics EXPRESS

9. M. A. Farahani, E. Castillo-Guerra, and B. G. Colpitts, "A Detailed Evaluation of the Correlation-Based Method Used for Estimation of the Brillouin Frequency Shift in BOTDA Sensors," IEEE Sens. J. **13**(12), 4589–4598 (2013).
10. F. Wang, W. Zhan, Y. Lu, Z. Yan, and X. Zhang, "Determining the change of Brillouin frequency shift by using the similarity matching method," J. Lightwave Technol. **33**(19), 4101–4108 (2015).
11. A. K. Azad, L. Wang, N. Guo, H. Y. Tam, and C. Lu, "Signal processing using artificial neural network for BOTDA sensor system," Opt. Express **24**(6), 6769–6782 (2016).
12. A. K. Azad, L. Wang, N. Guo, C. Lu, and H. Y. Tam, "Temperature sensing in BOTDA system by using artificial neural network," Electron. Lett. **51**(20), 1578–1580 (2015).
13. R. Ruiz-Lomber, J. M. Serrano, and J. M. Lopez-Higuera, "Automatic strain detection in a Brillouin optical time domain sensor using principal component analysis and artificial neural networks," in *Proceedings of IEEE Conference on* Sensors (IEEE, 2014), pp. 1539–1542.
14. F. N. Khan, Y. Yu, M. C. Tan, W. H. Al-Arashi, C. Yu, A. P. T. Lau, and C. Lu, "Experimental demonstration of joint OSNR monitoring and modulation format identification using asynchronous single channel sampling," Opt. Express **23**(23), 30337–30346 (2015).
15. F. Davò, S. Alessandrini, S. Sperati, L. D. Monache, D. Airoldi, and M. T. Vespucci, "Post-processing techniques and principal component analysis for regional wind power and solar irradiance forecasting," Elsevier J. Solar. Energy **134**, 327–338 (2016).
16. F. N. Khan, C. H. Teow, S. G. Kiu, M. C. Tan, Y. Zhou, W. H. Al-Arashi, A. P. T. Lau, and C. Lu, "Automatic modulation format/bit rate classification and signal-to-noise ratio estimation using asynchronous delay-tap sampling," Elsevier J. Com. Elect. Eng. **47**, 126–133 (2015).
17. M. C. Tan, F. N. Khan, W. H. Al-Arashi, Y. Zhou, and A. P. Tao Lau, "Simultaneous optical performance monitoring and modulation format/bit-rate identification using principal component analysis," J. Opt. Commun. Netw. **6**(5), 441–448 (2014).
18. M. Alem, M. A. Soto, and L. Thévenaz, "Analytical model and experimental verification of the critical power for modulation instability in optical fibers," Opt. Express **23**(23), 29514–29532 (2015).
19. R. Ruiz-Lombera, J. Urricelqui, M. Sagues, J. Mirapeix, J. M. López-Higuera, and A. Loayssa, "Overcoming nonlocal effects and Brillouin threshold limitations in Brillouin optical time-domain sensors," IEEE Photonics J. **7**(6), 6803609 (2015).
20. I. T. Jolliffe, *Principal Component Analysis* (Springer, 2002).
21. A. C. Rencher and W. F. Christensen, *Methods of multivariate analysis* (John Wiley and Sons, 2012).
22. M. Turk and A. Pentland, "Eigenfaces for recognition," J. Cogn. Neurosci. **3**(1), 71–86 (1991).
23. V. Perlibakas, "Distance measures for PCA-based face recognition," Pattern Recognit. Lett. **25**(6), 711–724 (2004).
24. A. Motil, R. Hadar, I. Sovran, and M. Tur, "Gain dependence of the linewidth of Brillouin amplification in optical fibers," Opt. Express **22**(22), 27535–27541 (2014).
25. X. Bao, A. Brown, M. Demerchant, and J. Smith, "Characterization of the Brillouin-loss spectrum of single-mode fibers by use of very short (<10-ns) pulses," Opt. Lett. **24**(8), 510–512 (1999).
26. S. Xie, M. Pang, X. Bao, and L. Chen, "Polarization dependence of Brillouin linewidth and peak frequency due to fiber inhomogeneity in single mode fiber and its impact on distributed fiber Brillouin sensing," Opt. Express **20**(6), 6385–6399 (2012).
27. M. A. Soto and L. Thévenaz, "Modeling and evaluating the performance of Brillouin distributed optical fiber sensors," Opt. Express **21**(25), 31347–31366 (2013).
28. A. Lopez-Gil, M. A. Soto, X. Angulo-Vinuesa, A. Dominguez-Lopez, S. Martin-Lopez, L. Thévenaz, and M. Gonzalez-Herraez, "Evaluation of the accuracy of BOTDA systems based on the phase spectral response," Opt. Express **24**(15), 17200–17214 (2016).
29. X. Wang, A. Mueen, H. Ding, G. Trajcevski, P. Scheuermann, and E. Keogh, "Experimental comparison of representation methods and distance measures of time series data," Data Min. Knowl. Discov. **26**(2), 275–309 (2013).

1. Introduction

Brillouin optical time domain analysis (BOTDA) sensor systems have attracted ample research interest during the last few decades as they can provide distributed temperature and strain measurement with high accuracy and resolution over a long sensing fiber [1–5]. These systems also offer special advantages such as longer durability, immunity to electromagnetic interference and can be used for remote sensing applications in hazardous environments [1]. In the Brillouin gain configuration of such systems, the pump pulse with higher frequency enters at the near end while the CW probe signal with lower frequency is launched at the far end of the fiber under test (FUT). The probe signal is amplified due to the interaction of these two counter-propagating signals through the process of stimulated Brillouin scattering (SBS). The gain experienced by the probe signal becomes maximum for a pump-probe frequency difference equal to the local Brillouin peak gain frequency, also called Brillouin frequency

Optics EXPRESS

shift (BFS) of the FUT. The local Brillouin gain spectrum (BGS) along the FUT can then be reconstructed after retrieving the BOTDA traces by scanning the pump-probe frequency difference step-by-step in the vicinity of local BFS.

In BOTDA sensing, the BFSs of the measured BGSs along the FUT are usually recovered by employing the curve fitting method (CFM) where a Lorentzian curve is fitted on to a measured BGS and its BFS is estimated to be the frequency of the maximum amplitude on the fitted curve [6–9]. The BFSs along the FUT obtained after employing CFM is then converted to temperature distribution using the BFS-temperature characteristics of the FUT. The accuracy of CFM depends not only on the level of noise in the measured BGSs but also on the initialization of parameters pertaining to the fitting process [8,9]. Moreover, this method requires the optimization of fitting parameters iteratively and hence takes significant time to process the measured BGSs. In order to overcome the limitations of CFM, many alternative processing techniques have been proposed in recent years for improving the accuracy and processing speed of temperature extraction in BOTDA systems. For example, the study conducted in [9] reports the use of cross-correlation based method and that in [10] exploits the use of correlation based spectrum matching for fast and accurate estimation of BFS. In addition, the studies conducted in [11,12] also demonstrate the use of artificial neural network as a signal processing tool to determine the temperature distribution along the FUT with good accuracy and processing speed without employing any curve fitting process.

Within this scope, in this work we propose a new technique to extract temperature distribution from the measured BGSs along the FUT by using principal component analysis (PCA) based pattern recognition. PCA has been widely employed as a data preprocessing tool for reducing the dimensionality of data in different fields of science and engineering in recent years [13–17]. In our proposed approach, first a reference database of relevant BGSs of known attributes (e.g., temperature) is constructed. Next, PCA is applied to extract the limited but most significant feature vectors of all the BGSs in this database. The compact representation of BGSs, obtained using PCA, aides in reducing the computational complexity of pattern recognition algorithm by several orders of magnitude. To extract temperature for a measured BGS attained from a BOTDA experiment, it is first processed using PCA to obtain its reduced-size feature vector and then its best match in the reference database is determined based on statistical distance measurement. Once the matching process is over, the known temperature attribute of the best-matched BGS in the reference database is considered to be the extracted temperature of the measured BGS. In this paper, the performance of proposed technique for extracting temperature distribution from the measured BGSs is compared with that of widely-used CFM. The results demonstrate that the temperature extraction using PCA-based pattern recognition leads to better accuracy and larger tolerance to measurement errors. Moreover, the use of reduced-size feature vectors obtained through PCA decreases the computational complexity (as well as time) of the matching process significantly, thereby making the proposed technique suitable for fast monitoring applications. Thus, the proposed PCA-based pattern recognition method can be used as a promising tool for accurate and fast monitoring of temperature distributions in BOTDA sensing systems.

2. Experimental setup

The experimental scheme of the BOTDA sensor is shown in Fig. 1. The pump and probe signals are generated in the two branches using the light from the continuous wave (CW) tunable laser operating at 1550 nm after being split by the coupler. The radio frequency (RF) signal applied to the electro-optic modulator (EOM1) in the upper branch is used to generate a double sideband suppressed carrier probe signal with extinction ratio above 30 dB. The variable optical attenuator (VOA) in the upper branch is used to control the probe power before launching into the FUT and the isolator is used to block the signal from the reverse direction. The CW signal at the lower branch is also modulated by another EOM2 driven by a pulse pattern generator (PPG) for the generation of pump pulses which are then boosted up by

Optics EXPRESS

the erbium-doped fiber amplifier (EDFA). The band pass filter (BPF) in the lower branch is used to filter out the amplified spontaneous emission (ASE) noise and then a polarization scrambler (PS) is utilized to minimize the polarization dependent fading of Brillouin gain. The pump and probe powers are adjusted to avoid undesirable spectral distortions arising from modulation instability and non-local effects [18,19]. The probe signal amplified through the process of SBS is detected by a photodetector (PD) after the unwanted higher-frequency sideband is filtered out by a fiber Bragg grating (FBG) filter. The BGSs along the FUT are reconstructed from the BOTDA time-domain traces obtained by scanning the pump-probe frequency difference at a given frequency step around the local BFS. The measured BGSs are then processed to extract temperature distribution along the FUT.

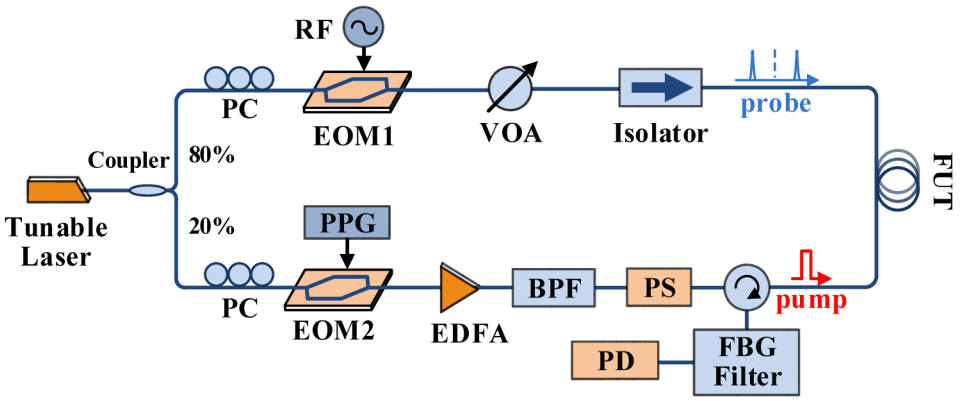

Fig. 1. BOTDA experimental setup, PC: Polarization Controller, EOM: Electro-Optic Modulator, RF: Radio Frequency, VOA: Variable Optical Attenuator, PPG: Pump Pattern Generator, EDFA: Erbium-doped Fiber Amplifier, BPF: Band Pass Filter, PS: Polarization Scrambler, FBG: Fiber Bragg Grating, PD: Photodetector, FUT: Fiber Under Test.

3. Operating principle

3.1 Principal component analysis based pattern recognition

Principal component analysis, also known as Karhunen-Loève transform (KLT), is a powerful statistical method for features extraction and data dimensionality reduction [20–23]. PCA decreases the large dimensionality of a data space to a comparatively smaller dimensionality of feature space. Such a reduction in dimensionality is obtained through orthogonal transformation to a new (and reduced-size) set of uncorrelated variables known as principal components (PCs). This enables PCA to attain a more concise and economical representation of data. In this work, we employ PCA and statistical distance measurement based pattern recognition to extract temperature distribution from the measured BGSs along the FUT obtained from BOTDA measurement. The acquisition of measured BGSs from BOTDA experiment requires the scanning of pump-probe frequency difference over a wide range (e.g., a few hundred MHz) using small frequency steps (e.g., 1 or 2 MHz). If the measured BGSs are obtained over a frequency range of $\upsilon_R = \upsilon_{stop} - \upsilon_{start}$ with a frequency step of $\Delta\upsilon$, each BGS will be of length $N = [\upsilon_R/\Delta\upsilon] + 1$ where N can be of the order of a few hundred depending on υ_R and $\Delta\upsilon$. Now, for the statistical distance measurement based matching process, we also need to construct the reference database R having ideal BGSs of same length N. For large values of N (e.g., a few hundred), the computational complexity of matching process will be quite high. The purpose of applying PCA to BGSs in database R as well as to measured BGSs along the FUT is to obtain their corresponding reduced-size feature vectors without significant loss of information. In this way, the computational complexity of the matching process will be reduced significantly.

Optics EXPRESS

In the proposed technique, the reference database R is constructed using a total of M ideal BGSs, each of length N. If the BGSs are represented as vectors g_i, where $i = 1, 2, \ldots, M$, all BGSs in the reference database can be expressed together as a matrix $R = [g_1\ g_2\ \ldots\ g_M]$ of size $N \times M$, where each BGS forms one column of R. Now, the mean vector $\bar{g}$ of R is defined as

$$\bar{g} = \frac{1}{M}\sum_{i=1}^{M} g_i \tag{1}$$

Next, we obtain zero-mean matrix $\Psi = [\psi_1\ \psi_2\ \ldots\ \psi_M]$ where $\Psi_i = g_i - \bar{g}$. The covariance matrix C of Ψ can be determined using

$$C = \frac{1}{M}\sum_{i=1}^{M} \psi_i \psi_i^T = \Psi\Psi^T \tag{2}$$

where C can have a total of N eigenvectors (also known as PCs) and related eigenvalues which can be obtained using

$$C\mu_j = \lambda_j \mu_j \quad \text{for } j = 1, 2, \ldots, N \tag{3}$$

where μ_j is the jth eigenvector and λ_j is the jth eigenvalue of C. Next, we rank the computed eigenvectors according to their eigenvalues and select only P (where $P << N$) eigenvectors pertaining to the P largest eigenvalues while the others are discarded. The number of eigenvectors chosen P is such that the parameter Ω satisfies the condition

$$\Omega = \sum_{j=1}^{P} \lambda_j \Big/ \sum_{j=1}^{N} \lambda_j > \delta \tag{4}$$

where the value of parameter δ in Eq. (4) is often selected to be higher than 0.9 in most practical applications [17,20,21]. Now, for any given BGS in the reference database R, the vector ψ can be approximated as a weighted sum of the P chosen eigenvectors, i.e.,

$$\sum_{k=1}^{P} w_k \mu_k \approx \psi \quad \text{for } k = 1, 2, \ldots, P \tag{5}$$

where μ_k, for $k = 1, 2, \ldots, P$, are orthogonal eigenvectors according to the basic property of PCA [22]. This means

$$\mu_l \mu_m = \begin{cases} 1 & \text{if} \quad l = m \\ 0 & \text{if} \quad l \neq m \end{cases} \tag{6}$$

Now, multiplying both sides of Eq. (5) with μ_k^T and then using Eq. (6), we get

$$w_k = \mu_k^T \psi \quad \text{for } k = 1, 2, \ldots, P \tag{7}$$

A vector r encompassing the weights w_k, i.e.,

$$r = \left[w_1\ w_2 \ldots w_P\right]^T \tag{8}$$

is called the feature vector of a given BGS. Hence, by using Eq. (8), we can determine feature vectors r_i for $i = 1, 2, \ldots, M$, for all the M BGSs in the reference database. In this way, each BGS of size N in the original database R can now be efficiently represented by its corresponding reduced-size feature vector r of size P. Similarly, by employing the P eigenvectors obtained for the reference database R, we can compute the feature vectors s_i (also of size P) for $i = 1, 2, \ldots, L$, for a set S composed of L measured BGSs along the FUT.

Next, for each feature vector s in S, we determine its best match r in R, i.e., the one which has the minimum statistical distance D with the given feature vector s. For the matching process, we adopt the widely-used Euclidean distance measure [17,23] defined as

$$D(s,r) = \|s - r\| = \sqrt{\sum_{k=1}^{P} \left(s^k - r^k\right)^2} \tag{9}$$

where s^k and r^k are the individual elements of the feature vectors s and r, respectively. The known temperature attribute of the best-matched feature vector r in the reference database R can then be considered as the extracted temperature of the measured BGS. The process of PCA-based pattern recognition to extract temperature distribution from the measured BGSs is illustrated in Fig. 2.

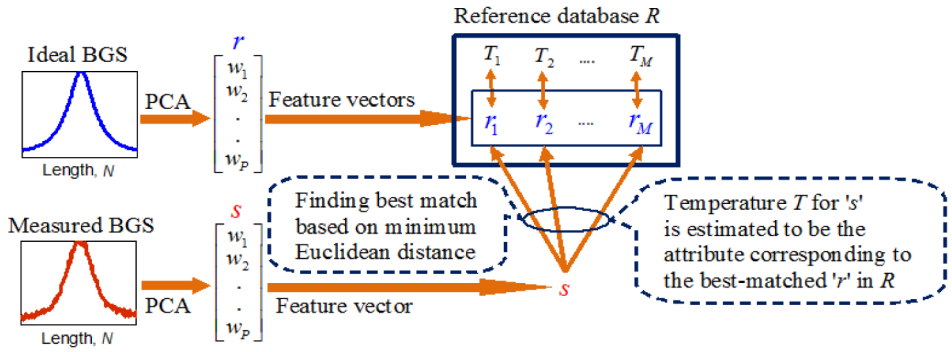

Fig. 2. Functional block diagram depicting temperature extraction from BOTDA measured BGS using PCA-based pattern recognition.

Since the BGS obtained from BOTDA measurement is ideally modeled by Lorentzian function [24,25], we construct reference database R using ideal BGSs with Lorentzian profile given by Eq. (10)

$$g(\upsilon) = \frac{g_B}{1 + 4\left[(\upsilon - \upsilon_B)/(\Delta\upsilon_B)\right]^2} \tag{10}$$

where g_B is the peak gain, υ_B is the BFS and $\Delta\upsilon_B$ is the linewidth of the spectrum. For obtaining the ideal BGSs corresponding to different temperatures, we simply use $g_B = 1$ in Eq. (10) while the values of υ_B are determined by the dynamic range of temperature extraction and that of $\Delta\upsilon_B$ depend on the experimental conditions, e.g., duration of pump pulse. We adopt the values of υ_B (BFSs of the ideal BGSs) in Eq. (10) from F = 10.834 GHz to 10.912 GHz at a step of α = 0.2 MHz. The BFSs of the ideal BGSs are then converted to their corresponding temperatures by using Eq. (11)

$$\upsilon_B(T) = C_T \Delta T + \upsilon_B(T_o) \tag{11}$$

where C_T is the temperature coefficient (slope) and ΔT is the difference in temperature for which the BFS of the local BGS changes from $\upsilon_B(T_o)$ to $\upsilon_B(T)$. For this conversion, we utilize the BFS-temperature characteristics of the FUT given in Fig. 3. The conversion provides a temperature range from ~0 °C to ~80 °C, i.e., the dynamic range of temperature extraction is ~80 °C. In this work, the local BGSs along the FUT are obtained by using the BOTDA setup shown in Fig. 1 and adopting the pump pulse of duration 20 ns. For such pump pulses, the linewidths $\Delta\upsilon_B$ of the measured BGSs are observed to have an average value of ~55 MHz and the linewidths vary around this average value due to several factors [6,24,26], especially at the end of the FUT.

Optics EXPRESS

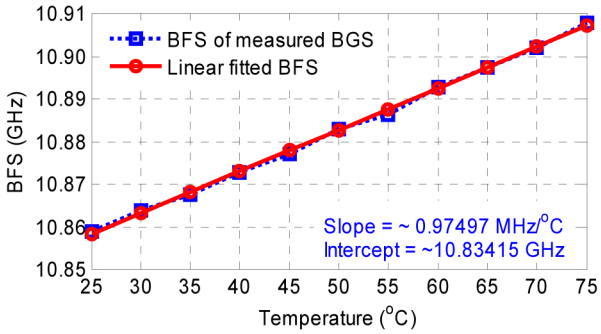

Fig. 3. BFS-temperature characteristics of the FUT. The whole span of a ~200 m FUT is heated inside the oven each time at different temperatures. Several local BGSs along the FUT for a given temperature are first normalized and then averaged to obtain one measured BGS for that specific temperature. The BFSs of the measured BGSs are determined by using CFM.

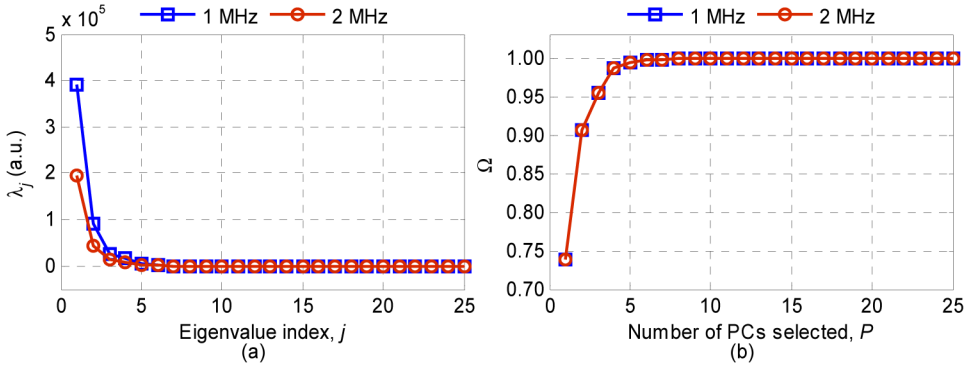

Fig. 4. (a) A total of $P = 25$ most significant eigenvalues λ_j plotted in descending order and (b) Parameter Ω versus number of PCs selected, P.

To compensate the effect of this variation in temperature extraction using PCA-based pattern recognition, we vary linewidth $\Delta\upsilon_B$ of each ideal BGS having a particular υ_B (from F = 10.834 GHz to 10.912 GHz, with step α = 0.2 MHz) from B = 40 MHz to 70 MHz, also at a step of α = 0.2 MHz, in Eq. (10). The frequency range υ_R for using Eq. (10) is chosen from υ = 10.76 GHz to 11.01 GHz, the same range we use to obtain the measured BGSs from BOTDA experiment. Hence, the reference database R consists of $M = [(F/\alpha) + 1] \times [(B/\alpha) + 1] = 391 \times 151$ ideal BGSs each having a particular temperature attribute for the chosen F and B ranges. PCA is then applied to represent each BGS of size N in the original database R with its corresponding feature vector r of much reduced-size P. Within the frequency range υ_R of 10.76 GHz to 11.01 GHz, we adopt two different frequency steps of $\Delta\upsilon$ = 1 and 2 MHz for constructing two separate databases so as to study the performance of the technique for the measured BGSs obtained using $\Delta\upsilon$ = 1 and 2 MHz, respectively. For both databases, the BGSs are normalized to make their peak gains equal to 1 before applying PCA to them separately. Figure 4(a) shows eigenvalues λ_j for a few PCs in descending order for the two reference databases from which it can be easily observed that the eigenvalues converge rapidly to zero. The parameter Ω given by Eq. (4) is also plotted as a function of number of PCs selected, P in Fig. 4(b) from which it is obvious that we can make use of only a few PCs (i.e., around 5 for both the cases under consideration) for the extraction of feature vectors r and can ignore the remaining without losing significant information. The use of reduced-size feature vectors r for the BGSs in the reference databases significantly reduces the computational complexity of the Euclidean distance-based matching process for pattern recognition.

Optics EXPRESS

Next, we apply PCA-based algorithm to extract temperature distribution from the measured BGSs along the FUT. For this purpose, we use BOTDA setup shown in Fig. 1 to obtain measured BGSs for the frequency range υ_R from υ = 10.76 GHz to 11.01 GHz, the same range we adopt to construct the reference databases. Within this range, we acquire BGSs with frequency step $\Delta\upsilon$ of 1 MHz which are then down-sampled to obtain BGSs for $\Delta\upsilon$ = 2 MHz. The BGSs obtained for both $\Delta\upsilon$ = 1 and 2 MHz frequency steps are normalized to make their peak gains equal to 1. Next, the reduced-size feature vectors s of the measured BGSs in S are obtained through PCA and are compared with the feature vectors r in the reference database R to directly extract the temperature distribution along the FUT without necessitating determination of BFS and the conversion from BFS to temperature.

In this study, we construct reference database R using ideal BGSs considering the BFS-temperature characteristics of the FUT shown in Fig. 3. For obtaining ideal BGSs using Eq. (10), the ranges of F and B are determined by the dynamic range of extracted temperature along the FUT (i.e., 80 °C) and the duration of pump pulse (i.e., 20 ns), respectively. Note that if another FUT with different BFS-temperature characteristics is used in the BOTDA setup shown in Fig. 1, the BFSs of local measured BGSs along the FUT will also be changed accordingly. Moreover, the linewidth of measured BGSs will also be different if the duration of pump pulse is changed [2,25]. In such cases, we will essentially need to reconstruct the reference database with modified F and B ranges based on the BFS-temperature characteristics of the new FUT, dynamic range of temperature extraction and the duration of pump pulse. However, we would like to mention that the construction of a new database is not a difficult or time consuming task. This is because the database is constructed with ideal BGSs which can simply be generated using Eq. (10). Hence, we can easily obtain a new database for a given configuration. Note that since the synthesis of database is performed offline and prior to actual temperature measurement, it does not really affect the response time of proposed method.

3.2 Curve fitting method

In order to compare the performance of PCA-based pattern recognition method with that of CFM, we also extract the temperature distributions along the FUT using least-square CFM where the ideal Lorentzian profile given by Eq. (10) is fitted on to the measured BGS and the frequency υ_B of the peak gain of the fitted curve is assumed to be the BFS of the BGS. Then we apply the linear relationship given by Eq. (11) to obtain temperature distributions along the FUT. A detailed description of the least-square CFM can be found in Ref [8,9,11].

4. Results and discussion

In our demonstration, a 38.2 km long FUT is used in the BOTDA setup shown in Fig. 1. A section of length ~600 m from the last part of the FUT is wound carefully by hand and placed inside an oven. The remaining part of the FUT is compactly wound on to the fiber mandrel and is put at room temperature (~25 °C) outside the oven. During each experiment, the temperature inside the oven is set at four different temperatures of 40 °C, 50 °C, 60 °C and 70 °C to acquire the BGSs along the whole span of the FUT. The temperature inside the oven is monitored by high precision thermometer. In the experiment, we adopt pump pulse of duration 20 ns which corresponds to a spatial resolution of 2 m. We acquire BGSs at a sampling interval of 0.4 m. To study the performance of two techniques, i.e., PCA-based pattern recognition and CFM for different levels of noise in the measured BGSs, we adopt different number of BOTDA trace averaging N_{av} starting from N_{av} = 100 to 1000 with a step of 100. For instance, the 3-D distribution of the measured BGSs acquired at 1 MHz frequency step and N_{av} = 1000 is shown in Fig. 5 for an oven temperature of 70 °C. In Fig. 5, the change of BFSs of the BGSs along the heated section is clearly distinguishable from the remaining part of the FUT placed at room temperature outside the oven.

Optics EXPRESS

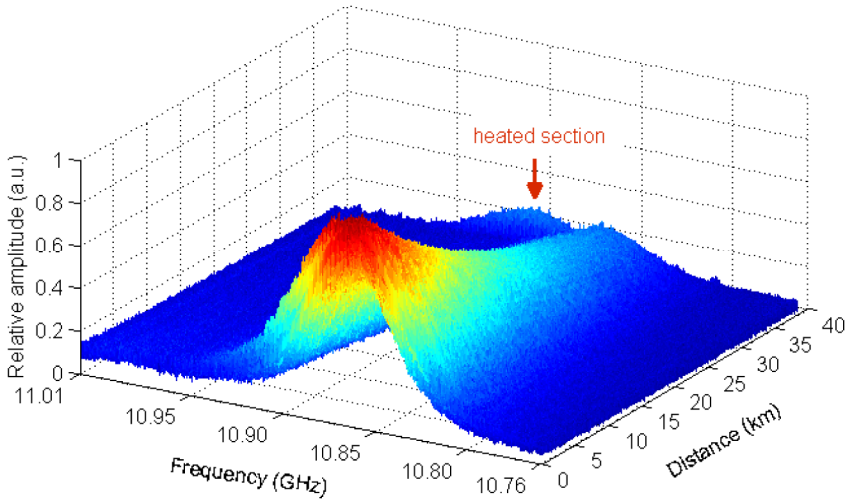

Fig. 5. Distribution of BGSs along the 38.2 km long FUT obtained from BOTDA measurement using pump pulse of duration 20 ns, trace averaging of 1000, sampling interval of 0.4 m and frequency step of 1 MHz with last ~600 m section of the FUT heated inside the oven at 70 °C.

In BOTDA sensors, the accuracy of temperature extraction is directly related to the level of noise in the measured BGSs. The averaging of BOTDA traces is a common method that helps to improve the signal-to-noise ratio (SNR) by the square root dependency on N_{av} [27]. In order to study the effect of noise on the accuracy of temperature extraction, we compute SNRs for the measured BGSs along the FUT by comparing each local noisy BGS with its fitted curve obtained after employing CFM. As depicted in Fig. 6, we consider peak gain g_B

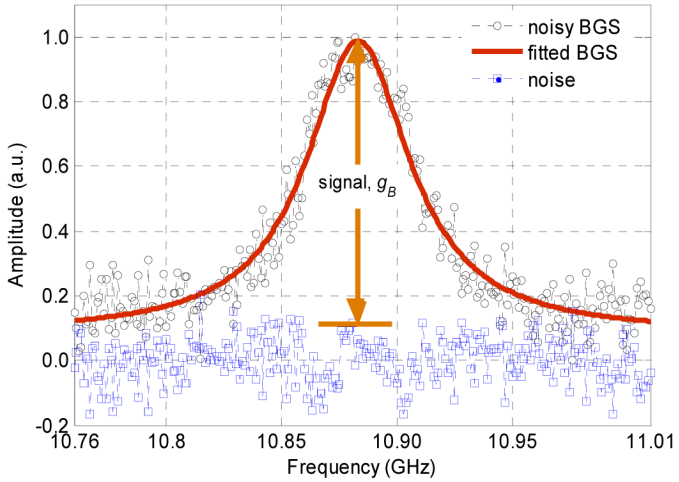

Fig. 6. A typical normalized noisy BGS near the end of 38.2 km long FUT and its fitted curve obtained after employing CFM. The noise is computed by subtracting the noisy BGS from the fitted one. The noisy BGS is obtained from BOTDA experiment using pump pulse of duration 20 ns, frequency step of 1 MHz and trace averaging of 1000. The fitted curve gives $g_B \approx 0.867$, $\upsilon_B \approx 10.8832$ GHz and $\Delta\upsilon_B \approx 54.46$ MHz.

of the fitted curve as the 'signal' while the residuals from the fitted curve are considered as 'noise' [9,28]. The SNR of a local BGS is then computed by using Eq. (12)

Optics EXPRESS

$$\mathrm{SNR} = \frac{g_B^2}{\sigma_n^2} \tag{12}$$

where σ^2_{n} is the variance of the noise. We determine SNR distributions for the measured BGSs along the FUT for each N_{av} with the last section of FUT heated inside an oven at four different temperatures i.e, 40 °C, 50 °C, 60 °C and 70 °C. For instance, the SNR distributions for the measured BGSs obtained at 1 MHz frequency step with the last ~600 m section of 38.2 km long FUT heated inside the oven at 40 °C are shown in Fig. 7(a) for different trace averaging, i.e., N_{av} = 100, 300, 500 and 1000. Next, for each N_{av} adopted in this study, we separately calculate average SNRs for the BGSs along the last 500 m (i.e., 37.7 km to 38.2 km) section of the FUT heated inside the oven at four different temperatures. Since for a particular N_{av}, the average SNRs for this section of the FUT heated at four different temperatures are almost the same, we take their average value to estimate the SNR for that particular N_{av}. The variation of SNR with N_{av} is shown in Fig. 7(b).

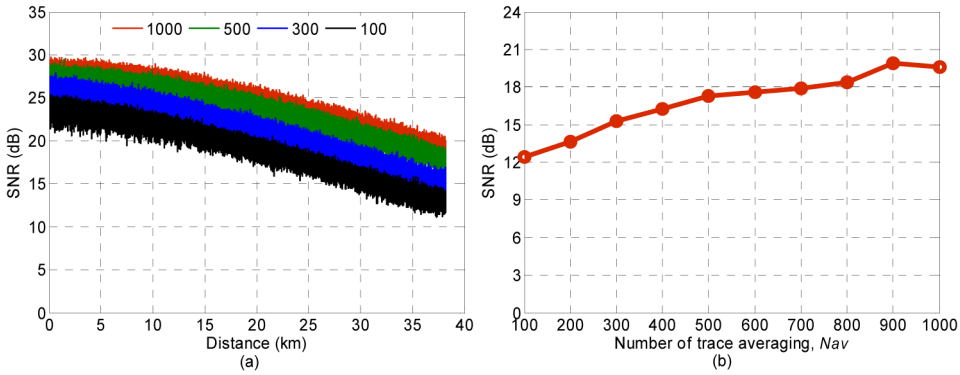

Fig. 7. (a) SNR distributions for the BGSs along the 38.2 km FUT with last ~600 m section of the FUT heated inside the oven at 40 °C, and (b) Average SNR calculated for the BGSs along the last 500 m (i.e., 37.7 km to 38.2 km) section of the FUT.

In order to extract temperature distributions from the measured BGSs using PCA-based pattern recognition, we synthesize feature vectors for the BGSs in the reference databases as well as for the measured BGSs obtained by using the BOTDA setup shown in Fig. 1. The number of PCs P used for the synthesis of feature vectors is selected based on the condition given by Eq. (4). It is clear from Fig. 4(b) that the value of parameter Ω is close to 1 for $P >$ 5. To select optimum number of PCs P, we adopt different P starting from 1 to 25 for the synthesis of feature vectors and extract temperature distributions for the measured BGSs obtained for each N_{av} and $\Delta\upsilon$. Then, for each P value, we compute root mean square error (RMSE) for the estimated temperatures along the last 500 m section of the FUT (which is heated inside the oven) by comparing the actual temperature measured by the thermometer and the one extracted by PCA-based pattern recognition method. For a particular P, N_{av} and $\Delta\upsilon$, four nearly equal RMSE values are obtained corresponding to four different fiber temperatures of 40 °C, 50 °C, 60 °C and 70 °C which are then averaged to compute the mean RMSE for that particular P, N_{av}, and $\Delta\upsilon$. For instance, the RMSEs corresponding to different P values for N_{av} = 100, 300, 500 and 1000 are plotted in Figs. 8(a) and 8(b) for the measured BGSs obtained using two different frequency steps of $\Delta\upsilon$ = 1 and 2 MHz, respectively. It is observed from Fig. 8 that the RMSEs increase gradually with a decrease in N_{av} (i.e., lower SNR) and are the worst for N_{av} = 100 which is the lowest trace averaging adopted in our experiment. On the other hand, the RMSEs for a particular N_{av} decrease sharply when the number of PCs selected is increased from P = 1 to 5 and become almost stable for $P \geq 8$. Therefore, in order to ensure minimum computational complexity for the matching process as

Optics EXPRESS

well as to obtain optimum accuracy, we finally chose P = 9 for the synthesis of feature vectors for the BGSs in the two reference databases (constructed for two different frequency steps $\Delta\upsilon$ of 1 and 2 MHz) as well as for the measured BGSs along the FUT obtained also for $\Delta\upsilon$ = 1 MHz and 2 MHz. The temperature distributions attained by using PCA-based pattern recognition method with P = 9 are shown in Figs. 9 and 10 for the BGSs obtained with N_{av} = 1000 for two different frequency steps of $\Delta\upsilon$ = 1 MHz and 2 MHz, respectively. The temperature distributions given by CFM are also plotted in Figs. 9 and 10 for comparison purpose. It can be observed from Figs. 9 and 10 that PCA-based pattern recognition method can successfully extract temperature distribution along the FUT and the temperature distributions along the last section of the FUT put inside the oven at controlled temperatures are very stable. Since the temperature outside the oven is not strictly controlled, the temperature distributions along the section of the FUT placed at room temperature outside the oven vary a little over time as can be seen from Figs. 9 and 10.

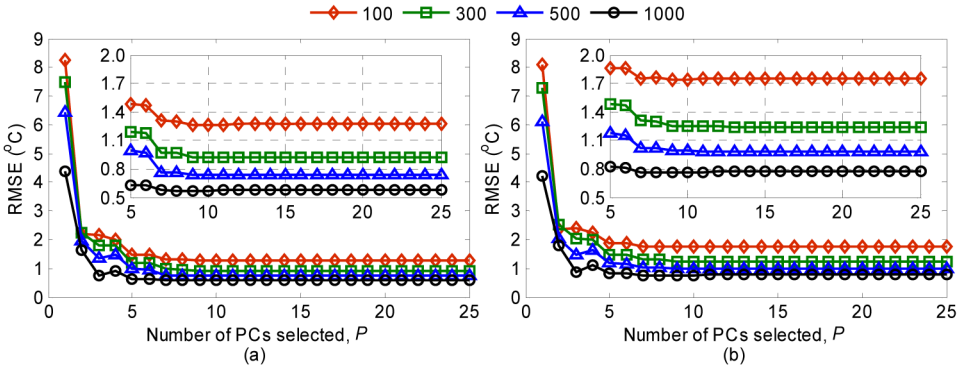

Fig. 8. RMSE as a function of number of PCs selected to synthesize the feature vectors of BGSs obtained using two different frequency steps of (a) 1 MHz and (b) 2 MHz.

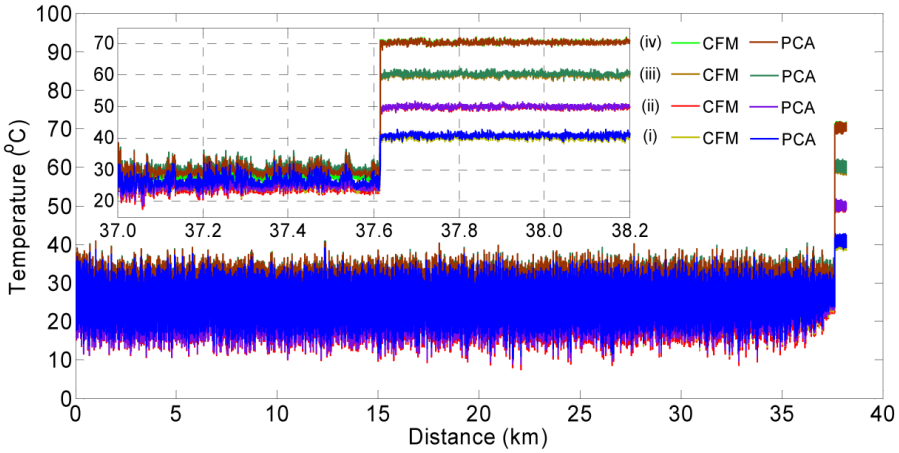

Fig. 9. Temperature distributions along 38.2 km FUT determined using PCA-based pattern recognition and CFM for the BGSs obtained using 1 MHz frequency step, 1000 trace averaging and the last ~600 m section of the FUT heated at (i) 40 °C, (ii) 50 °C, (iii) 60 °C and (iv) 70 °C; inset: temperature distributions for 1.2 km section from 37 km to 38.2 km.

Optics EXPRESS

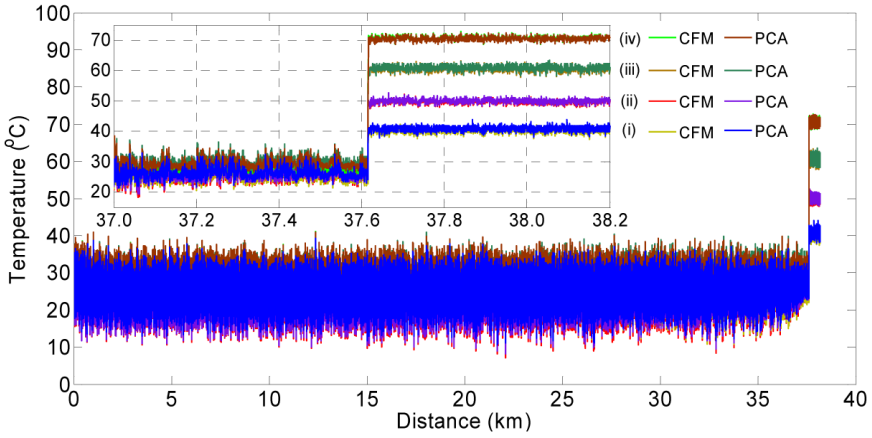

Fig. 10. Temperature distributions along 38.2 km FUT determined using PCA-based pattern recognition and CFM for the BGSs obtained using 2 MHz frequency step, 1000 trace averaging and the last ~600 m section of the FUT heated at (i) 40 °C, (ii) 50 °C, (iii) 60 °C and (iv) 70 °C; inset: temperature distributions for 1.2 km section from 37 km to 38.2 km..

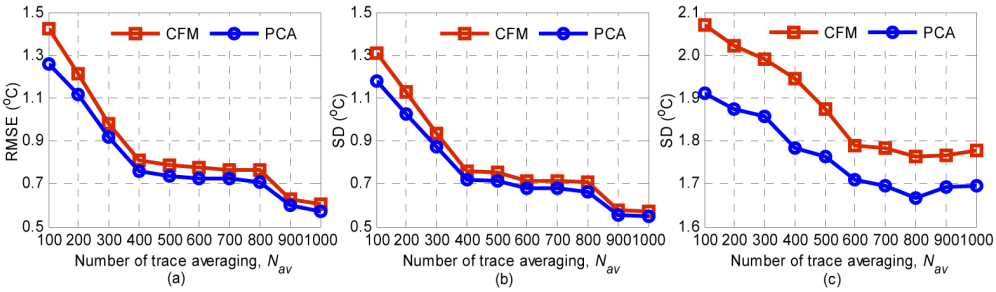

Fig. 11. (a) RMSE, (b) SD computed for the extracted temperatures along the last 500 m (i.e., 37.7 km to 38.2 km) section of the FUT and (c) SD computed for the extracted temperatures along the 500 m (i.e., 37 km to 37.5 km) section of the FUT. The measured BGSs are obtained using 1 MHz frequency step.

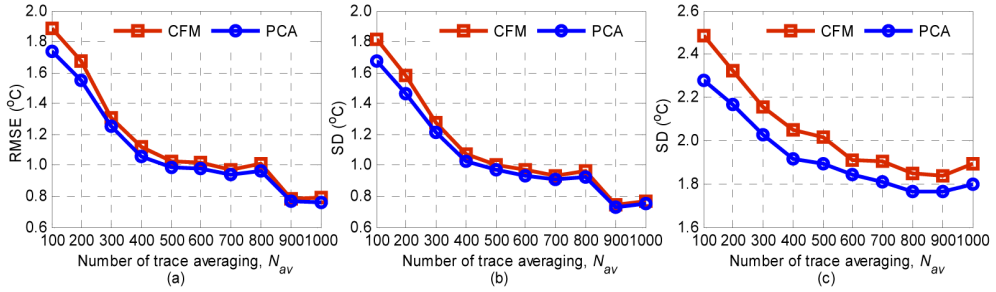

Fig. 12. (a) RMSE, (b) SD computed for the extracted temperatures along the last 500 m (i.e., 37.7 km to 38.2 km) section of the FUT and (c) SD computed for the extracted temperatures along the 500 m (i.e., 37 km to 37.5 km) section of the FUT. The measured BGSs are obtained using 2 MHz frequency step.

In order to compare the temperature extraction accuracies of PCA-based pattern recognition method and CFM, we compute RMSE as well as standard deviation (SD) of the extracted temperatures along the last 500 m (i.e., 37.7 km to 38.2 km) section of the FUT heated inside the oven. We also determine SD of the extracted temperatures along a 500 m (i.e., 37 km to 37.5 km) section near the end of FUT which is put outside the oven at room temperature for performance comparison. The reason for choosing this 500 m section near the

Optics EXPRESS

end of FUT is that the SNR for the measured BGSs along this section is almost similar to that along the last 500 m section of the FUT. The RMSE and SD values for the selected sections of the FUT are computed for each of the four temperatures under consideration, i.e., 40 °C, 50 °C, 60 °C and 70 °C. The determined RMSE and SD values corresponding to these four temperatures are then averaged to calculate the mean RMSE and SD for a particular trace averaging N_{av} and frequency step $\Delta\upsilon$. The experimental results for different N_{av} are shown in Fig. 11 for the BGSs obtained using $\Delta\upsilon$ = 1 MHz. It is evident from Fig. 11 that for each N_{av}, PCA-based pattern recognition provides better accuracy than CFM and the difference in accuracy increases gradually while using lower N_{av} (i.e., lower SNR) to obtain the measured BGSs. For example, the RMSE values at N_{av} = 1000 in Fig. 11(a) for PCA-based pattern recognition and CFM are 0.571 °C and 0.603 °C, respectively which are comparable to each other but for N_{av} = 100, the RMSE value of 1.261 °C for PCA-based pattern recognition is lower than that of 1.421 °C for CFM. Similarly, the RMSE values in Fig. 12(a) given by the two methods at $\Delta\upsilon$ = 2 MHz while adopting higher N_{av} are also comparable to each other but at lower trace averaging, such as N_{av} = 100, PCA-based pattern recognition performs better with an RMSE value of 1.737 °C than the CFM having an RMSE of 1.890 °C. Similarly the SD values computed for the extracted temperatures along the heated section of last 500 m of FUT as shown in Figs. 11(b) and 12(b) for $\Delta\upsilon$ = 1 and 2 MHz, respectively, also validate better estimation performance for PCA-based method. Finally, the SD values provided by the two techniques for the extracted temperatures along 500 m (i.e., 37 km to 37.5 km) section near the end of the FUT which is put at room temperature outside the oven also confirms the superiority of PCA-based pattern recognition method over CFM as clear from Figs. 11(c) and 12(c) for the measured BGSs obtained using $\Delta\upsilon$ = 1 and 2 MHz, respectively. It can be seen from Figs. 11 and 12 that the errors in temperature extraction using both techniques increase when lower N_{av} is adopted to obtain the measured BGSs. This is due to the fact that SNR is roughly proportional to the square root of N_{av} as observed in Fig. 7(b). Therefore, the use of lower N_{av} results in lower SNR which in turn leads to larger temperature estimation errors. In addition, it can also be noticed from Figs. 11 and 12 that the estimation errors for both methods are higher when BGSs are obtained using $\Delta\upsilon$ = 2 MHz as compared to the case when $\Delta\upsilon$ = 1 MHz, for each given value of N_{av}. This is because a larger frequency step (i.e., less data points on the BGSs) leads to less accurate temperature extraction with a factor proportional to the square root of frequency step $\Delta\upsilon$ [27]. It is worth mentioning that the SD values computed for the extracted temperatures along the 500 m (i.e., 37 km to 37.5 km) section near the end of the FUT which is compactly wound on to the fiber mandrel are higher than those for the last 500 m (i.e., 37.7 km to 38.2 km) section of the FUT which is wound carefully by hand with diameter relatively larger than that of the fiber mandrel. The reason for this is the presence of uneven strain due to compact packing of fiber on to the mandrel [11].

Optics EXPRESS

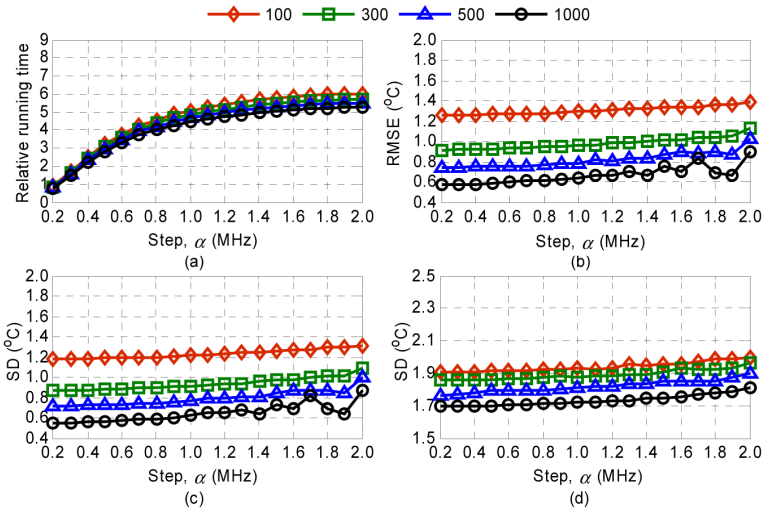

Fig. 13. (a) Relative running time, (b) RMSE, (c) SD for the extracted temperatures along the last 500 m (i.e., 37.7 km to 38.2 km) section and (d) SD for the extracted temperatures along the 37 km to 37.5 km section of the FUT using PCA-based pattern recognition with $P = 9$. The BGSs are obtained using $\Delta\upsilon = 1$ MHz frequency step.

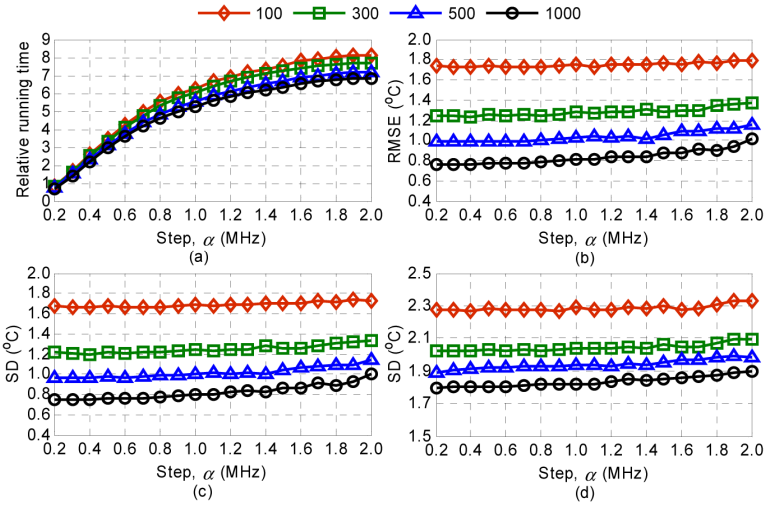

Fig. 14. (a) Relative running time, (b) RMSE, (c) SD for the extracted temperatures along the last 500 m (i.e., 37.7 km to 38.2 km) section and (d) SD for the extracted temperatures along the 37 km to 37.5 km section of the FUT using PCA-based pattern recognition with $P = 9$. The BGSs are obtained using $\Delta\upsilon = 2$ MHz frequency step.

Finally, we compare computational complexity and processing speed for extracting temperature distributions employing PCA-based pattern recognition and CFM. The computational complexity of CFM is on the order of $O(\beta N^2)$ where N is the length of measured BGS and β is the total number of iterations required by the method for the fitting process [9,10]. On the other hand, the computational complexity of Euclidean distance measurement for two sequences of length N is on the order of $O(N)$ [29]. Since we need to compute Euclidean distance between a feature vector (of length P) corresponding to a given measured BGS and all the M feature vectors (also of length P) in reference database R, the computational complexity after performing PCA becomes on the order of $O(MP)$. For a fixed BFS range F and linewidth range B, the number of feature vectors M depends on the step α

Optics EXPRESS

adopted to construct the reference database R. We can construct R using fewer BGSs by adopting a higher step α so as to improve the data processing speed. However, using such smaller database will adversely affect the accuracy of temperature extraction. Therefore, there is a tradeoff between processing speed and accuracy in temperature extraction while using PCA-based pattern recognition method. To analyze the processing speed and error performance of this method while using reference database R of different sizes M, we adopt different steps α within the F and B ranges for the construction of R. For this purpose, the step α is incremented gradually by an amount of 0.1 MHz in the range between 0.2 MHz and 2 MHz. For each reference database R constructed using a given step α, we employ PCA-based pattern recognition method to determine the running times while extracting temperature distributions from the measured BGSs obtained using a particular trace averaging N_{av} and frequency step $\Delta\upsilon$ for four different temperatures under consideration, i.e., 40 °C, 50 °C, 60 °C and 70 °C. The running times corresponding to these four temperatures are then averaged to obtain the running time for a particular α, N_{av} and $\Delta\upsilon$. Note that the running time calculated for each case also includes the time elapsed to obtain the feature vectors (of length $P = 9$) from the measured BGSs using PCA as well as the time required for the Euclidean distance measurement based matching process. In addition to running time, we also compute mean temperature errors for the measured BGSs obtained at four different temperatures in terms of RMSE and SD values of the extracted temperatures along the last 500 m (i.e., 37.7 km to 38.2 km) section of the FUT as well as SD values of the extracted temperatures for the 500 m (i.e., 37 km to 37.5 km) section near the end of the FUT for each α, N_{av} and $\Delta\upsilon$. On the other hand, we also determine mean running time while using CFM to process the measured BGSs obtained at four different temperatures for each N_{av} and $\Delta\upsilon$. For a particular N_{av} and $\Delta\upsilon$, we compute the relative running time (RRT) by dividing the running time of CFM with that of PCA-based pattern recognition computed for a specific step α. Figure 13 shows RRT and temperature errors as a function of step α for the measured BGSs acquired using $\Delta\upsilon = 1$ MHz frequency step and adopting four different trace averaging, i.e., 100, 300, 500 and 1000. It can be observed from Fig. 13(a) that temperature extraction using PCA-based pattern recognition method, with R constructed using step $\alpha = 0.2$ MHz, is slightly slower than the one obtained using CFM. On the other hand, employing $\alpha = 0.2$ MHz enables best estimation accuracy among the steps α adopted in our study. For higher step α, the processing speed of PCA-based method can be several times faster than the CFM. However, a larger α value results in a slight decrease in estimation accuracy as shown in Figs. 13(b)-13(d). We also obtain similar results for the measured BGSs attained using $\Delta\upsilon = 2$ MHz as shown in Fig. 14. For instance, RRT in Fig. 13(a) for the measured BGSs obtained using $\Delta\upsilon = 1$ MHz and $N_{av} = 500$ is ~4.2 for a step $\alpha = 0.8$ (which corresponds to a reference database R of size $M = 98 \times 38$), i.e., PCA-based pattern recognition method is ~4.2 times faster than CFM. For this case, temperature errors employing PCA-based method in Figs. 13(b)-13(d) are 0.766 °C, 0.747 °C and 1.798 °C which are slightly smaller than 0.789 °C, 0.750 °C and 1.874 °C obtained using CMF as can be seen from Figs. 11(a)-11(c), respectively. Similarly, for the measured BGSs acquired using $\Delta\upsilon = 2$ MHz and $N_{av} = 500$, the processing speed employing PCA-based method in Fig. 14(a) is ~4.8 times faster while the temperature errors of 1.003 °C, 0.986 °C and 1.929 °C in Figs. 14(b)-14(d) for $\alpha = 0.8$ are also slightly better than 1.023 °C, 1.001 °C and 2.015 °C attained using CFM in Figs. 12(a)-12(c), respectively. We can also obtain similar processing speeds and error performances by using other N_{av} adopted in this study. This implies that we can construct reference databases for PCA-based pattern recognition by adopting $\alpha = 0.8$ to obtain comparable error performance but the processing speed more than 4 time faster than that of CFM for the measured BGSs obtained using $\Delta\upsilon = 1$ and 2 MHz. The processing speed of PCA-based pattern recognition method can be further improved by constructing the reference databases with fewer relevant BGSs, thus making this technique desirable for use in applications which require faster processing speeds but can tolerate relatively large errors, e.g., to detect fire occurrence. It can be observed from Figs. 13(a) and 14(a) that the RRT

Optics EXPRESS

using $\Delta\upsilon$ = 2 MHz is higher than the case when $\Delta\upsilon$ = 1 MHz for each given N_{av}. This is because the length N of BGSs obtained using $\Delta\upsilon$ = 2 MHz is almost half compared to that acquired using $\Delta\upsilon$ = 1 MHz. Note that PCA-based pattern recognition method utilizes feature vectors of same length P for the measured BGSs obtained using $\Delta\upsilon$ = 1 and 2 MHz during the Euclidean distance measurement-based matching process. However, the method takes relatively shorter running time for extracting the feature vectors of BGSs obtained using $\Delta\upsilon$ = 2 MHz as compared to the case when $\Delta\upsilon$ = 1 MHz is used. On the other hand, the processing time for CFM does not reduce significantly as the method requires more iterations (i.e., higher β) to process the BGSs obtained using $\Delta\upsilon$ = 2 MHz as compared to that using $\Delta\upsilon$ = 1 MHz. In addition, for a particular frequency step $\Delta\upsilon$, the running time for PCA-based method is independent of trace averaging N_{av}. In contrast, the CFM requires more iterations and hence long running time to optimize the curve fitting parameters for the measured BGSs obtained using lower N_{av} (i.e., lower SNR). That is why, the RRT at a particular frequency step $\Delta\upsilon$ in Figs. 13(a) and 14(a) also increases when lower N_{av} is adopted to obtain the measured BGSs from the BOTDA experiment.

It is worth mentioning that the proposed PCA-based pattern recognition method is independent of the shape of BGS since it relies on the degree of similarity between BGSs obtained from BOTDA measurement and the ones available in the reference database. This means that for any measured BGS shape under consideration, we can simply construct a reference database corresponding to that particular BGS shape and then use it for temperature extraction by applying PCA and minimum distance measurement based matching process. The BGSs with specific shapes needed for the construction of reference database can be obtained either from BOTDA experiment or through appropriate theoretical modeling. This remarkable feature of proposed method opens up the possibility of accurate temperature extraction from BOTDA measured BGSs with special shapes, e.g., BGSs having multiple peaks resulting from non-local effects [3,19].

5. Conclusions

In this paper, we proposed a PCA and statistical distance measurement based pattern recognition technique to extract temperature distributions along the fiber from the local BGSs measured by using BOTDA system. The proposed technique can successfully extract temperature distributions without employing any curve fitting process and can provide better sensing accuracy along with faster data processing speed as compared to conventional CFM, especially in scenarios where the BGSs are measured by incorporating lower BOTDA trace averaging to reduce the acquisition time. The data processing speed of this technique can also be customized to make it suitable for applications with special timing requirements. Therefore, the proposed PCA-based technique can be an attractive alternative for extracting temperature distributions along the fibers in future BOTDA sensors.

Funding

National Science Foundation of China (61377093, 61435006); Hong Kong Government General Research Fund (PolyU5208/13E, PolyU152658/16E, PolyU152757/16E); The Hong Kong Polytechnic University (1-ZVFL).

Acknowledgments

Abul Kalam Azad acknowledges the support from the Research Grant Council, Hong Kong under the Hong Kong PhD Fellowship Scheme 2012/13.