Avoiding the side-effect of high-performance spectrometers with coded apertures

Mingbo Chi,[1,2] Peng Hao,[1] Yihui Wu,[1,*] Yongshun Liu,[1] and Ping Zhang[1]

[1]State Key Laboratory of Applied Optics, Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun 130033, China

[2]University of Chinese Academy of Sciences, Beijing 100039, China

*Corresponding author: yihuiwu@ciomp.ac.cn

Received 2 May 2013; revised 2 July 2013; accepted 5 August 2013; posted 5 August 2013 (Doc. ID 189774); published 3 September 2013

In this report, a high-resolution, high-signal-to-noise-coded aperture spectrometer is introduced that replaces the traditional single slit with two-dimensional array slits manufactured by microelectromechanic system technology. The encoding and decoding principle of this coded aperture spectrometer is described, as well as the instrument structure. We then discuss the side-effect, which is caused by sub-aperture manufacturing errors in size and position and the smear noise in the imaging CCD. The side-effect adversely affects the decoding wavelength accuracy of this spectrometer, so we present some effective ways to avoid this phenomenon and to increase the decoding wavelength accuracy of the spectrometer. In the end, we present our experimental results.

OCIS codes: (120.6200) Spectrometers and spectroscopic instrumentation; (300.6190) Spectrometers; (300.6320) Spectroscopy, high-resolution.

http://dx.doi.org/10.1364/AO.52.006467

1. Introduction

Optical spectroscopy is generally considered to be a mature technology. It is used to measure the properties of physical objects based on their interaction with light that has naturally evolved. As to a spectrometer, its performance is quantified by a few critical parameters: bandpass (the range of detectable wavelengths), spectral resolution (the narrowest spectral feature that can be independently detected), and the signal-to-noise ratio (SNR). However, virtually all spectrometer designs available today have to deal with trade-offs among these critical parameters. For example, the narrow slit which can enhance wavelength resolution in conventional grating spectrometers results in loss of light, which reduces the SNR of the whole system [1,2]. This trade-off limits the conventional spectrometer's use in the detection of weak signals that scarcely differ from the detector noise.

Several technologies and techniques have tried to balance the contradiction between the spectral resolution and SNR in a dispersive spectrometer. One approach is to utilize the optical-image slicer technology, which uses plane mirrors to divide a round image into narrow strips. A purely reflective based optical-pupil slicer has since been developed as a commercial product by tornado spectral systems, which is called the high throughput virtual slit (HTVS) [3,4]. The HTVS differs from image slicers in that it preserves the spatial information of the input aperture. Another approach to solve this problem is through coded aperture spectroscopy—the replacement of the input slit with a more complicated pattern of aperture coded in a proper matrix. The earliest coded aperture spectrometer was coded in a Hadamard matrix, and it realized the multimeasurement by moving the mask [5]. However, the movable mask would bring in notable mechanical

1559-128X/13/266467-07$15.00/0

errors and break down frequently, so it was not widely used. With the advent of a high-performance imaging CCD detector, a new class of coded aperture spectrometers, multimodal multiplex spectroscopy (MMS), was proposed, whose mask is completely static designed and is simpler, cheaper, and more robust than the dynamic designs of most HT spectrometers [6,7].

In a coded aperture spectrometer, the encoding mask with array slits and the image CCD function together to implement the encoding process, and they have a bad impaction on the SNR and decoding wavelength accuracy of the spectrometer. The influences of different encoding matrixes and the misalignment between the mask and CCD on the SNR of the spectrometer have been taken into consideration in previous research [8]. In this paper, we focus on the influence of sub-apertures' manufacturing errors and the smear noise of the imaging CCD detector, which has not been discussed before. The manufacturing errors of the aperture will break down the foundation of the encoding process; the apertures in the same column generate the same spectral distribution on the detector. The smear noise will generate a crosstalk among apertures in the same column. Their existence will decrease the decoding wavelength accuracy of the spectrometer. In the detection of discrete spectra, which suffers a lot from the manufacturing errors and the smear noise, there will be intense vacillations generated in the spectral intervals beside the spectral peaks so we call this phenomenon "side-effect." After the discussion, we present some effective ways to avoid the side-effect so that we can increase the decoding wavelength accuracy of the coded aperture spectrometer.

2. Main Structure of the Coded Aperture Spectrometer

In a traditional spectrometer shown in Fig. 1(a), the input signal is modulated and separated by the grating system into a series of images of the slit, which spread on the detecting plane by wavelength. We replace the single slit with a two-dimensional (2D)-coded aperture [Fig. 1(b)], which is encoded in a cyclic S-matrix [9]. As shown in Fig. 1(c), one row of the 2D coded aperture corresponds to pixels in a specific row of the detector. Meanwhile we call one independent column of the apertures a "channel" and mark it with η_n just because the apertures in the same column theoretically generate the same spectral distribution on the detecting plane, and the spectrum of different channels should be different.

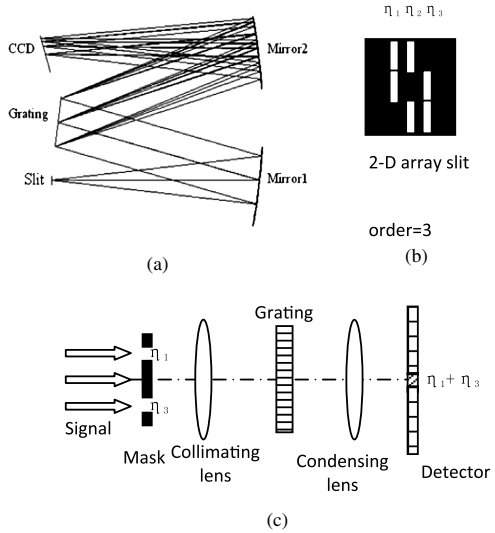

Fig. 1. (a) Schematic of a traditional grating spectrometer. (b) Schematic of 2D slit array (order = 3). (c) Schematic of 2D slit-array HT spectrometer.

As we can see from Fig. 1(c), which shows the encoding process of the second row of the coded aperture in Fig. 1(b), the spectra generated by channels 1 and 3 overlap on the detecting plane and the overlap from each row of the aperture can be seen as a multi-measurement of different channels selected by the coding (1 or 0) of the corresponding sub-aperture. The results of each multi-measurement can be represented by ψ_n and the encoding process can be expressed as follows:

$$\begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \end{bmatrix} = \begin{bmatrix} \psi_1 \\ \psi_2 \\ \psi_3 \end{bmatrix}. \tag{1}$$

If we denote $\eta = [\eta_1, \eta_2, \eta_3]^T$, $\psi = [\psi_1, \psi_2, \psi_3]^T$, and use W to represent the encoding matrix, Eq. (1) can be written as

$$W\eta = \psi. \tag{2}$$

Apply W^- to both sides of Eq. (2), and then we can get the spectral distribution of each channel η:

$$\eta = W^-\psi. \tag{3}$$

3. Side-Effect Caused by the Sub-aperture Manufacturing Errors

The foundation of the encoding process is the assumption that the sub-apertures in the same column generate the same spectral distribution. However, because of the manufacturing errors, the mask cannot be ideal and the system cannot be unbiased. If there are significant size (sub-slit height and sub-slit width) and location (column direction and row direction) errors between the sub-slits in the same column, the spectra generated will be quite different from each other (Fig. 2). If we implement the decoding algorithm using the ideal encoding matrix, the decoded spectrum will be distorted.

A. Sub-slit Width Error

For example, on a 3×3 cyclic S-matrix mask [Fig. 1(b)], for one pixel on the detector, sub-slits in each column correspond to a certain wavelength, λ_1, λ_2, λ_3, and the intensities at these wavelengths are denoted as $\phi_{\lambda 1}$, $\phi_{\lambda 2}$, $\phi_{\lambda 3}$, respectively. When the two sub-slits in column 1 have width errors, their

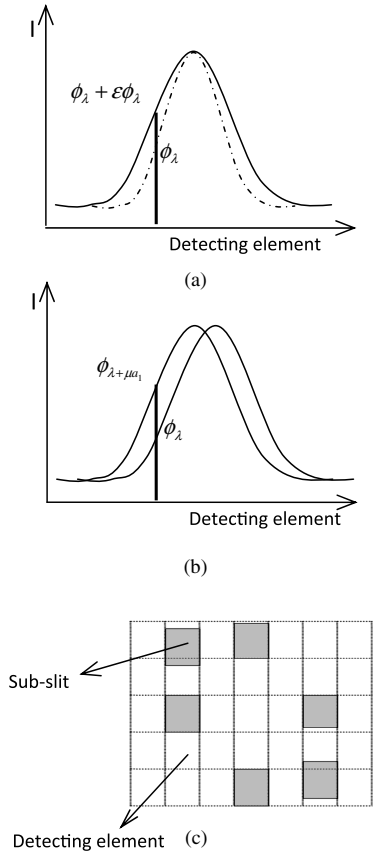

Fig. 2. Impaction of mask errors. (a) Sub-slit width error, (b) sub-slit position error in dispersive (column) direction, and (c) sub-slit height and position error in perpendicular dispersive (row) direction.

corresponding spectral resolution full width at half-maximum (FWHM) suffers tremendously [Fig. 2(a)]. Impacted by these errors, the encoding process can be represented by a modified matrix such as Eq. (4)

$$\begin{bmatrix} \eta_1' \\ \eta_2' \\ \eta_3' \end{bmatrix} = \begin{bmatrix} 1+\varepsilon_1 & 1 & 0 \\ 1+\varepsilon_2 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} \phi_{\lambda_1} \\ \phi_{\lambda_2} \\ \phi_{\lambda_3} \end{bmatrix} = \begin{bmatrix} \eta_1 + \varepsilon_1\phi_{\lambda_1} \\ \eta_2 + \varepsilon_2\phi_{\lambda_1} \\ \eta_3 \end{bmatrix} \tag{4}$$

with $\eta_1', \eta_2', \eta_3'$, the real measurement on the detector, η_1, η_2, η_3, the ideal measurement. If implementing the decoding algorithm with the ideal matrix, the detecting errors are

$$\begin{cases} \delta_{\phi_{\lambda_1}} = \frac{1}{2}(\varepsilon_1 + \varepsilon_2)\phi_{\lambda_1} \\ \delta_{\phi_{\lambda_2}} = \frac{1}{2}(\varepsilon_1 - \varepsilon_2)\phi_{\lambda_1} \\ \delta_{\phi_{\lambda_3}} = \frac{1}{2}(-\varepsilon_1 + \varepsilon_2)\phi_{\lambda_1} \end{cases}. \tag{5}$$

So, due to the sub-slit width errors, the spectral resolution (FWHM) of the spectral channel corresponding to this column (column 1) is changed ($\delta_{\phi_{\lambda_1}}$) and the spectra of other spectral channels (columns 2 and 3) will be affected by this spectral channel as $\delta_{\phi_{\lambda_2}}$ and $\delta_{\phi_{\lambda_3}}$ are both proportional to ϕ_{λ_1}. In fact, all the channels have manufacturing errors, and their spectra at different wavelengths will interfere with each other, which will lead to a decrease of the decoding wavelength accuracy.

B. Sub-slit Column Position Error

In another example, on a 3×3 cyclic S-matrix mask [Fig. 1(b)], if the two sub-slits in column 1 have position errors in dispersive direction, their spectra will generate wavelength shifts which are proportional to their position errors, respectively, [Fig. 2(b)]. So the position deviation of sub-slits in the same column will bring in a deviation of their spectra, and then the decoding wavelength accuracy of spectral channels corresponding to the same pixel element will be affected by these errors. The position errors in dispersive direction of the two sub-slits in column 1 are denoted as a_1 and a_2, respectively. The proportional coefficient between the wavelength shift and the position error is denoted as μ. Similar to the effect of the width errors, if we implement the decoding algorithm with the ideal matrix, the decoding errors are

$$\begin{cases} \delta_{\phi_{\lambda_1}} = \frac{1}{2}(\phi_{\lambda_1+\mu a_1} + \phi_{\lambda_1+\mu a_2}) \\ \delta_{\phi_{\lambda_2}} = \frac{1}{2}(\phi_{\lambda_1+\mu a_1} - \phi_{\lambda_1+\mu a_2}) \\ \delta_{\phi_{\lambda_3}} = \frac{1}{2}(-\phi_{\lambda_1+\mu a_1} + \phi_{\lambda_1+\mu a_2}) \end{cases}. \tag{6}$$

We can see that each spectral channel corresponding to this column of pixels is influenced by other spectral channels and then the wavelength accuracy is decreased.

C. Sub-slit Height and Row Position Error

In yet another example, on a 3×3 cyclic S-matrix mask, for a detecting element on the CCD, sub-slits in each column correspond to a certain wavelength. As the same as the sub-slit width error, the deviation of sub-slits height and location in perpendicular dispersive direction in column 1 will bring in a deviation of intensities of their spectra, and the encoding process can also be represented by Eq. (4). Thus, the decoded spectrum corresponding to some spectral channel (column 2 or 3) will be influenced by the spectral channel corresponding to column 1 [Eq. (5)].

D. Experimental Demonstration and Proposed Mitigation Technique

As we can see, the manufacturing accuracy of the 2D silicon array slits will influence the spectrometer's decoding wavelength accuracy and it is the most important in coded aperture spectrometer's implementation issues. We tested two kinds of 2D silicon array slits, which are made by microelectromechanic system technology. The first one's lithography mask is made by laser phototypesetting technology and the manufacture error (including size and position error) is approximately ± 3 μm. The other is made by e-beam direct writing technology and the manufacture error is approximately ± 0.3 μm. The manufacturing

of the lithography mask by the e-beam method is more expensive and takes a little longer time than the one using the laser phototypesetting technology. However, the lithography mask made by the e-beam method has the advantage of repeated use besides its high precision, while the one using laser phototypesetting technology can be used only a few times. In Fig. 3 we plot the detected results of an Hg lamp from the different 2D silicon array slits.

Having compared the two reconstructed spectra, we can see that the spectrum reconstruction has been distorted by the side-effect caused by large manufacture errors. Increasing the manufacturing precision of the coded apertures can effectively avoid this phenomenon and increase the decoding wavelength accuracy. Except for this, we can place a completely opaque row of 1 CCD pixel height between each row of apertures and make the aperture size a little higher than the CCD pixel size. This avoids the need for the sub-aperture height and row position manufacturing precision. It also decreases the need for the apertures' positioning accuracy toward the CCD pixel. As for the sub-apertures' width errors, we can measure them with the help of a high-resolution microscope. Then we can use the result of measurement to correct the encoding matrix as shown in Eq. (4). As for the sub-slit column position error, its effect on the wavelength accuracy changes with the input signal so it is hard to correct. Detection of discrete spectrum suffers a lot from the column position error as its intensity changes rapidly with wavelength. If the system is unbiased, the intensities in the intervals near the peak should be zero. So we can quantify the amount of decoded error through measuring the maximum spectral intensity in these intervals. We have acquired the spectrum of an Hg lamp for 50 times with a coded aperture spectrometer, whose sub-slit's position error is about ± 0.3 μm. Having taken the average of the 50 decoded spectra, we found that the maximum decoded error in the intervals near the peak at 546 nm is about 3% of the peak intensity at 546 nm. So we think the decoded error can be almost negligible when the position error is below 0.3 μm.

What's more, we have detected the spectrum of an Hg lamp with the coded aperture shown in Fig. 4(a)

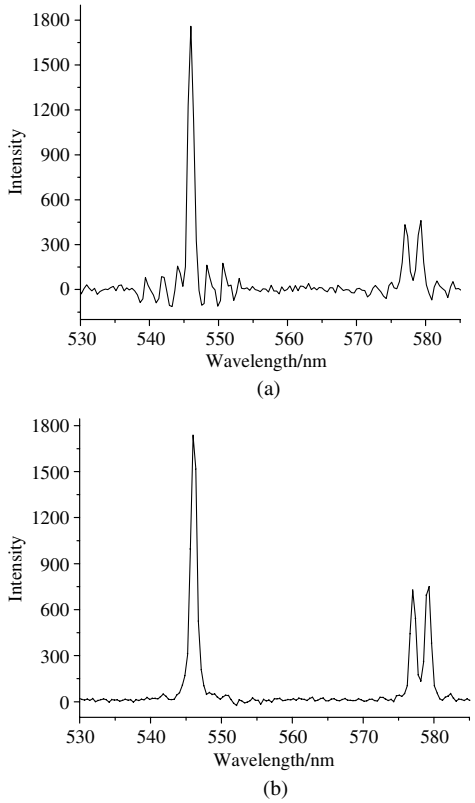

Fig. 3. Comparison between the reconstructed spectrums from (a) 2D silicon-slit-array used in lithographic exposure mask made by laser phototypesetting technology and (b) another 2D silicon-slit-array used in lithographic exposure mask made by e-beam direct writing technology.

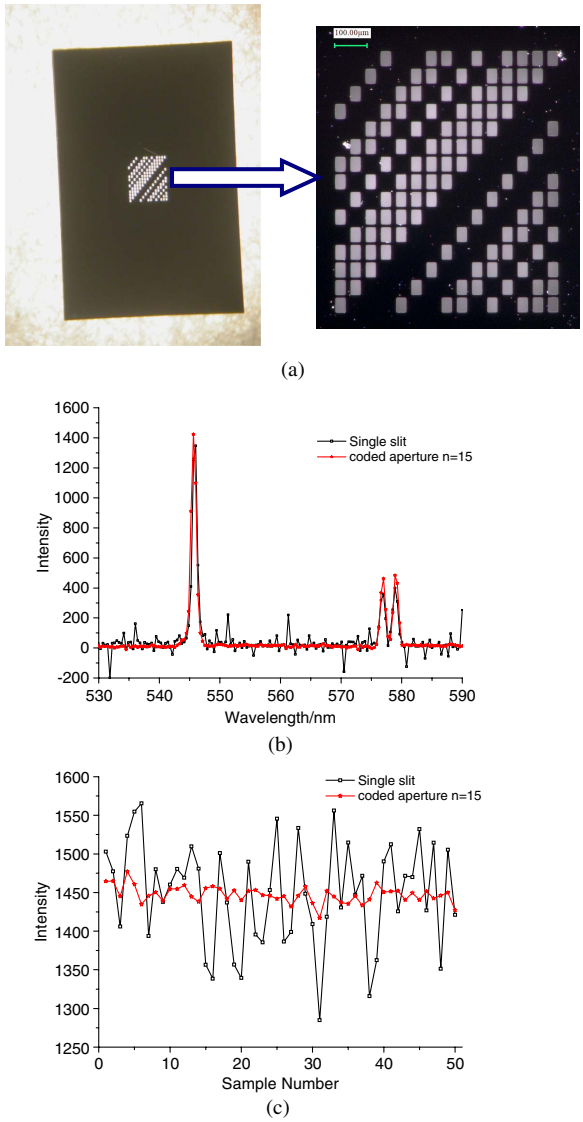

Fig. 4. (a) 2D-coded apertures (cyclic-S matrix code with order 15) and (b) Spectrum detected by the spectrometer with a conventional single slit and a $15 * 15$-coded aperture. (c) Peak intensities at 546 nm in the 50 detections of an Hg lamp.

and a single slit, respectively. The single slit is in the same size of the sub-slit in the coded aperture. The schematic of our system is shown in Fig. 1(a). The two detections have the same source illumination, focusing optics, detector, and integration time, etc. The only difference lies in the slit. We have acquired the coded aperture spectrum and the slit-only spectrum for 50 times, respectively. Figure 4(b) shows the contrast of the slit-only spectrum and the coded aperture spectrum. Figure 4(c) shows the fluctuations of the peak intensity at 546 nm in the two detections. We then compute the mean and standard deviation of the peak intensity at 546 nm to quantify the SNR of the two detections. We found that the SNR of the coded aperture spectrum is 6.8 times of the slit-only spectrum's SNR. The average FWHM of the coded aperture spectrum is 1.2 nm, while the one with the slit-only spectrum is 1.1 nm. Then we can understand that the spectrometer using a coded aperture can achieve almost the same resolution and higher SNR compared to the one using a single slit.

4. Side-Effect Caused by the Smear Noise of the Imaging CCD

In the encoding process, the modulation of each row of apertures can be considered as a multi-measurement of the input signal and every two of the multi-measurements should be independent. To acquire all the encoded data at the same time, we choose a 2D imaging CCD as the detector. However, smearing [10], which is an inevitable phenomenon that occurs during the charge transfer process of CCD imaging sensors, has generated a crosstalk among the multi-measurements. The smearing is especially significant when the intensity of some pixel is much higher than the pixels around it, so the detection of discrete spectrum suffers a lot from this phenomenon. As shown in Fig. 5(a), smear noise is white line generated when the light on some pixel is more intense than other pixels in the same column.

As the pixels in the same column share nearly the same intensity of the smear noise (represented by ε), which is proportional to the intensity of the input signal, the encoding process can be described as follows:

$$W\eta + \varepsilon = \phi. \tag{7}$$

If having ignored the smear noise in the decoding process, the reconstructed spectrum η' can be described as

$$\eta' = \eta + W^{-}\varepsilon. \tag{8}$$

In our spectrometer, we choose the S matrix as the encoding matrix W. If all the elements in each row of W^{-} are binned, the sum is $2/(n+1)$, where n is the order of W [11]. The decoded error in each channel is proportional to the intensity of the smear noise and shares the same intensity.

As for one column of CCD pixels, each spectral channel on the mask corresponds to a certain wavelength, and their intensity differs a great deal from each other in the detection of discrete spectra. So the side-effect caused by the smear noise is more serious on the intervals with low spectral intensities when they multiply with the intervals with high intensity, which are the main cause of the smear noise in this column. In other words, the spectral peak will interfere with the spectral intervals near it through the crosstalk caused by the smear noise in the encoding process, which will decrease the wavelength accuracy of the decoded spectrum. As shown in Fig. 5(b), there are significant side-effects in the intervals near the peak at 546 nm of the Hg-lamp spectrum, which is caused by the smear noise.

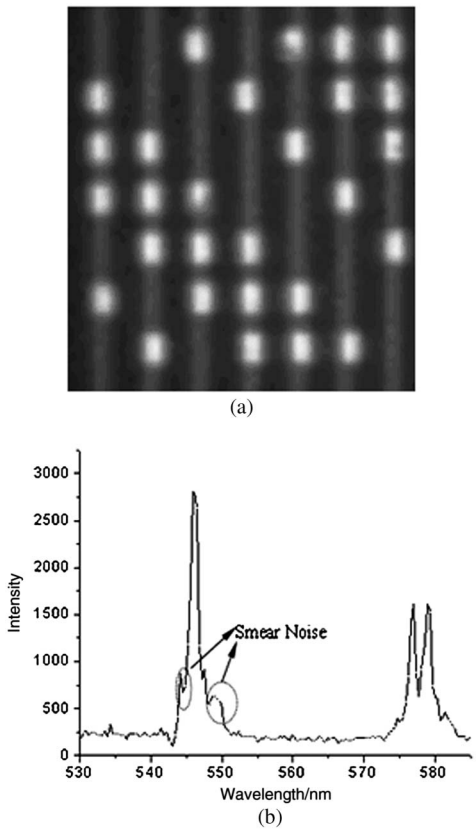

Fig. 5. (a) Smear noise on the CCD detector. (b) Side-effect caused by smear noise in the decoding spectrum.

Before implementing the decoding algorithm, we should remove the smear noise to avoid the side-effect and assure the wavelength accuracy of the decoded spectrum. We proposed a smear-removal algorithm based on wavelet transform that has a good localization characteristic property in time domain and frequency region [12]. The original signal in one column of the imaging CCD is composed of smear noise, encoded spectral signal, and the random noises. The smear noise in each pixel of this column has almost the same intensity, so its frequency is much lower than the frequencies of the encoded spectral signals and the random noises which change rapidly. In the higher-level decomposition of the original signal based on the Mallat algorithm, smear noise that has a lower frequency mainly contributes to

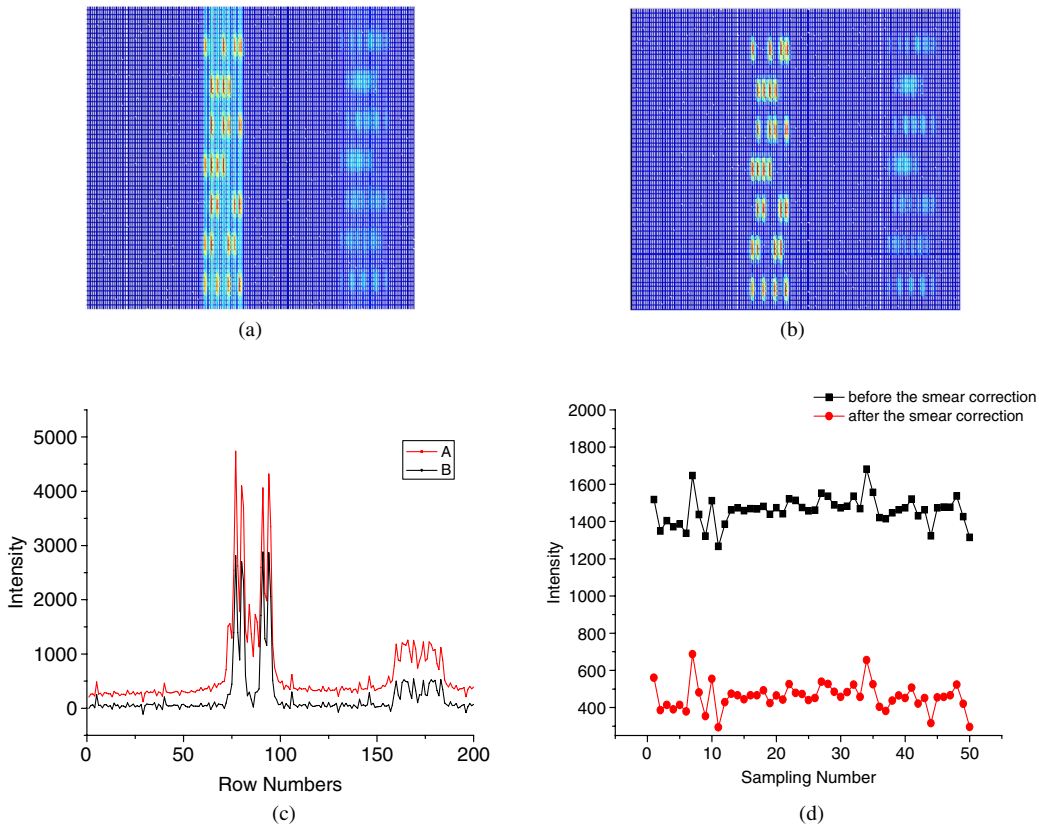

Fig. 6. (a) Signals on CCD before removal of smear noise. (b) Signals on CCD after implementing the removal algorithm. (c) Spectral distribution of the fifth row of the apertures before smear correction (curve A) and after smear correction (curve B). (d) The intensities of one pixel on the detector before and after the smear correction.

Table 1. Standard Deviation of the Pixel Intensities Before and After the Smear Correction

Pixel number	1	2	3	4	5	6	7	8	9	10
Before the smear correction	58.45	47.91	77.16	66.06	61.26	78.41	75.99	60.41	64.67	72.25
After the smear correction	56.10	49.70	73.64	59.36	60.12	75.60	71.85	54.77	61.11	69.60

the approximation coefficients, while the detail coefficients represent the random noise and the spectral signal. If we just use the detail coefficients to reconstruct the signal, the smear noise will be removed, and the coded spectral signal as well as the random noises will be reserved. So our algorithm will not increase the random noise level while removing the smear noise as shown in Figs. 6(a)–6(c).

We have acquired the spectrum of an Hg lamp for 50 times. Before and after the smear correction, we recorded the intensities of 10 pixels in the same column where there are intense smear noises. The intensities of one pixel in the ten before and after the correction are shown in Fig. 6(d). We compute the standard deviation of each pixel's intensities to quantify the random noise level. The experimental result is shown in Table 1, and we can find that the standard deviation of each pixel stays almost the same before and after the smear correction. It can be derived that our smear removal algorithm will not increase the random noise level while removing the smear noise.

5. Conclusion

Because of the manufacturing errors, the decoding algorithm cannot be implemented accurately, and the spectral channels multiplying on the same column of pixels will interfere with each other, which will decrease the wavelength accuracy. The smear noise in the image CCD will also decrease the wavelength accuracy through the crosstalk that is generated among the sub-apertures in the same column. To avoid the interference and increase the decoding wavelength accuracy, we should guarantee the manufacturing precision of the coded apertures and remove the smear noise before the decoding process. The removal algorithm we proposed will not increase the random noise level while removing the smear noise. The design of the opaque row between two rows of sub-apertures and the calibration of sub-aperture's width error will help increase the wavelength accuracy. In future research, we will try hard to deal with the sub-aperture column position errors

except for increasing the manufacturing precision as far as possible.

We thank the National Natural Science Foundation of China (Nos. 61102023 and 11034007), the development plan project of Jilin Province Science and Technology (20120329), and the project supported by the National High Technology Research and Development Program of China (Grant No. 2012AA040503) for financial support.

References

1. M. T. E. Golay, "Multi-slit spectrometry," J. Opt. Soc. Am. **39**, 437–444 (1949).
2. H. Suto and H. Takami, "Waveguide image slicer," Appl. Opt. **36**, 4582–4586 (1997).
3. J. T. Meade, B. B. Behr, and A. R. Hajian, "A new high-resolution, high-throughput spectrometer: first experience as applied to Raman spectroscopy," Proc. SPIE **8374**, 83740V (2012).
4. E. Beckert, K. G. Strassmeier, M. Woche, R. Eberhardt, A. Tunnermann, and M. Andersen, "Waveguide image-slicers for ultrahigh resolution spectroscopy," Proc. SPIE **7018**, 70182J (2008).
5. J. A. Decker and M. Harwit, "Experimental operation of a Hadamard spectrometer," Appl. Opt. **8**, 2552–2554 (1969).
6. M. E. Gehm, S. T. McCain, N. P. Pitsianis, D. J. Brady, P. Potuluri, and M. E. Sullivan, "Static 2D aperture coding for multi-modal multiplex spectroscopy," Appl. Opt. **45**, 2965–3183 (2006).
7. S. McCain, M. Gehm, Y. Wang, N. Pitsianis, and D. Brady, "Coded aperture Raman spectroscopy for quantitative measurements of ethanol in a tissue phantom," Appl. Spectrosc. **60**, 663–671 (2006).
8. A. A. Wagadarikar, M. E. Gehm, and D. J. Brady, "Performance comparison of aperture codes for multimodal, multiplex spectroscopy," Appl. Opt. **46**, 4932–4942 (2007).
9. M. Harwit and N. Sloane, *Hadamard Transform Optics* (Academic, 1979), pp. 57–61.
10. Y. S. Han, E. Choi, and M. G. Kang, "Smear removal algorithm using the optical black region for CCD imaging sensors," IEEE Trans. Consum. Electron. **55**, 2287–2293 (2009).
11. S. P. Love, "Programmable matched filter and Hadamard transform hyperspectral imagers based on micro-mirror arrays," Proc. SPIE **7210**, 721007 (2009).
12. I. Daubechies, "The wavelet transform, time-frequency localization, and signal analysis," IEEE Trans. Inf. Theory **36**, 961–1005 (1990).