OUKA

**The University of Osaka
Institutional Knowledge Archive**

Title	Matching information terminal based on a spatially coded moiré technique
Author(s)	Nitta, Kouichi; Togo, Hiroyuki; Tanida, Jun
Citation	Optical Engineering. 2001, 40(11), p. 2386-2391
Version Type	VoR
URL	https://hdl.handle.net/11094/3248
rights	Copyright 2001 Society of Photo-Optical Instrumentation Engineers. One print or electronic copy may be made for personal use only. Systematic reproduction and distribution, duplication of any material in this paper for a fee or for commercial purposes, or modification of the content of the paper are prohibited.
Note	

Matching information terminal based on a spatially coded moiré technique

Kouichi Nitta
Hiroyuki Togo
Jun Tanida, MEMBER SPIE
Osaka University
Department of Material and Life Sciences
Graduate School of Engineering
2-1 Yamodaoka Suita
Osaka 565-0871, Japan
E-mail: nitta@mls.eng.osaka-u.ac.jp

Abstract. We report on the construction of a matching information terminal based on a spatially coded moiré technique, to be used as visualization equipment for genome analysis. The terminal provides the result of a string matching operation, which is achieved by overlapping of spatially coded patterns with a small intersection angle to generate moiré fringes. The fringes reveal the matching between two string sequences coded as spatial patterns. Due to its flexibility in image display, the matching information terminal is able to deal with long strings such as a gene sequence. The constructed terminal achieves string matching for 256 bases of DNA sequences and displays about 25,000 point-to-point matching results at a time. © *2001 Society of Photo-Optical Instrumentation Engineers.* [DOI: 10.1117/1.1408313]

Subject terms: moiré technique; string matching; photonic bioinformatics.

1 Introduction

Optical computing techniques utilizing the inherent parallelism and broad bandwidth of light offer effective means for high-performance information processing that cannot be achieved by electronics alone.[1] However, we should also recognize that very few methods based on these techniques are useful at present for practical applications. There are several reasons for this, but a fundamental one is that we have not been able to find suitable applications for optical computing. If sophisticated applications can be demonstrated, the situation can be expected to change.

As is well known, bioinformatics is becoming an important research field as genome sequencing is performed on more and more species.[2] Bioinformatics is a broad science that treats genome sequences and protein structure and function. Due to progress in genome sequencing, information processing on the genome can greatly reduce the required amount of biological experimentation. As a result, we can accelerate the investigation of molecular biology and of life itself. The amount of genome information is already huge, and the relations within it are complex. Therefore, a powerful information processor is crucial for bioinformatics. In addition, intuitive judgment or inspiration by a human operator is also required for the task of genome analysis. In this situation, information visualization is important as a human-machine interface.

Optical computing systems are considered as promising for the above requirement. Their parallelism and broad bandwidth should be useful for processing large amounts of information, and the visibility of light provides effective means of information visualization. Therefore, optical computing for bioinformatics—we might call it *photonic bioinformatics*—is a good candidate for the application that demonstrates the capability of the optical computing. Some attempts have already been presented along these lines.[3,4]

As an instance of the application of photonic bioinformatics, the authors have presented a method for string data matching based on a spatially coded moiré technique.[5] String matching is one of the essential operations in bioinformatics.[6] For example, string matching between DNA sequences provides useful information for prediction of gene function and classification of species. A feature of the moiré matching method is its capability of parallel extraction of the locations where deletion, insertion, and reiterated sequences exist. Visualization of the optical computing technique is also utilized effectively. Sequences of nucleotides and amino acids can be treated by the method.[7]

This paper presents a prototype matching information terminal based on a spatially coded moiré technique. The terminal is a visualization device that indicates the correspondence between two different string sequences with the help of personal computer. The constructed terminal executes the matching operation by physical overlapping of two coded images, which are displayed on different devices with a small intersection angle. The terminal has excellent features such as flexibility to test various types of spatial code patterns, dynamic change of the target strings, and easy access to genome databases.

In Sec. 2, the principle of string matching using the spatially coded moiré technique is explained. In Sec. 3, some characteristic patterns are described for extracting information on deletion, insertion, and reiterated sequences. In Sec. 4, the matching information terminal is described with respect to its features and structures, and the specification of a prototype using a liquid-crystal display is presented. In Sec. 5, an experimental result obtained with the matching information terminal is presented. As a variation of spatial coding, an inverse coding rule is demonstrated on the information terminal, which improves the visibility of the moiré fringes. In Sec. 6, the performance of the matching information terminal is evaluated in terms of matching numbers and detectable length of insertions or deletions. To
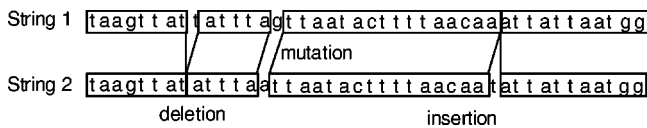
Fig. 1 String matching of two DNA sequences.

clarify the features of the proposed method, comparison between the moiré matching method and the dot-matrix methods is made.

2 String Matching Based on a Moiré Technique

String matching is an operation to evaluate the similarity of one one-dimensional sequence of data to another sequence. Although this operation is a fundamental one in the context of genome analysis, where huge numbers of strings must be processed, no effective method is yet available. Figure 1 shows an example of string matching between two DNA sequences. Due to the occurrence of exchange, deletion, and insertion of bases, the matching operation must be executed with consideration of the possibility of lateral shift of the elements in the strings.

In a mathematical form, this operation is described as follows:

$$O[i,j] = S_1[i] \otimes S_2[i+j] \quad (i=1,\ldots N; \; j=-K,\ldots,K),$$

(1)

where $S_1[i]$ and $S_2[i]$ are the i'th elements of two target strings S_1 and S_2, and $\otimes$ is a matching operator. The function $O[i,j]$ is the result of the matching. The value of $O[i,j]$ is "match" when the i'th element of S_1 and the $i+j$'th element of S_2 are identical, or "mismatch" when these are different. Finally, N is the length of the string S_1, and K is the maximum amount of lateral shift allowed in detecting deletion or insertion of elements.

String matching for DNA sequences provides various information. Deletion, insertion, and exchange between the sequences extracted by the string matching provide various information on life. For example, the function of a sequence can be predicted by comparison with a sequence whose function is known. A characteristic subsequence can be extracted by comparison of the sequences appearing in similar species.

To achieve the above operation effectively, an optical method based on a spatially coded moiré technique has been presented.[5] Figure 2 shows the processing procedure for data matching between strings S_1 and S_2. First, each element in the target strings is converted into one of the code patterns. Figure 3(a) shows the code patterns for the DNA bases adenine (A), guanine (G), cytosine (C), and thymine (T). Figure 3(b) shows the inverse coding used to improve visibility. The rule for this coding is explained in detail in Sec. 5. The coded images are overlapped with a small intersection angle to generate moiré fringes. On the moiré fringes a bright segment indicates a position where the elements in the two strings are coincident. The output moiré pattern shows the result of matching operations for different relative shifts between the strings on a two-
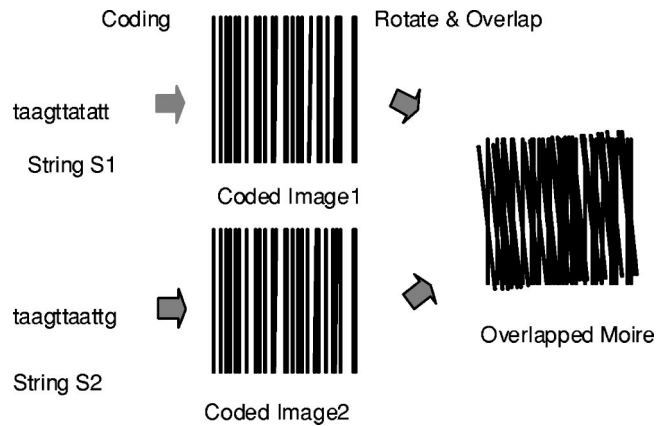

Fig. 2 Processing procedure for string matching using spatial coding and moiré technique.

dimensional plane in parallel. As shown in Eq. (1), the output results $O[i,j]$ for multiple combinations of i and j are obtained at the same time.

Figure 4 shows the pattern observed when the coded images, printed on transparent sheets, are overlapped. As test sequences, about 300 bases extracted from the 16S rRNA sequences of KOD1 (S_1) and *Pyrococcus furiosus* (S_2), kinds of archaebacteria, are used. As a reference, Fig. 5 shows contents of the sequences, as aligned by CLUSTALW.[8] In Fig. 5 an asterisk (*) under elements of the two sequences expresses that the elements are identical. A hyphen (-) expresses deletion or insertion between the two sequences. Figure 4(b) shows the output pattern with a larger insertion angle than that of Fig. 4(a). The larger the intersection angle is, the more data are processed at a time, but with less sensitivity. Figure 4(c) depicts an observed pattern with lateral shift of one coded image. In this case, the positional correspondence between the sequences can be changed and the bright segments are shifted downward.

The performance of the moiré matching method is evaluated by the number P of matching points that can be found at a time. It is approximated by
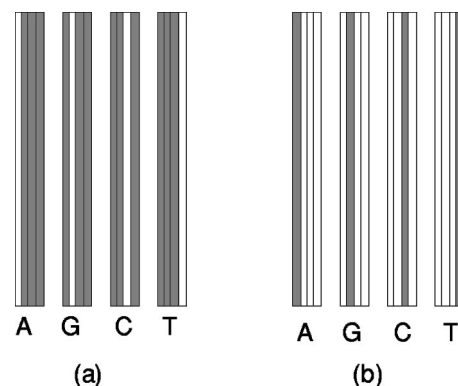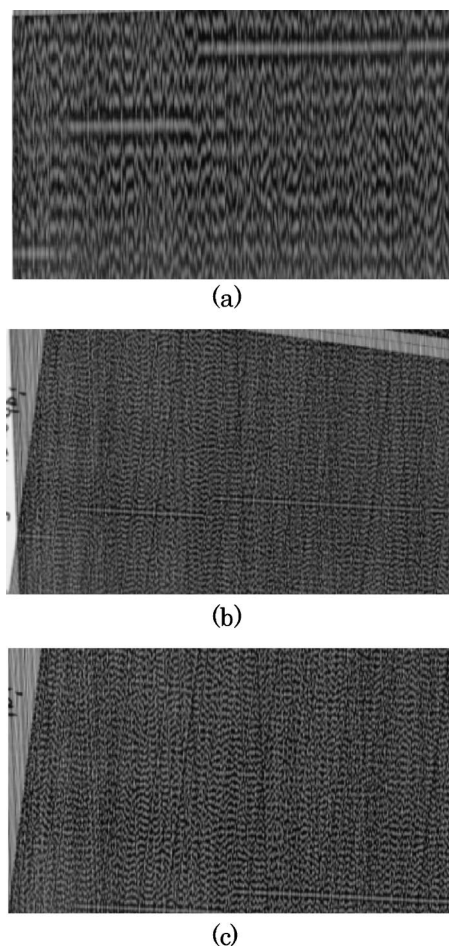
$$P = LM \sin \alpha,$$

(2)

Fig. 3 Code patterns for DNA bases: (a) original coding, (b) inverse coding.

(a)

(b)

(c)

Fig. 4 Photograph of matching results with printer outputs: (a) small insertion angle, (b) large insertion angle, and (c) lateral shift of one coded image.

where M is the number of the elements on the processing plane, L is the height of the code pattern divided by the code width d, and α is the intersection angle between two coded images.[7] The condition for generation of moiré fringes is $\alpha \ll \pi/3$.

Another performance measure is the detectable length of insertion and deletion. The length Q is approximated by the following equation[7]:

$$Q = 2L \tan \frac{\alpha}{2}. \tag{3}$$

3 Interpretation of Moiré Fringes

In the moiré matching method, deletion and insertion between two sequences and reiterated sequences are extracted as characteristic patterns. Deletion and insertion provide important information for examining the process of evolution, because crossover or mutation in the DNA sequences is considered as a driving force of evolution.[9] Reiterated sequences are useful for identifying individuals.

Figure 6 is a feature-extracted image of Fig. 4(a) to explain how to interpret the output pattern. In the figure, a line segment corresponds to a bright fringe in the original image.

```
KOD1        ATTCCGGTTGATCCTGCCGGAGGCCACTGCTATGGGGGTCCGACTAAGCCATGCGAGTCA
P. furiosus ATTCCGGTTGATCCTGCCGGAGGCCACTGCTATGGGGGTCCGACTAAGCCATGCGAGTCA
            ************************************************************

KOD1        TGGGG-CGCGCTCTGCGCG----CACCGGCGGACGGCTCAGTAACACGTCGGTAACCTAC
P. furiosus AGGGGGCGTCCCTTCTGGGACGCCACCGGCGGACGGCTCAGTAACACGTCGGTAACCTAC
            **** ** * * **   ********************************

KOD1        CCTCGGGAGGGGGATAACCCCGGGAAACTGGGGCTAATCCCCCATAGGCCTGAGGTACTG
P. furiosus CCTCGGGAGGGGGATAACCCCGGGAAACTGGGGCTAATCCCCCATAGGCCTGGGGTACTG
            *************************************************** ******

KOD1        GAAGGTCCTCAGGCCGAAAGGGGCATCT---GCCCGCCCGAGGATGGGCCGGCGGCCGAT
P. furiosus GAAGGTCCCCAGGCCGAAAGGGAGCCGTAAGGCCCGCCCGAGGATGCGGGCGGCGGCCGAT
            ******* ************       *  *************** * ***********

KOD1        TAGGTAGTTGGTGGGGTAACGGCCCACCAAGCCGAAGATCGGTACGGGCCATGAGAGTGG
P. furiosus TAGGTAGTTGGTGGGGTAACGGCCCACCAAGCCGAAGATCGGTACGGGCCGTGAGAGCGG
            **************************************************** ****** **

KOD1        GAGCCCGGAGATGGACACTGAGACACGGGTCCAGGCCCTACGGGGCGCAGCAGGCGCGAA
P. furiosus GAGCCCGGAGATGGACACTGAGACACGGGTCCAGGCCCTACGGGGCGCAGCAGGCGCGAA
            ************************************************************

KOD1        ACCTCCGCAATGCGGGCAACCGCGACGGGGGGACCCCCAGTGCCGTGGCATAGCCACGGC
P. furiosus ACCTCCGCAATGCGGGAACCGCGACGGGGGGACCCCCAGTGCCGTGCCTCTGGCACGGC
            ****************  ****************************** *  * ******

KOD1        TTTTCCGGAGTGTAAAAAGCTCCGGGAATAAGGGCTGGGCAAGGCCGGTGGCAGCCGCCG
P. furiosus TTTTCCGGAGTGTAAAAAGCTCCGGGAATAAGGGCTGGGCAAGGCCGGTGGCAGCCGCCG
            ************************************************************
```

Fig. 5 String alignment between DOD1 and *Pyrococcus furiosus* using CLUSTALW.

One of the most characteristic patterns is long segments running from the upper right to the lower left corner. These segments indicates close correspondence between the sequences, with two large gaps. In this case, vertical shift of a segment shows insertion or deletion of elements, and the amount of the shift is proportional to the number of the inserted or deleted elements. The horizontal length of the segment indicates the length of the matching sequences.

The actual number of the elements can be figured out from the stripe patterns in the vertical direction. Such a pattern is generated where the same kind of elements are repeated, e.g., GGGG. Thus the vertical spacing between the short segments can be a measure of one-element insertion or deletion. Thus, in Fig. 6, we find that five and three elements are inserted in one of the sequences at two positions, which is equivalent to a deletion in the other sequence.

The short segments themselves also provide information on reiterated sequence. Although one-element reiteration is not useful except for setting the scale, several-element reiterations (known as a microsatellites) can be utilized for polymorphism analysis.[9] In this case, the length of the reiteration is obtained from the vertical spacing, and the reiteration number is given by the number of segments arranged vertically. Unfortunately, the observed sequences in this example do not contain microsatellites.

Along with the long segments, there are several discontinuities in the horizontal direction. Such a discontinuity
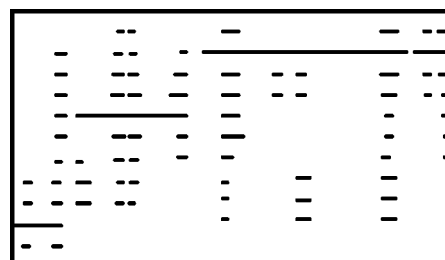
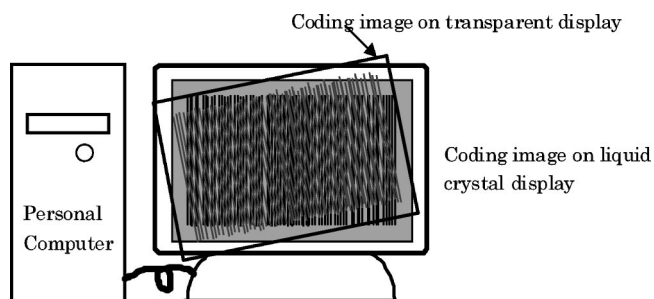**Fig. 6** Characteristic patterns appearing in moiré fringes.

Fig. 7 Schematic diagram of moiré matching terminal.

(a)

(b)

Fig. 8 Matching result provided by the terminal: (a) the observed pattern and (b) the corresponding bases.

indicates complete replacement of elements. In Fig. 6, one clear instance appears at the right in the longest segment near the top, and another one is found near the left end of the same segment.

As shown in the above interpretation process, the output data of the moiré matching method appear in a continuous analog format, so that the result is somewhat fuzzy and seems to lack precision. However, we regard this as a result of data reduction and find that the appropriate information has been extracted. These features will be discussed in Sec. 6.

4 Matching Information Terminal

To execute the moiré matching operation effectively, a matching information terminal was constructed. Figure 7 shows a schematic diagram of the terminal.

The terminal achieves string matching between two DNA sequences. On it, two coded images are displayed on different devices and physically overlapped with a small intersection angle. It may seem that both coded images could be displayed on the computer terminal, but the finite resolution of the display prevents us from realizing such a method. If a coded image is drawn at a small angle, the pattern itself generates moiré patterns with the pixel grid of the display, which interfere the significant moiré fringes for the matching. Therefore, we conclude that an optical method with two devices is a reasonable implementation.

With the help of the personal computer, many features can be added to the matching operation. First of all, we can test various types of spatially coded patterns to improve the sensitivity of the operation. Since the terminal operator observes the displayed moiré fringes and extracts useful information from them, the pattern must be informative and suitable for human inspection. For such study, the matching information terminal is a powerful tool.

As a second feature provided by the computer support, dynamic change of the target strings is mentioned. A moiré matching process using transparent sheets made by a high-resolution printer enables us to observe a broad area at one time. However, an A4 (30 cm×21 cm) sheet treats only about 3500 bases using a printer of 1200-dot/in. resolution. A genome sequence has millions of bases or more, and contains hundreds of thousands of deletions and insertions. Therefore, dynamic change of the coded image is necessary for a genome sequence matching.

The third feature is easy access to genome databases. Up-to-date results in life science are made available as public genome databases, so that access to the database is in-
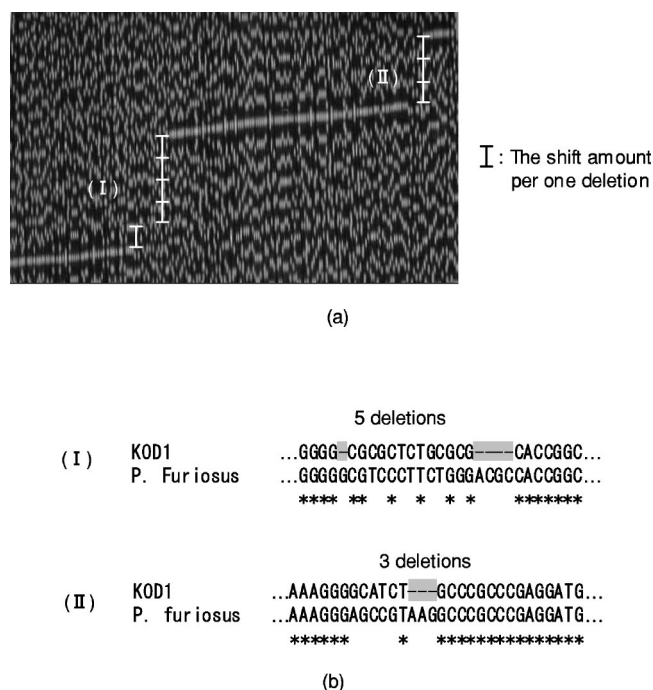
dispensable as a tool for genome analysis. The communication capability of the personal computer is effectively utilized for that requirement.

As a prototype of the matching information terminal, a liquid-crystal display and a transparent sheet are employed to generate moiré fringes. One coded image printed on a transparent sheet is set in front of the liquid-crystal display. The transparent sheet is sandwiched between two acrylic plates to keep it flat. In this case, we can dynamically manipulate only one coded image, and the operational flexibility is reduced to some extent. However, this implementation is sufficient for basic research on the method and is especially attractive by the reason of its easy fabrication.

5 Experimental Results from the Matching Information Terminal

Figure 8 shows the matching results obtained with the prototype terminal for the sequences in Fig. 5. Figure 8(a) is the observed pattern, on which the annotation for the insertion is superimposed. Comparing Fig. 4(a) with Fig. 8(a), it is found that similar patterns are obtained except for a segment tilt. The tilt is caused by slight mismatch of the size of the coded images on the liquid-crystal display and the transparent sheet.

To demonstrate flexibility in the coding of the matching information terminal, an inverse coding method is employed. Figure 3(b) shows the inverse coding rule, which consists in interchange of the opaque and transparent parts of the original coding in Fig. 3(a). With this coding rule, we can improve the visibility of the moiré fringes. Figure 9 shows the string matching results obtained by application of the inverse coding for (a) one and (b) both of the coded images. For the two cases, the inverse coding rule is adopted for the coded image on the transparent sheet, and

(a)

(b)
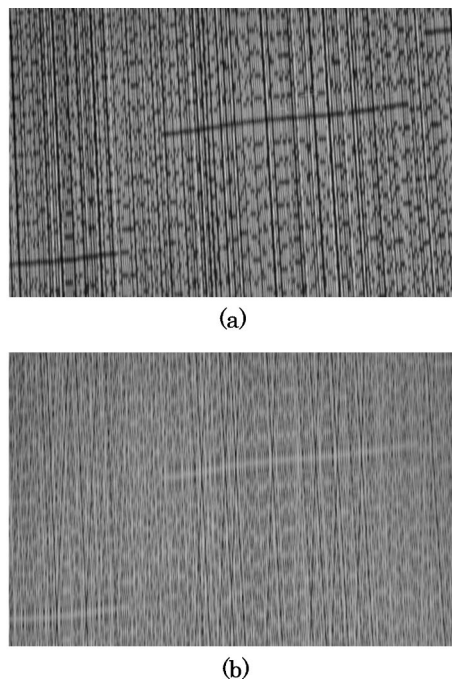
Fig. 9 Photograph of matching results using the inverse coding method: (a) single inversion and (b) two inversions.

Table 1 Feature comparison between the moiré matching and the dot matrix method.

Feature	Moiré matching method	Dot-matrix method
Output data format	Analog Continuous	Digital Discrete
Senstivity control	Intersection angle Coding pattern Moving	Marker coloring Data grouping
Data reduction	Averaging effect	Filtering
No. of data	$8N$	N^2

the coding rule for the display is changed. Comparison between the results and those in Fig. 8(a) indicates that the visibility of the fringes is improved for the single-inversion case. With the single coding inversion, a single mutation can be detected from the moiré fringes.

6 Discussion

For the prototype, the pixel pitch of the liquid-crystal display is 0.297 mm, and the pixel number is 1024×768. Thus M and L in Eq. (2) are 256 and 192, respectively, because a code pattern requires 4 pixels in the horizontal direction. Consequently, from Eq. (2), the prototype can display about 25,000 matching results at a time if the intersection between two coded images is 30 deg. From Eq. (3), Q is about 100 for $\alpha = 30$ deg. This means that we can detect 100-base lengths of insertion or deletion.

Comparison between the moiré matching method and the dot-matrix method is made to clarify the features of the proposed method. The dot-matrix method is a simple but very powerful technique that is widely used.[9] The method seems to execute an equivalent operation to the moiré method. Although a variety of modifications are possible on the dot matrix method, the basic concept remains the same. Thus, only one elementary method is considered in the comparison.

In the dot-matrix method, two sequences to be compared are assigned to the rows and the columns of a matrix. The combinations between the elements in the sequences correspond to the cells in the matrix. If elements of a combination are identical, a marker is plotted at the cell. After applying the same procedure, we can obtain the matching information between the sequences as a pattern composed of the markers. A diagonal line segment indicates a run of the matching elements, and a horizontal or vertical shift of

the segment shows the location of deletion or insertion of the elements. Marker coloring and element grouping are typical modifications of the method, which extend its applicability.

Table 1 summarizes the feature comparison between the moiré matching method and the dot-matrix method. Four characteristic points are compared.

The first point is the format of the output data. The moiré matching method generates continuos analog data, whereas the dot-matrix generates discrete digital ones. Although the latter seems to be easy to handle, the former contains rich information. From the point of the view of visualization, the output of the moiré matching method is more informative and more likely to induce human inspiration than the dot matrix.

The second point is the sensitivity control method. In the moiré matching method, the matching sensitivity is controlled by the intersection angle between the coded images and changing of the coding patterns. By moving one coded image, we can increase the sensitivity. In the dot-matrix method, changing the marker coloring and grouping a run of the elements are used for that purpose. It is difficult to judge between them in general, but the way in the moiré matching method is more intuitive and easier to handle.

The third point is the kind of data reduction that is required to extract useful information from the flood of data expected in the genome analysis. Because of its analog character, the averaging effect of the moiré phenomenon is effectively utilized in the moiré matching method. On the other hand, the dot-matrix method indicates the matching points unconditionally, so that desired matching information tends to be buried in the noise, especially for the DNA sequences. Usually, a filtering process, in which isolated markers and short strips are eliminated, is introduced for the data reduction. However, such an artificial process may lose important information. On this point, the moiré matching method is preferable. In it, information on the individual sequence elements is preserved in the coded patterns and can be reflected in the output with a simple manipulation, viz., by changing the intersection angle.

The final point is the total number of data to be handled in the method. The moiré matching method treats the data of two sequences of total length $2N$, where N is the length of one sequence. Each element of the sequence is converted into four data points, so that the total number of data becomes $8N$. On contrary, the dot-matrix method needs to

handle all the combinations of the sequences, N^2. Since the genome information is huge and long sequences must be processed, economy in the data handling becomes important.

In the previous paper,[5] we commented that the proposed method is insensitive to exchange of the bases, which seems a disadvantage. However, we find that sensitivity to exchange can be achieved by changing the coding method. As a result, we can extend the application field of the proposed method. With regard to the relation between the coding rule and the sensitivity, studies on moiré interferograms[10] are useful.

We recognize that the moiré matching method is not a replacement technique for the conventional methods such as the dot-matrix method. Instead, the proposed method is a useful complement for them. We hope the analog nature of the output information will open up a new methodology in genome analysis.

7 Conclusion

In this paper, a prototype of the matching information terminal based on a spatially coded moiré technique has been presented. The prototype executes string matching by overlapping the coded image drawn on a liquid-crystal display and that printed on a transparent sheet. The features of the prototype are flexibility in coding method, dynamic change of the test target, and accessibility of genome databases. The prototype terminal achieves string matching between 256 bases of DNA sequences and displays about 25,000 matching results at a time. The usefulness of the proposed method has been shown by comparison with dot-matrix methods.

As a future refinement, we should tackle the construction of a terminal capable of changing both coded images for long sequences. For the purpose, a transparent display with high resolution is required. We expect that such a matching information terminal will help us to explore further the capability of the spatially coded moiré matching technique in genome information processing.

References

1. J. Tanida and Y. Ichioka, "Digital optical computing," in *Progress in Optics XL*, pp. 77–114, North-Holland, Amsterdam (2000).
2. F. H. Arnold, "Combinational and computational challenge for biocatalyst design," *Nature (London)* **24**, 253–257 (2001).
3. J. M. Kinser, "Mining DNA data in an efficient 2D optical architecture," *Proc. SPIE* **4089**, 104–110 (2000).
4. G. A. Betzos, P. A. Mitkas, and C. W. Wilmsen, "Design of an optoelectronic biomolecular scanning system using smart VCSEL and photodetecter arrays," in *Optics in Computing,* OSA Technical Digest, pp. 38–40, Optical Society of America, Washington, DC (1999).
5. J. Tanida, "String data alignment by spatial coding and moiré technique," *Opt. Lett.* **409**, 1681–1683 (1999).
6. J. Meidanis and J. C. Setubal, *Introduction to Computational Biology,* PWS-Kent, Boston (1996).
7. J. Tanida and K. Nitta, "String data matching based on a moiré technique using 1-D spatial coded patterns," *Proc. SPIE* **4089**, 16–23 (2000).
8. "Clustal W Multiple Sequence Alignment," http://clustalw.genome.ad.jp/
9. T. A. Brown, *Genomes* (2000).
10. S. Yokozeki and K. Patorski, "Moiré fringe profile prediction method and its application to fringe sharpening," *Appl. Opt.* **17**, 2541 (1978).

Kouichi Nitta received the ME from Osaka University, Suita, Osaka, Japan, in 2000. He is a member of JSAP and OSJ.

Hiroyuki Togo received the ME from Osaka University, Suita, Osaka, Japan, in 2001.

Jun Tanida received the ME and DE from Osaka University, Japan, in 1983 and 1986, respectively. His research interests are in optical digital computing, optical analog information processing, and photonic bioinformatics. He is a member of JSAP, OSJ, OSA, SPIE, and IPSJ.