Chromosome Image Recognition with Local Band Patterns

Toru Abe [1]   Chieko Hamada [2]   Tetsuo Kinoshita [1]

[1] Cyberscience Center, Tohoku University   2-1-1 Katahira, Aoba, Sendai, 980-8577 Japan
[2] Graduate School of Information Sciences, Tohoku University   2-1-1 Katahira, Aoba, Sendai, 980-8577 Japan

Abstract

To make a visual examination of a chromosome image for various chromosome abnormalities, individual chromosome regions have to be determined in the subject image and classified into the distinct chromosome types. We propose a subregion based method to improve this process. The proposed method regards each chromosome region as a series of subregions (local band patterns), and iterates a search for subregions in the image consecutively.

1   Introduction

A visual examination of chromosome images for various chromosome abnormalities plays an important role in many clinical practices, including treatment and prevention of genetic disorders, radiation dosimetry, toxicology, etc.

Usually, the visual chromosome examination requires the following procedures [1]:

1. staining a set of chromosomes in a cell nucleus and capturing its image,

2. determining individual chromosome regions in the subject image,

3. classifying the determined regions into the 24 distinct chromosome types (1, 2, . . . , 22, X, and Y).

With proper staining methods (e.g. G-banding method etc.), a characteristic series of light and dark bands appears along the longitudinal axis of a chromosome (Figure 1). The band appearance on a chromosome is called a band pattern, and it is unique to each type of chromosome. For determining and classifying the chromosome regions in an image, many methods have been proposed. In most of them, individual chromosome regions are extracted from the subject image, the longitudinal profile of intensity on each region is acquired as its band pattern, and the region is classified according to the band pattern [1–4]. The classification result is inspected for abnormalities of number, where there are one or more entire chromosomes additional to or missing from the normal complement. The band pattern on each region is visually examined for abnormalities of structure, where part of the bands are lost (deletion), repeated (duplication), shifted (translocation), etc [6].
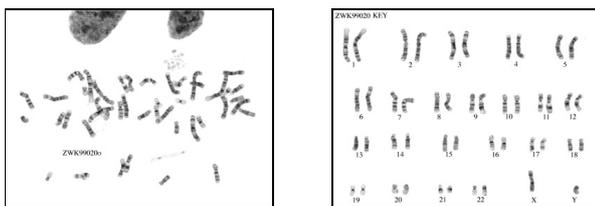
Chromosome regions in an image frequently touch or overlap each other, and have some parts difficult to distinguish them from the background. Consequently, the accurate extraction of individual chromosome regions is not easy. However, most of the existing methods for chromosome region classification require that individual chromosome regions are extracted accurately from the subject image in advance. When aberrant bands occur partly on a chromosome region because of various reasons (not only region extraction failure but also region overlap, structural chromosome abnormalities, etc.), it is difficult to classify the region into a correct chromosome type.

To overcome this problem, we propose a novel chromosome image recognition method based on local band patterns. The proposed method regards each chromosome region as a series of subregions (local band patterns), and iterates a search for subregions in the subject image consecutively. As a result, chromosome region classification can be performed simultaneously with its determination for each subregion, and the results of preceding subregion searches can be utilized for the following subregion searches.

2   Chromosome Image Recognition with Local Band Patterns

In the proposed method, a reference band pattern is prepared for every chromosome type. Each reference band pattern for an entire chromosome region is divided into several parts, and they are used as the templates for recognizing (determining and classifying) the subregions of the chromosome. In the following, the divided parts are referred to as local band patters (lbp), and the m th local band pattern on the chromosome type i is denoted by $lbp_m^{(i)}$ (Figure 2 (a)).

Firstly, the subject image is searched for the subregion corresponding to a template. If the subregion corresponding to $lbp_m^{(i)}$ is detected, secondly, the neighborhood of the detected subregion is searched for the next subregion corresponding to the adjoining $lbp_{m-1}^{(i)}$ or $lbp_{m+1}^{(i)}$ (Figure 2 (b)).

Figure 1: G-banded chromosome image (left), and its classification result (right) [5].
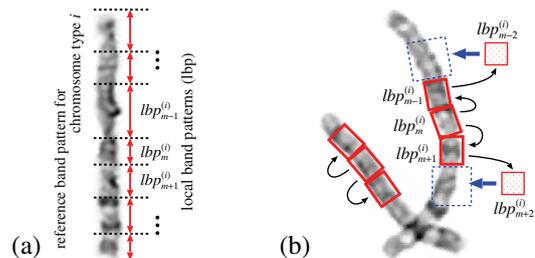
Figure 2: (a) local band patterns (lbp), (b) chromosome region recognition with local band patterns.

By iterating the search for corresponding subregions consecutively, with the first detected subregion as the starting point, one subregion after another is detected. Through this iteration, the entire region of a chromosome is determined in the image, and at the same time the chromosome type i is assigned to the region.

2.1 Subregion search

The subregion search is made by scanning an image with a template and seeking for subregions where the mean-squared-error (MSE) to the template are sufficiently small. Since the dimensions and intensities of chromosome regions vary with every image, the proposed method adjusts the dimensions (widths and lengths) of templates to those of chromosome regions in the subject image and adapts the intensities in the subject image to those of the templates.

Suppose that the reference image I_r is used for making templates and the subject image I_s is searched with these templates. By binarizing I_r on the basis of intensities, the mode w_r of region widths and the sum A_r of region areas are measured in the binarized image. From these measurements, the sum L_r of region lengths can be estimated by $L_r = A_r/w_r$. Similarly, w_s, A_s, L_s are determined in the binarized I_s. To adjust the template dimensions, the proposed method magnifies the template width and length by w_s/w_r and L_s/L_r, respectively.

The MSE to a template lbp is computed at every position within a search area set in I_s. As shown in Figure 3 (a), when lbp is set at (x, y) and rotated by θ in I_s, the MSE $e^2(x, y)$ is calculated by:

$$e^2(x, y) = \frac{1}{UV} \sum_{u=0}^{U-1} \sum_{v=0}^{V-1} (I_s'(x', y') - lbp(u, v))^2, \quad (1)$$

$$I_s'(x', y') = \alpha I_s(x', y') + \beta, \quad (2)$$
$$x' = x + u\cos\theta - v\sin\theta, \quad (3)$$
$$y' = y + u\sin\theta + v\cos\theta, \quad (4)$$

where U and V are the length and width of the template lbp, respectively. For every position, the rotation angle θ is set to minimize $e^2(x, y)$, and to adapt the intensities in I_s, the intensities of pixels overlapped with lbp are transformed by (2), where α and β are set to minimize $e^2(x, y)$.

2.2 Search area

When the subregion corresponding to a template $lbp_m^{(i)}$ is detected, the area should be searched for the subregion corresponding to the following $lbp_{m-1}^{(i)}$ can be narrowed.

As shown in Figure 3 (b), the proposed method sets a sector for the preceding subregion search result, circumscribes a rectangle about the sector, and uses this rectangle
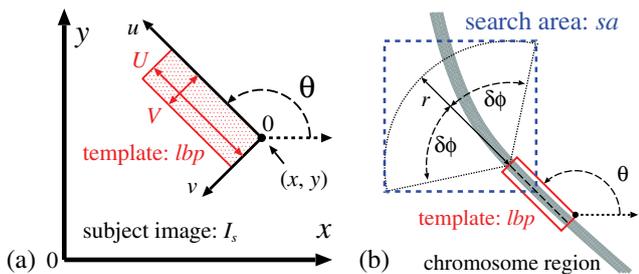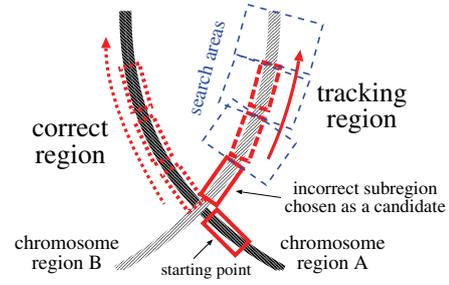
Figure 4: An error in subregion searches and tracking incorrect chromosome region ($k = 1$).

as the search area sa for the following subregion search. The sector is set at the end of the preceding detected subregion with radius r, direction θ, and central angle $2\delta\phi$, where r is determined from the length of the following template and θ is set to the rotation angle of the preceding template.

2.3 Combination of candidate subregions

The MSE to a template is computed for every subregion in a search area. Assume that subregions in the search area are sorted by their MSE in ascending order, and consider that the top k subregions are chosen as the candidates to correspond with the template.

When the subregion of the least MSE is chosen as the only candidate ($k = 1$) and an incorrect subregion is chosen as a candidate, inappropriate search areas are set and an incorrect chromosome region is tracked in the rest of the subregion searches (Figure 4). To deal with this difficulty, the proposed method chooses more than one candidate subregion in the search area for every template, and determines the optimum combination of the candidate subregions for each chromosome region.

Suppose that the subregion corresponding to a template $lbp_m^{(i)}$ is detected as a starting point and the search area $sa_{m-1}^{(i)}$ for $lbp_{m-1}^{(i)}$ is made from it (Figure 5 (a)). When the top k ($k > 1$) candidate subregions are chosen in $sa_{m-1}^{(i)}$, an individual search area $sa_{m-2}^{(i)}$ for $lbp_{m-2}^{(i)}$ is made from each candidate subregion. The proposed method circumscribes a rectangle about all of $sa_{m-2}^{(i)}$, and uses this larger rectangle $SA_{m-2}^{(i)}$ as the search area for $lbp_{m-2}^{(i)}$ (Figure 5 (b)).

To determine the optimum combination of the candidate subregions for each chromosome region, the proposed method selects a subregion from the set of candidate subregions for every lbp so as to minimize the overall MSE on the combination of the selected subregions, and imposes the

Figure 3: (a) template lbp set at (x, y) and rotated by θ in I_s, (b) search area sa made from the preceding search result.
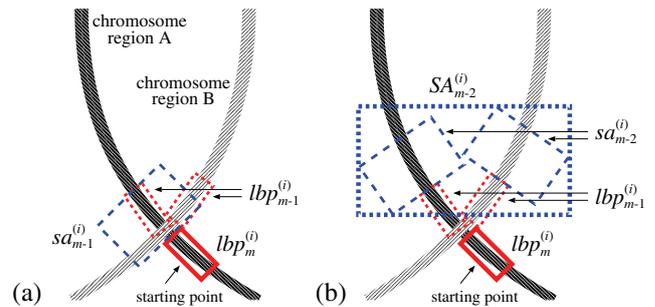
Figure 5: (a) search area sa made from the preceding search result, (b) larger search area SA generated from all of sa.
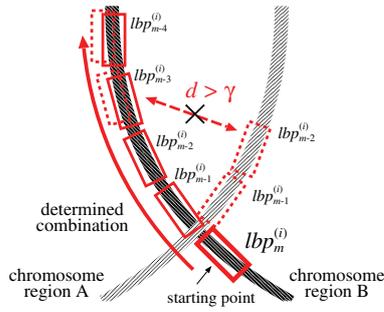
Figure 6: Combination of candidate subregions.

Figure 8: lbp used in the experiments as templates.

condition that the distance d between two adjoining subregions is less than a given threshold γ (Figure 6).

3 Experiments

We implemented a prototype system based on the proposed method, and conducted chromosome region recognition experiments.

3.1 Chromosome images

The experiments were carried out on the chromosome images that are opened to public by the website at the Wisconsin State Laboratory of Hygiene and ZooWeb [5]. This site provides not only original chromosome images but also their classification results. Examples of the original chromosome image and its classification result are shown in Figure 1. Although the proposed method can be applied to the original chromosome images, it is difficult to evaluate the recognition results in them. Therefore, the experiments were conducted on the classification result images, where every chromosome region was extracted, classified, and arranged in standard order.

Figure 7 shows examples of the classification result and its binarized image. Thirty-one classification result images were used in the experiments. Each image was 768×576 pixels in size, and characters in it were removed. These images were divided into two sets: one (15 images) was used as subject images, and the other (16 images) was used as reference images employed for making templates (lbp).

3.2 Templates

To make templates for subregion searches, firstly, chromosome regions were extracted from the reference images. On the medial axis determined in each extracted region, average intensities were taken perpendicularly to the medial axis, and they were used as an intensity profile. As a result, for each chromosome type i, intensity profiles $P_n^{(i)}$ ($n = 1, 2, \ldots, N^{(i)}$) were made. Since region extraction and medial axis determination failed for some chromo-

some regions, the number $N^{(i)}$ of intensity profiles differed depending on the chromosome type i.

Secondary, for each chromosome type i, the longest intensity profile $Pl^{(i)} = P_l^{(i)}$ was determined, and the lengths of $P_n^{(i)}(n \neq l)$ were extended to that of $Pl^{(i)}$. The average intensity profile was made from all $P_n^{(i)}$, and it was used as the reference band pattern $R^{(i)}$ of the chromosome type i.

Finally, by dividing each $R^{(i)}$ into several parts, templates $lbp_m^{(i)}$ for subregion searches were made. To adjust the dimensions of $lbp_m^{(i)}$, the mode w_r of region widths and the sum L_r of region lengths were determined in the reference image where $Pl^{(i)}$ was acquired. As shown in Figure 8, lbp were made for the chromosome types 1, 2, $\ldots$, 5, and a single lbp (marked with '$*$' in Figure 8) was selected to use for the starting point in each chromosome type.

3.3 Experimental results

In the experiments, the parameters r and $\delta\phi$ of a sector (described in 2.2), the threshold γ of the distance between two adjoining subregions, and the number k of chosen candidate subregions (described in 2.3) were set as follows:

$$
\begin{aligned}
r &= 1.5 \times \text{the length of the following template} \\
2\delta\phi &= 180° \ (\delta\phi = 90°) \\
\gamma &= 15\text{pixels} \\
k &= 1, \ 3
\end{aligned}
$$

To find starting points for consecutive subregion searches, the entire area of an image should be searched for the subregions corresponding to a lbp. Currently, we have not attained necessary accuracy in finding starting points. For this reason, in the experiments, starting points were set manually in the subject images. As shown in Figure 9, for each chromosome region in every subject image, we set a single lbp manually at the corresponding position as the starting

Figure 7: Classification result and its binarized image.
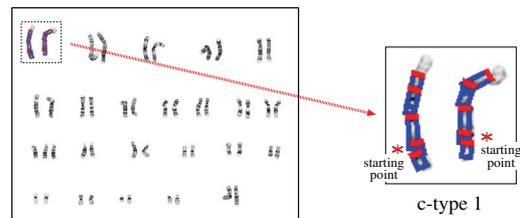
Figure 9: Examples of chromosome region recognition results ($k = 3$).

(a) c-type 2　　(b) c-type 3　　(c) c-type 4　　(d) c-type 5
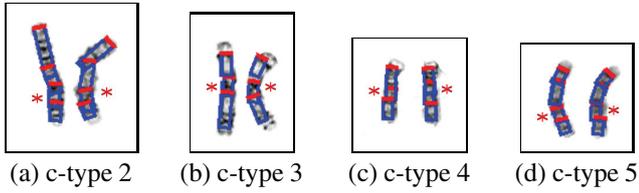
Figure 10: Examples of success in recognition ($k = 3$).

(a) c-type 1　　　　　(b) c-type 5
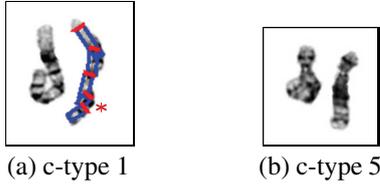
Figure 11: Examples of failure in recognition ($k = 3$).

point, and started the iteration of a search for subregions from it. In Figure 9, three candidate subregions at most were chosen for each lbp ($k = 3$), and the subregions selected as the optimum combination are represented by the rectangles (the starting points are marked with '$*$').

For a region of the chromosome type i, if candidate subregions were detected to all $lbp_m^{(i)}$ and the optimum combination of the subregions was determined, we decided that recognition succeeded on this region. Otherwise, we decided that chromosome region recognition failed. Examples of success in chromosome region recognition ($k = 3$) are shown in Figure 10, and examples of recognition failure are shown in Figure 11. The optimum combination of subregions was determined correctly for the right chromosome region in Figure 11 (a). However, because the left region in Figure 11 (a) and the left region in (b) curved windingly, some candidate subregions was not detected correctly, and thus the optimum combination could not be determined for them. The optimum combination was not determined either for the right region in Figure 11 (b), because the band pattern on this region was not clear. For 150 chromosome regions (2 chromosome regions × 5 chromosome types × 15 images), in a case where $k = 1$, the success rate of chromosome region recognition was $101/150 = 67\%$. Compared to this, in a case where $k = 3$, the success rate was increased to $124/150 = 83\%$.

Figure 12 and 13 show examples of chromosome region recognition results on an original image and a synthesized image, respectively. In Figure 12, although a chromosome region touched with another region, it was recognized correctly. In Figure 13, a region of chromosome type 6 overlapped with two regions of chromosome type 1, and a search for subregions started from $lbp_4^{(1)}$. As shown in Figure 13 (b) and (c), the subregions of chromosome type 1 and 6 were chosen as candidates for $lbp_2^{(1)}$. In this case, although the MSE of the candidate region on chromosome type 6 was less than that of the candidate region on chromosome type 1, the region of chromosome type 1 was recognized correctly through the determination of the optimum combination of the candidate subregions.

4　Conclusions

In this paper, we have proposed a subregion based method for improving the recognition of chromosome regions in an image. The proposed method regards each chromosome region as a series of subregions (local band
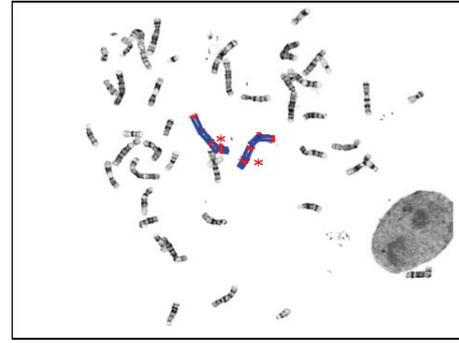
Figure 12: Chromosome region recognition results on an original chromosome image ($k = 3$, c-type 1).

(a) synthesized image　(b) candidate subregions　(d) recognition results

(c) search areas
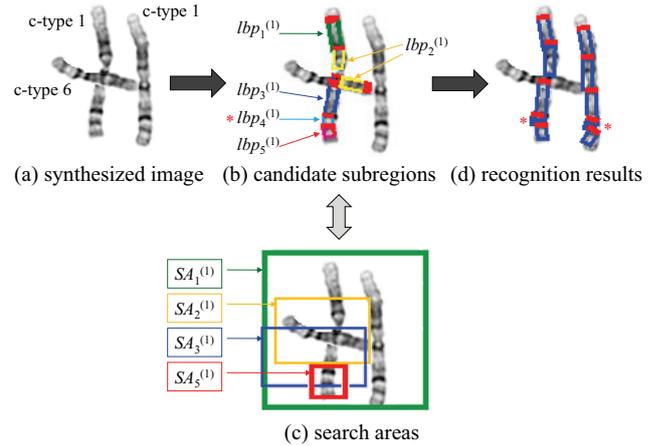
Figure 13: Chromosome region recognition results on a synthesized image ($k = 3$).

patterns), and iterates a search for subregions in the image consecutively. In the proposed method, the success rate of chromosome region recognition can be increased by choosing more than one candidate subregion for every template and determining the optimum combination of the candidate subregions for each chromosome region.

In future work, we plan to develop following methods and conduct further experiments on chromosome images:

- A method to search efficiently for the starting points in the subject image.

- A method for determining effective subregion search templates in each reference band pattern.

References

[1] J.Graham and J.Piper: "Automatic karyotype analysis," *Meth. Mol. Biol.*, vol.29, pp.141–185, 1994.

[2] A. Carothers and J. Piper: "Computer-aided classification of human chromosomes: a review," *Statistics Computing*, vol.4, no.3, pp.161–171, 1994.

[3] Q. Wu, Z. Liu, T. Chen, Z. Xiong, and K.R. Castleman: "Subspace-based prototyping and classification of chromosome images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol.14, no.9, pp.1277–1287, 2005.

[4] M. Moradi and S.K. Setarehdan: "New features for automatic classification of human chromosomes: a feasibility study," *Pattern Recognit. Lett.*, vol.27, no.1, pp.19–28, 2006.

[5] ZooWeb — Karyotypes home page: http://worms.zoology. wisc.edu/zooweb/Phelps/karyotype.html.

[6] S.L. Gersen and M.B. Keagle, eds.: The principles of clinical cytogenetics, 2nd ed., Human Press, 2004.