RG-001

# 染色体画像中での部分領域探索法 A Method of Searching for Subregions in Chromosome Images

阿部 亨 <sup>†</sup> Toru Abe <sup>†</sup> 濱田 智栄子‡

Chieko Hamada <sup>‡</sup>

# 1 はじめに

光学顕微鏡画像に基づき,染色体の種々の異常を視覚的 に診断する作業は,染色体異常に起因する疾病の予防や治 療,染色体異常を誘発する物質の判定試験などにおいて重 要な役割を果たす[1].近年,染色体の視覚的診断に関す る知見が深まり,その重要性が高まるにつれ,視覚的診断 に必須の処理である染色体画像解析(染色体領域の抽出・ 同定)の自動化が以前にも増して強く求められている.対 象画像から個々の染色体領域を抽出し,各領域が何番の染 色体に対応するかを同定する手法については,多くの研究 が行われており,一部は製品化もなされている.しかし, 十分な精度が得られる手法は未だ実現されておらず,視覚 的診断に際しては,専門家による手作業で染色体領域の抽 出・同定が行われる場合が多い.

これに対し,筆者らは,1本の染色体領域を複数の部分 領域の連なりと見なし,部分領域を対象画像中で探索する 処理を順に繰り返すことで,染色体領域全体を高精度に抽 出・同定する手法を提案している[2].本稿では,この提 案手法中,染色体の部分領域を探索する処理を効果的に実 現する方法の検討を行い,染色体画像を対象とした実験に より,提案する部分領域探索手法の有効性を示す.

### 2 染色体画像解析

# 2.1 染色体画像解析の一般的手順

染色体は,遺伝情報を担う生体物質であり,ヒトの正常な細胞核1個には,常染色体44本と性染色体2本の計46本が含まれる.常染色体は22種類各2本で構成され, 各々1,2,...,22の番号で区別される.性染色体はX,Y で区別され,正常な女性はXを2本,正常な男性はXと Yを1本ずつ持つ.各染色体は,動原体と呼ばれる箇所 でくびれており,そこで短腕と長腕に二分される[1].

染色体の視覚的診断には,染料により縞(バンド)を染め出した染色体(細胞核1個分)の光学顕微鏡画像が用いられる(図1(a)).染料としてギムザを使用するG分染法などの一般的な染色法を用いた場合,視覚的診断は,以下の手順で作業が進められる[3].

- 1. 対象画像から個々の染色体領域を抽出
- 2. 各領域が何番の染色体に対応するかを同定
- 3. 各領域の状態から染色体の異常を視覚的に診断

通常,まず,対象画像から個々の染色体領域を抽出し,

次に,抽出した各領域の特徴に基づき,各々が何番の染色体に対応するか同定を行う(図1(b)).同じ番号の染色体でも,その大きさや輝度値は,撮影条件や染色状態等により変化するため,染色体領域の長さ,幅,輝度値は画像毎に異なる(同じ画像中では,異なる番号の染色体でも領域幅はほぼ一定となる).これに対し,染色体領域の相対的長さ,各領域での短腕と長腕の長さの比,各領域でのバンドの並び方(バンドパターン)は,異なる画像でも染色体番号毎にほぼ一定となるため,染色体領域の同定には,これらの特徴が用いられる[5].同定結果に基づき,染色体の本数が通常とは異なる数的異常の診断が行われる.さらに,同定された各領域のバンドパターンを観察することで,遺伝情報の欠失(一部のバンドが欠ける),重複(余分なバンドが生じる),転座(バンドの位置が通常とは異なる)など,染色体の構造異常が診断される.

2.2 染色体画像解析のための従来手法

対象画像から個々の染色体領域を抽出し,得られた各領 域が何番の染色体に対応するか同定を行う処理について は,その重要性から種々の研究がなされている.

一般に,個々の染色体領域の抽出は,同定とは独立の前 処理として,対象画像に対する輝度値の2値化により行 われる[6-11].画像中の染色体領域には,領域同士の重 なりや接触が生じる箇所,背景との区別が難しい箇所が存 在する.このため,単なる2値化では,複数本の染色体が 連結した1つの領域として抽出されたり,逆に,1本の染 色体が分離した複数の領域として抽出されたりする場合 があり,個々の染色体領域の正確な抽出は非常に困難であ る.そこで,2値化の結果を分離・連結し個々の染色体領 域を決定するために,バンドパターン等の情報を用いる手 法が提案されている[12,13].しかし,この手法では,前 処理(抽出)の段階で,領域の特徴に基づく同定結果の一 部が必要になるという問題が生じる.

染色体領域の同定には,領域の相対的長さなどの特徴を 利用できるが,染色体番号を一意に決定するためには,各 領域のバンドパターンが特徴として必要となる.バンド



1 < () (b)

図1 (a) 染色体画像(G分染法), (b) 抽出·同定結果[4]

<sup>†</sup> 東北大学サイバーサイエンスセンター

<sup>‡</sup> 東北大学大学院情報科学研究科

パターンに基づく手法では,多くの場合,染色体番号毎に 標準のバンドパターン(参照バンドパターン)を用意し, 抽出された領域のバンドパターン(領域の中心線に沿った 輝度値)と比較することで同定が行われる[7,8].その際, 染色体領域の大きさや輝度値は画像毎に異なるため,その ままでは,各領域のバンドパターンと参照バンドパターン とを正しく比較することができない.そこで,従来手法で は,抽出された領域の長さの中央値に基づき画像毎に領域 長を正規化し,さらに,輝度値の平均・分散を領域毎に正 規化する方法がとられている[14].しかし,これらの手 法では,個々の染色体領域の正確な抽出が事前に必要であ り,他にも,1本の染色体領域内で長さや輝度値が局所的 に変動している場合への対応が困難という問題がある.

また,中心線の決定誤り,領域の重なり,構造異常など により,見掛けの一部に異常が生じたバンドパターンで は,領域全体(1本分)のバンドパターンを比較に用いる と,正確な同定が行えない場合がある.そこで,染色体領 域の部分的特徴(特徴的な一部のバンドなど)のみを用い 同定を行う手法が提案されている[6,9].この手法では, 通常とは見掛けの異なるバンドの影響を抑えることが可 能となるものの,利用できる特徴自体が少なくなるため, 一般に,同定精度が低くなるという問題がある[8].

3 提案手法における処理の流れ

前述のように,染色体画像を解析(抽出・同定)する従 来手法には,次の問題があった.

(P1) 前処理による染色体領域個々の正確な抽出が困難

(P2) 領域長や輝度値が変動した場合への対応が困難

(P3) バンドの見掛けに異常が生じた場合への対応が困難

これら従来手法の問題の解決を図り,染色体領域の高精度 な抽出・同定を実現するために,筆者らは,1本の染色体 領域を複数の部分領域の連なりと見なし,部分領域を対象 画像中で探索する処理を順に繰り返すことで,染色体領域 全体を抽出・同定する手法を提案している[2].

提案手法では,抽出・同定を行う際のテンプレートとして,図2に示すように,染色体番号毎に用意した参照 バンドパターンを各々複数の部分に分割したものを用いる.以降,これを部分バンドパターン (lbp: local banding pattern)と呼び,*i* 番染色体の*m* 番目の部分バンドパター ンを $lbp_m^{(i)}$ で表す.

まず,部分バンドパターンを1つ選択し,図3に示す ように,これをテンプレートに用い,類似する箇所(部分 領域)を対象画像中で探索する.テンプレート $lbp_m^{(i)}$ に類 似する部分領域を検出した場合,次に,その近傍に探索範 囲を設定し,隣接するテンプレート $lbp_{m-1}^{(i)}$ , $lbp_{m+1}^{(i)}$ に類 似する部分領域を探索する.この反復により,最初に検出 した箇所を起点に,隣接部分バンドパターンに類似する部 分領域を順に探索し,1本の染色体領域全体を芋蔓式に抽 出・同定する.

以上のアプローチにより,提案手法では,従来手法の問題(P1)~(P3)に対し,次の対応を図る.



図 2 参照バンドパターンと部分バンドパターン (lbp)



図3 部分領域の探索による染色体画像解析

- (S1)部分領域単位で抽出と同定を同時に実行 部分領域の探索・探索範囲の設定を繰り返すことで, 直前の探索結果を利用しながら,部分領域単位で抽出 と同定を同時に(個々の染色体領域を事前に抽出する ことなく)実行.
- (S2)部分領域単位で領域長・輝度値を補正 部分領域を探索する際,部分領域単位で領域長・輝度 値を補正することで,個々の染色体領域を事前に抽出 することなく,領域長・輝度値の変動に対応.
- (S3) 信頼できる部分領域の特徴を統合し利用 探索に失敗した部分領域を,通常とは見掛けの異なる バンドが生じている箇所として除き,残りの部分領 域の特徴を統合し利用することで,高精度な同定を 実現.

## 4 染色体画像中の部分領域探索手法

前節で述べた提案手法の抽出・同定精度は,対象画像 中での部分領域探索精度に大きく左右される.本節では, lbpをテンプレートに用い,対象画像中の部分領域を効果 的に探索する手法について説明する.以下では,探索を行 う対象画像を $I_S$ で,参照バンドパターン(及びlbp)の作 成に用いる参照画像を $I_R$ で表す.また, $I_S$ , $I_R$ の各々を 輝度値で2値化し,染色体領域を大まかに求めた画像を  $B_S$ ,  $B_R$ で表す.

4.1 *lbp* の大きさの補正

テンプレートとなる lbp は,前述の通り,染色体番号毎 に用意した参照バンドパターンを分割し作成する.N 枚 の参照画像  $I_{\text{R}n}$  (n = 1, 2, ..., N) から参照バンドパター ンを作成するためには,まず,各  $I_{\text{R}n}$  から個々の染色体 領域を抽出し,図2に示すように,各領域で中心線を決定 する.中心線上の各点で,中心線に直交する方向での輝度 値の平均を求め,これを領域のバンドパターンとして用 いる.次に,染色体番号毎に,バンドパターンの長さを揃 え,各点での輝度値の平均を求めることにより参照バンド パターンを作成する.

提案手法では、部分領域の探索を行う際、 $I_{\rm S}$ 中の染色体 領域の大きさへlbpの大きさを対応させるために、 $I_{\rm S}$ 中で の染色体領域の幅の最頻値と長さの総和に基づき、lbpの 大きさの補正を行う.まず、 $B_{\rm R}$ と $B_{\rm S}$ の各々で、領域の 総面積(総画素数) $A_{\rm R}, A_{\rm S}$ と領域幅の最頻値 $w_{\rm R}, w_{\rm S}$ を 求める、染色体領域の幅が各画像中で一定と仮定すれば、 染色体領域の長さの総和は、個々の染色体領域を抽出する ことなく、各々 $L_{\rm R} = A_{\rm R}/w_{\rm R}, L_{\rm S} = A_{\rm S}/w_{\rm S}$ で推定でき る[15].そこで、提案手法では、lbpの幅 V として $w_{\rm S}$ を 用い、長さUとして、lbpの元の長さを $L_{\rm S}/L_{\rm R}$ 倍したも のを用いる.

#### 4.2 部分領域の探索

染色体の部分領域を探索する際には, $I_{\rm S}$ 中で領域と判定された各箇所( $B_{\rm S}$ から決定)にテンプレートである lbpの原点を設定し,lbpと $I_{\rm S}$ との輝度値の差の2乗平均(MSE: mean-squared-error)を求め,MSE が小さい順に 最大k箇所をlbpに対応する部分領域の候補として選択 する.

図 4 (a) に示すように, *lbp* (長さ U, 幅 V)の原点 (u,v) = (0,0)が $I_{\rm S}$ 中の(x,y)に位置する場合, *lbp*を回 転させながら次式により各  $\theta$  で  $e^2(x,y)$ を求め,  $e^2(x,y)$ が最小となる  $\theta$  での値を(x,y) での MSE とする.

$$e^{2}(x,y) = \frac{1}{UV} \left( E_{\rm c}^{2}(x,y) + E_{\rm b}^{2}(x,y) \right)$$
(1)

$$E_{\rm c}^2(x,y) = \sum_{(u,v) \in O_{\rm c}} \left( \alpha I_{\rm S}(x',y') + \beta - lbp(u,v) \right)^2$$
(2)

$$E_{\rm b}^2(x,y) = \sum_{(u,v)\in O_{\rm b}} \left( I_{\rm S}(x',y') - lbp(u,v) \right)^2 \tag{3}$$

$$x' = x + u\cos\theta - v\sin\theta \tag{4}$$

$$y' = y + u\sin\theta + v\cos\theta \tag{5}$$

各  $\theta$  では,図 4 (b) に示すように,lbp の各箇所 (u,v)が, $I_{\rm S}$ 中の領域と重なる箇所  $O_{\rm c}$  であるか,背景と重なる箇所  $O_{\rm b}$  であるかを  $B_{\rm S}$  から決定し,各々で別々に得られた輝度値の差の2乗和  $E_{\rm c}^2(x,y), E_{\rm b}^2(x,y)$ から,式(1)により  $e^2(x,y)$ を求める. $E_{\rm c}^2(x,y)$ を  $O_{\rm c}$ で求める際には,染色体領域の輝度値の変動による影響を軽減するた





め,式(2)に示すように, $I_{\rm S}$ 中の輝度値を部分領域単位で 補正する.具体的には,式(6),(7)で得られる $\alpha,\beta$ を用い た線形変換を行い,各 $x, y, \theta$ の組に対し $E_c^2(x, y)$ の最小 化を図る(式(6),(7)中, $|O_c|$ は, $O_c$ の画素数を表す).

$$\alpha = \frac{|O_{c}| \sum_{(u,v) \in O_{c}} I_{S}(x',y') lbp(u,v)}{-\sum_{(u,v) \in O_{c}} I_{S}(x',y') \sum_{(u,v) \in O_{c}} lbp(u,v)}{|O_{c}| \sum_{(u,v) \in O_{c}} I_{S}^{2}(x',y') - \left(\sum_{(u,v) \in O_{c}} I_{S}(x',y')\right)^{2}}$$
(6)  
$$\beta = \frac{1}{|O_{c}|} \left(\sum_{(u,v) \in O_{c}} lbp(u,v) - \alpha \sum_{(u,v) \in O_{c}} I_{S}(x',y')\right)$$
(7)

 $E_{\rm b}^2(x,y)$ を $O_{\rm b}$ で求める際は, $I_{\rm S}$ 中の輝度値を $O_{\rm c}$ と同様に補正すると,lbpが染色体領域から大きくはみ出している場合でも MSE が小さくなり,lbpに対応すると誤って判定される箇所が増加する.そのため, $O_{\rm b}$ では,式(3)に示すように,輝度値を補正しない.

4.3 部分領域の探索に適した *lbp* の選択

参照バンドパターンを分割し作成した *lbp* は,その分割 位置・長さにより,部分領域探索に用いた場合の効果が異 なる.バンドパターン中の特徴的(染色体領域の他の箇所 との区別が容易)な箇所をテンプレートとして求める手 法としては,Lockwoodらが,対応箇所以外との相関が小 さいものを有効なテンプレートとして選択する手法を提 案している[16].しかし,この選択手法は,染色体の模式 図(標準的なバンドパターンを2値で描いた図)を対象と したものであり,個々の染色体領域における特徴量(領域 長,輝度値)の変動を考慮していない.

そこで,本稿では,実際の染色体画像を対象に,個々の 染色体領域における特徴量の変動も考慮したlbp選択手 法を提案する.提案手法では,各 $I_{Rn}$ から得られたバン ドパターン(屈曲を除き直線状に補正したもの)を対象 に,各lbpをテンプレートとして用いた場合に探索を誤る 頻度を推定し,誤りが少ないlbpを部分領域の探索に有効 なテンプレートとして選択する.具体的には,まず,lbpとの MSE  $e^2$ を,  $2I_{Rn}$ から得られた全バンドパターン (数的異常が無いN 枚の $I_{Rn}$ を用いた場合は46 × N本 のバンドパターン)の全箇所で計算し,正解箇所(当該 lbp に対応する部分領域)での MSE の頻度分布 $f_c(e^2)$  と 正解箇所以外での頻度分布 $f_i(e^2)$ を求め正規分布で近似 する.多くの場合,両者は図5に示すような関係となる ため, $f_c(e^2_h) = f_i(e^2_h)$ となる $e^2_h$ を決定し,式(8)で得ら



図 5 MSE e<sup>2</sup> の頻度分布

れる *F<sub>e</sub>* (図 5 中の斜線部分の面積)を正解箇所の探索を 誤る頻度の推定値として用いる.

$$F_{e} = \int_{0}^{e_{\rm b}^{2}} f_{\rm i}(t) \,\mathrm{d}t + \int_{e_{\rm b}^{2}}^{\infty} f_{\rm c}(t) \,\mathrm{d}t \tag{8}$$

実際の画像中での染色体領域は屈曲しているため,部分 領域の探索を行う際には,ほとんどの場合,テンプレート として用いた lbpの一部が染色体領域からはみ出し,背景 と重なる箇所  $O_b$ が生じることになる. $O_c$ , $O_b$ の画素数  $|O_c|, |O_b|$ は,対象とする  $I_S$ ,あるいは,その中での箇所 により変化するものの,各  $I_S$ での平均 $\overline{|O_c|}, \overline{|O_b|}$ は,用 いる lbpと同じ長さ・幅のマスクにより  $I_S$ を2値化した  $B_S$ を走査すれば求めることができる.そこで,各 lbpに 対する  $F_e$ を計算する際,対象とする  $I_S$ 毎に $\overline{|O_c|}, \overline{|O_b|}$ を 求め,MSEの頻度分布を補正することを考える.具体的 には,次式で補正した MSE  $e'^2$ の頻度分布から  $F_e$ を計算 する.

$$e'^{2} = \frac{|O_{\rm c}|e^{2} + |O_{\rm b}|e_{0}^{2}}{|O_{\rm c}| + |O_{\rm b}|}$$
(9)

ここで, $e^2$ は, $O_b$ が生じない場合の MSE(屈曲を除き直線状に補正したバンドパターンとlbpとの MSE)を, $e_0^2$ は,背景とlbpとの MSE を各々表す.

# 5 実験

本節では,提案する部分領域探索手法の有効性を検証す るために行った実験について述べる.

# 5.1 実験に用いた染色体画像

実験には、ZooWebとThe Wisconsin State Laboratory of Hygiene により公開されている染色体画像を用いた[4]. この画像には、原画像と共に、抽出・同定結果の画像も含 まれており、今回は、提案手法による部分領域探索結果の 評価を行うために、抽出・同定結果の画像を実験に使用し た(原画像に対しては、各領域の染色体番号が与えられて いないため、探索結果の正誤判定が困難となる).

実験に使用した画像は計 38 枚(数的異常,構造異常を 生じている例を含む)で,画像サイズは 768 × 576 画素で ある.これら 38 枚の画像を 19 枚ずつ 2 つのセット A, B に分割し,参照バンドパターンの作成と部分領域の探索に 交互に用い実験を行った.A には合計 878 本の染色体領 域が,Bには合計 889 本の染色体領域が含まれている.

実験に際しては,各画像を輝度値で2値化し,画像中の 領域箇所の決定,及び,染色体領域幅の推定と染色体領域 長の総和の推定に用いた.また,各画像中,領域と判定さ れた箇所での輝度値の平均と分散が一定となるよう正規 化を行った.さらに,画像中の全染色体領域で中心線を決 定し,参照バンドパターンの作成と探索結果の評価に使用 した.中心線の抽出には文献[10]の手法を用いたが,全 ての染色体領域で中心線を自動的に決定することは困難 であったため(自動的に決定できたのは7割程度であっ た),自動抽出に失敗した場合は手動で中心線を決定した. 図6に,実験に用いた染色体画像,2値化した画像,各染 色体領域で決定された中心線,実験に用いた参照バンドパ ターンの例を各々示す.

5.2 評価方法

探索結果として得られた部分領域候補の正誤判定を行 う際は,まず,正解箇所として,テンプレートであるlbpに対応する範囲を当該染色体領域の中心線上で決定し,始 点と終点を結ぶベクトル $cr_e - cr_s$ を求める.図7に示す ように,部分領域候補が正解箇所の中心線を60%以上覆 い,かつ,始点と終点を結ぶベクトル $sc_e - sc_s$ とベクト  $\nu cr_e - cr_s$ とのなす角が $90^\circ$ 未満の場合を正解(探索成 功)とし,それ以外の場合を誤り(探索失敗)とした.

探索精度の評価には,次式で定義される再現率 R(Recall)と適合率 P(Precision)を用いた. R は,全正解箇 所に対する探索成功箇所の割合を,P は,全部分領域候補 に対する探索成功箇所の割合を表している.

$$R = |C \cap D| / |C| \tag{10}$$

$$P = |C \cap D|/|D| \tag{11}$$

ここで,Cは,探索対象画像 $I_S$ での正解箇所の集合を, Dは,探索結果として $I_S$ で得られた部分領域候補の 集合を, $D \cap C$ は,探索成功箇所の集合を表す.また,  $|C|, |D|, |D \cap C|$ は,各々 $C, D, D \cap C$ の箇所数を表す. 5.3 実験結果

まず,4.2節で述べた提案手法(部分領域探索手法)の 効果を検証するために,以下の値に基づき部分領域を探索



図 6 (a) 実験に用いた染色体画像の例,(b) 2 値化した 画像,(c) 各染色体領域で決定された中心線,(d) 実験に 用いた参照バンドパターンの例



FIT2010(第9回情報科学技術フォーラム)



図 8 順位 k を変化させた場合の再現率 R と適合率 P (部分領域探索手法の比較)

する実験を行った.

- (a) *I*<sub>S</sub> 中の各箇所と*lbp* との相関
- (b) I<sub>S</sub>中の各箇所と lbp との MSE (輝度値補正なし)
- (c)  $I_{\rm S}$  中の各箇所と lbp との MSE( $O_{\rm c}, O_{\rm b}$  で輝度値補正)
- (d) *I*<sub>S</sub> 中の各箇所と *lbp* との MSE (*O*<sub>c</sub> でのみ輝度値補 正,提案手法)

探索に用いるテンプレートは,参照バンドパターンから箇 所・長さを変化させて多数の*lbp*を作成し,その中から以 下の方法により選択した.

- 上記(a)の場合は,各 *lbp* について,全参照バンドパターンを対象に,対応箇所以外での相関の最大値を求め,染色体番号毎に,相関の最大値が小さい順に3つの *lbp* を選択.
- 上記 (b)~(d) の場合は,各 lbp について,全参照バンドパターンを対象に,対応箇所以外での MSE の最小値を求め,染色体番号毎に,MSE の最小値が大きい順に3つの lbp を選択.なお,MSE を求める際は,部分領域を探索する場合と同様に(b)~(d) 各々の方法で輝度値を補正した(参照バンドパターンは,屈曲が除かれ直線状に補正されているため,lbp のとの MSEを求める際に Ob は生じない.従って,(c)と(d)では同じ lbp がテンプレートとして選択される).

各  $I_{\rm S}$  での探索結果は,各テンプレートについて,(a)の場合は相関の降順,(b)~(d)の場合はMSEの昇順に上位k位以内の部分領域候補を式(10),(11)中のDに,その中の探索成功箇所を $C \cap D$ に用いた.なお,探索に際しては,4.1節で述べた手法により,lbpの大きさの補正を行った.染色体番号毎に3つテンプレートを用い各 $I_{\rm S}$ で探索を行うため,(a)~(d)各々で $|C| = (878 + 889) \times 3 = 5301$ , $|D| = 38 \times 24 \times 3 \times k = 2736k$ となる.

順位 k を 1 から 50 の範囲で変化させて D を求め, R の平均と P の平均を横軸と縦軸に各々プロットした結果 を図 8 に示す.この結果から,(a)に比べ,(b),(c),(d)で は, R, P 共に値が向上していることが分る.これは,(a) では,テンプレートが染色体領域から大きくはみ出して重







図 10 部分領域探索結果の例((e)により得られた正解 箇所, k = 2)

なっている箇所でも相関が大きくなり,部分領域候補と 誤って選択される例が多かったためである.また,各画像 中,領域と判定された箇所での輝度値の平均と分散が一 定となるよう正規化を行っているにもかかわらず,(b)に 比べ,(c),(d) でR, P 共に値が向上していることから, $I_{\rm S}$ 中の輝度値を部分領域単位で補正した効果が確認できる. さらに,(c)よりも高いR, P の値が(d)で得られており, このことは, $I_{\rm S}$ 中の領域箇所 $O_{\rm c}$ でのみ輝度値の補正を 行う提案手法の有効性を示していると考えられる.

次に,4.3 で述べた提案手法(*lbp* の選択手法)により,探 索誤り頻度の推定値 $F_e$ に基づき選択された*lbp* で部分領 域の探索を行った結果を図9に示す.図9中の(d)は,先 の実験(図8)での(d)と同じものであり,(e)は,各染色 体番号に対し $F_e$ が小さい順に3つの*lbp*をテンプレート として選択し,(d)と同様の手法で部分領域を探索した結 果である.この結果から,(d)に比べ,(e)ではR,P共に 値が向上していることが確認でき,これは, $F_e$ に基づき *lbp*を選択する提案手法の有効性を示すものである.(e) で得られた部分領域探索結果の例を図 10に示す.この図 では,k = 2の場合の探索成功箇所(上位2位以内の部分 領域候補で正解のもの)を黒い矩形で表し,対象画像 $I_S$ を2値化した $B_S$ に重ね表示している.さらに,*lbp*の一 部が染色体領域からはみ出す場合(背景と重なる箇所 $O_b$ が生じる場合)を考慮し,補正した $e^{\prime 2}$ の頻度分布から求めた $F_e$ に基づき,各染色体番号に対し3つのlbpをテンプレートとして選択し,(d)と同様の手法で部分領域を探索した結果を図9(f)に示す.この結果から,(e)に比べ,(f)では, $k \ge 2$ の場合,R, P共に値が向上していることが確認できる.これは,lbpを選択する際, $O_b$ による影響を考慮した効果を表すものと考えられる.

なお,1つのlbpをテンプレートに用い,1枚の $I_{\rm S}$ 中で 部分領域探索を行う際に要する処理時間は, $I_{\rm S}$ 中で染色 体領域と判断された箇所の面積(画素数),大きさを補正 後のlbpの長さ・幅,探索の実装方法(lbpを回転する際 の単位回転角の設定)などの要因により異なる.今回行っ た実験の例では,いずれの場合も数秒程度を処理(1つの lbpで1枚の $I_{\rm S}$ を探索)に要している.

# 6 おわりに

本稿では,部分バンドパターン(*lbp*)をテンプレート に用い,染色体画像中で部分領域を効果的に探索する手法 (部分領域の探索手法,及び,探索に適した*lbp*の選択手 法)を提案し,さらに,染色体画像を対象にした部分領域 探索実験によりその効果を確認した.

今後は,実用的な染色体画像解析手法の実現を目指し, 以下の追加実験,検討を進める予定である.

- 今回の実験では,探索精度の評価が容易という理由から,各染色体領域が整列された状態の画像を対象に用いている.従って,今後は,図1(a)に示した画像のように,各染色体領域が抽出・同定される前の原画像を対象とした評価実験を行い,提案手法の効果に関するより詳細な検討を行う予定である.
- 今回提案した lbp の選択手法は,個々の染色体領域に おける領域長や輝度値の変動を考慮したものとなっ ている.しかし,染色体領域の屈曲や重なり等につい ては十分考慮されていないため,今回の提案手法を, これらの影響を考慮に入れたものに拡張する必要が ある.
- また、1枚の対象画像に含まれる全染色体領域を抽出・同定するためには、各染色体領域に対し少なくとも1つの部分領域を探索の起点として決定する必要がある.この要件を満たすために必要なlbpの条件(各染色体番号当り何個のlbpをテンプレートとして用意し、上位何位以内の探索結果を部分領域候補として選択するか)について検討を進める必要がある.

# 参考文献

- S. L. Gersen and M. B. Keagle Eds.: "The principles of clinical cytogenetics", Humana Press, 2nd edition (2004).
- [2] T. Abe, C. Hamada and T. Kinoshita: "A chromosome image recognition method based on subregions", Proc. The 9th Asian Conf. Computer Vision (2009).
- [3] J. Graham and J. Piper: "Automatic karyotype analysis", Meth. Mol. Biol., 29, pp. 141–185 (1994).

- [4] "ZooWeb Karyotypes home page", http://worms.zoology. wisc.edu/zooweb/Phelps/karyotype.html.
- [5] D. G. Harnden and H. P. Klinger Eds.: "ISCN1985: An international system for human cytogenetic nomenclature (1985)", S. Karger AG. (1985).
- [6] F. C. A. Groen, T. K. Kate, A. W. M. Smeulders and I. T. Young: "Human chromosome classification based on local band descriptors", Pattern Recognit. Lett., 9, 3, pp. 211–222 (1989).
- [7] G. Ritter and C. Pesch: "Polariy-free automatic classification of chromosomes", Computational Statistics Data Analysis, 35, pp. 351–372 (2001).
- [8] Q. Wu, Z. Liu, T. Chen, Z. Xiong and K. R. Castleman: "Subspace-based prototyping and classification of chromosome images", IEEE Trans. Pattern Anal. Machine Intell., 14, 9, pp. 1277–1287 (2005).
- [9] M. Moradi and S. K. Setarehdan: "New features for automatic classification of human chromosomes: A feasibility study", Pattern Recognit. Lett., 27, 1, pp. 19–28 (2006).
- [10] J. hong Kao, J. hui Chuang and T. Wang: "Chromosome classification based on the band profile similarity along approximate medial axis", Pattern Recognit., 41, 1, pp. 77–89 (2008).
- [11] B. Legrand, C. Chang, S. Ong, S.-Y. Neo and N. Palanisamy: "Chromosome classification using dynamic time warping", Pattern Recognit. Lett., 29, 3, pp. 215–222 (2008).
- [12] G. C. Charters and J. Graham: "Trainable grey-level models for disentangling overlapping chromosomes", Pattern Recognit., 32, 8, pp. 1335–1349 (1999).
- [13] W. Schwartzkopf, A. Bovik and B. Evans: "Maximumlikelihood techniques for joint segmentation-classification of multispectral chromosome images", IEEE Trans. Med. Imag., 24, 12, pp. 1593–1610 (2005).
- [14] J. Piper and E. Granum: "On fully automatic feature measurement for banded chromosome classification", Cytometry, 10, pp. 242–255 (1989).
- [15] 阿部,濱田,木下: "染色体画像解析のための領域長正規化 手法",信学論 D, **J91-D**, pp. 2983–2987 (2008).
- [16] D. H. Lockwood, D. A. Johnston, V. M. Riccard and S. O. Zimmerman: "The use of subchromosome-length unique band sequences in the analysis of prophase chromosomes", Amer. J. Hum. Genetics, 43, pp. 934–947 (1988).