

染色体画像中の部分領域探索手法

阿部 亨^{†1} 濱田 智栄子^{†2} 木下 哲男^{†1}

光学顕微鏡画像を用い、染色体の種々の異常を視覚的に診断する際には、対象画像から個々の染色体領域を抽出し、各領域が対応する染色体番号を同定する作業が必要となる。筆者らは、この作業の自動化を図るために、1本の染色体領域を複数の部分領域（部分バンドパターン）の連なりと見なし、部分領域の探索を対象画像中で順に繰り返すことにより、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。本稿では、提案手法中、染色体の部分領域を探索する処理に関して、これを効果的に実現する方法の検討を行い、染色体画像を対象に部分領域探索実験を行った結果を述べる。

A Subregion Search Method for Chromosome Images

TORU ABE,^{†1} CHIEKO HAMADA^{†2}
and TETSUO KINOSHITA^{†1}

To make a visual examination of various chromosome abnormalities, individual chromosome regions have to be determined in a subject image and classified into the distinct chromosome types. We have proposed a subregion based method to improve this process. The proposed method regards each chromosome region as a series of subregions (local banding patterns), and iterates a search for subregions in the subject image consecutively. In this article, we investigate an effective method to search for subregions in chromosome images, and present the results of subregion search experiments conducted on chromosome images.

^{†1} 東北大学サイバーサイエンスセンター
Cyberscience Center, Tohoku University

^{†2} 東北大学大学院情報科学研究科
Graduate School of Information Sciences, Tohoku University

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同定する手法については、多くの研究がなされており、一部は製品化も行われている。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、染色体の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。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に関して、これを効果的に実現する方法の検討を行い、染色体画像を対象に部分領域探索実験を行った結果を述べる。

2. 染色体画像解析

2.1 染色体画像解析の一般の手順

染色体は、遺伝情報を担う細胞核内の生体物質であり、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、各々は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、図1に示すように、動原体と呼ばれる箇所にくびれており、ここで短腕と長腕に二分される¹⁾。

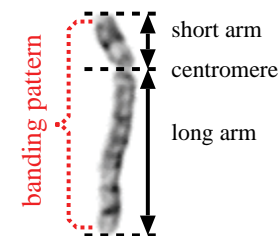


図1 染色体(5番)

光学顕微鏡画像に基づき、染色体の異常を視覚的に診断するには、一般に、以下の手順で作業が進められる²⁾。

- (1) 染色体を染色し光学顕微鏡画像を撮影
- (2) 対象画像中から個々の染色体領域を抽出
- (3) 各領域が何番の染色体に対応するかを同定
- (4) 各領域の状態から染色体の異常を視覚的に診断

染色体の視覚的診断には、図 2 (a) に例を示すように、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。染色には、いくつかの方法があるが、中でも、染料としてギムザを使用する G 分染法が広く用いられている。

通常、まず、対象画像から個々の染色体領域を抽出し、次に、抽出され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。同じ番号の染色体でも、その大きさや輝度値は、細胞周期での時期の違いや染色の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（領域幅は、異なる番号の染色体でも、同じ画像中ではほぼ一定となる）。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、各領域におけるバンドの並び方（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⁴⁾。図 2 (a) 中の各染色体領域を抽出・同定し染色体番号の順に並べた結果を図 2 (b) に示す。

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れば、遺伝情報の欠失（一部のバンドが欠け

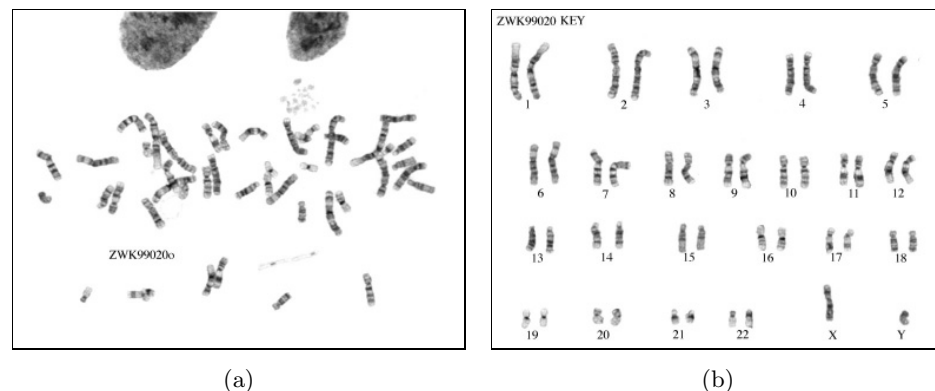


図 2 (a) 染色体画像 (G 分染法), (b) 抽出・同定結果³⁾

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, 染色体の構造異常の視覚的診断が可能となる。

2.2 染色体画像解析のための従来手法

光学顕微鏡画像から個々の染色体領域を抽出し、抽出さ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、対象画像に対する輝度値の 2 値化により、同定とは独立の前処理として行われている⁵⁾⁻¹⁰⁾。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 値化された結果を、バンドパターン等に基づき再度分離・連結することで、個々の染色体領域を決定する手法が提案されている^{11),12)}。しかし、この手法では、前処理 (抽出) の段階で、バンドパターン等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、各領域の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。バンドパターンに基づく手法では、一般に、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) とこれらとの比較により同定が行われる^{6),7)}。その際、正確に抽出できなかった領域や屈曲が大きい領域では、中心線が正しく決定されない場合が多く、バンドパターンを正確に求めることが困難となる。また、染色体領域の大きさや輝度値は画像毎に異なるため、各領域のバンドパターンと参照バンドパターンとの比較が正しく行えるよう、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、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¹³⁾。しかし、これらの正規化手法では、個々の染色体領域の正確な抽出が事前に必要であり、また、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。

さらに、領域抽出の失敗、あるいは、領域の重なり、構造異常などにより、バンドの一部に異常が生じている染色体領域では、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パターンなど) のみを用い同定を行う手法が提案されている^{5),8)}。この手法では、通常とは異なるバンドの影響を抑制す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⁷⁾。

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。

提案手法での処理は、以下の3つに大きく分けられる。

- (1) 部分領域の探索
- (2) 探索範囲の設定
- (3) 染色体領域（部分領域の組み合わせ）の決定

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3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を用いる。以下では、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4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。部分バンドパターン $lbp_m^{(i)}$ 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、隣接する部分バンドパターン $lbp_{m-1}^{(i)}, lbp_{m+1}^{(i)}$ 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。

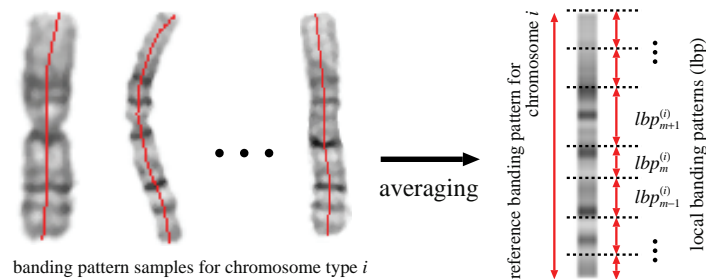


図3 参照バンドパターンと部分バンドパターン (lbp)

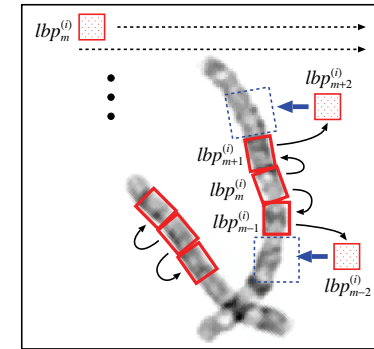


図4 部分領域の探索・探索範囲の設定

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。

- (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、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）実行し、さらに、直前の探索結果を次の探索に利用。

- (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、領域長・輝度値の局所的な変動にも対応。

- (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、高精度な抽出・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、染色体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 をテンプレートとして用い、染色体画像中の部分領域を効果的に探索する手法について述べる。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と 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決定した画像を B_S, B_R で表す。

4.1 部分領域の探索

染色体の部分領域を探索する際には、 I_S 中で領域と判定された各箇所 (B_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_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 5 (a) に示すように、 lbp (長さ U , 幅 V) の原点 $(u, v) = (0, 0)$ が I_S 中の (x, y) に位置する場合、次式により、 lbp を回転させながら各 θ で $e^2(x, y)$ を求め、 $e^2(x, y)$ が最小となる θ での値を (x, y) での MSE とする。

$$e^2(x, y) = \frac{1}{UV} (E_c^2(x, y) + E_b^2(x, y)) \quad (1)$$

$$E_c^2(x, y) = \sum_{(u,v) \in O_c} (\alpha I_S(x', y') + \beta - lbp(u, v))^2 \quad (2)$$

$$E_b^2(x, y) = \sum_{(u,v) \in O_b} (I_S(x', y') - lbp(u, v))^2 \quad (3)$$

$$x' = x + u \cos \theta - v \sin \theta \quad (4)$$

$$y' = y + u \sin \theta + v \cos \theta \quad (5)$$

提案手法では、図 5 (b) に示すように、 lbp の各箇所 (u, v) が、 I_S 中の領域と重なる箇所 O_c であるか、領域と重なる箇所 O_b であるかを区別し (B_S から決定)、各々で別々に得られた輝度値の差の 2 乗和 $E_c^2(x, y), E_b^2(x, y)$ から、式 (1) により $e^2(x, y)$ を求める。その際、染色体領域の輝度値が画像毎 (あるいは箇所毎) に異なる場合の影響を軽減するため、 O_c で

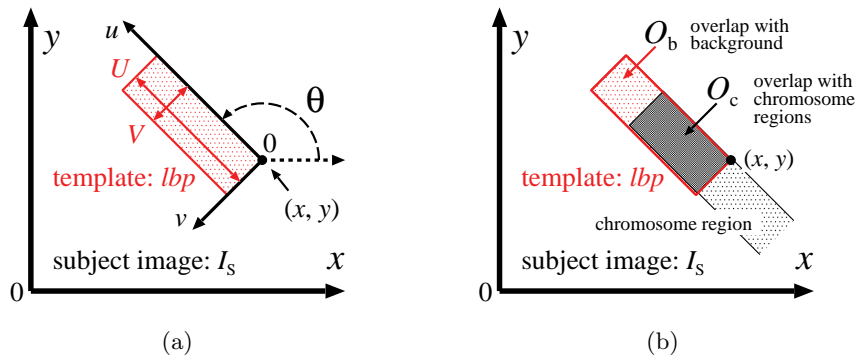


図 5 (a) 対象画像 I_S 中のテンプレート lbp , (b) 領域に重なる lbp の箇所 (O_c) と背景に重なる lbp の箇所 (O_b)

は、式 (2) に示すように、 I_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 α, β を用いた線形変換を行い $E_c^2(x, y)$ の最小化を図る (式 (6), (7) 中、 $|O_c|$ は、 O_c の画素数を表す)。

$$\alpha = \frac{|O_c| \sum_{(u,v) \in O_c} I_S(x', y') lbp(u, v) - \sum_{(u,v) \in O_c} I_S(x', y') \sum_{(u,v) \in O_c} lbp(u, v)}{\left(\sum_{(u,v) \in O_c} I_S(x', y') \right)^2} \quad (6)$$

$$\beta = \frac{1}{|O_c|} \left(\sum_{(u,v) \in O_c} lbp(u, v) - \alpha \sum_{(u,v) \in O_c} I_S(x', y') \right) \quad (7)$$

MSE を求める際、 O_c と同様に O_b でも補正を行うと、 lbp が染色体領域から大きくはみ出して重なっている場合でも 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_b では、式 (3) に示すように、 I_S 中の輝度値の補正を行わない。

4.2 lbp の大きさの補正

提案手法では、 lbp をテンプレートに用い部分領域の探索を行う際、 I_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_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 の大きさを補正する。

まず、 lbp の作成に用いた I_R を 2 値化した B_R と I_S を 2 値化した B_S の各々で、領域の総面積 (総画素数) A_R, A_S と領域幅の最頻値 w_R, w_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 $L_R = A_R/w_R, L_S = A_S/w_S$ で推定できる¹⁴⁾。そこで、提案手法では、 lbp の幅 V として w_S を用い、長さ U として、 lbp の元の長さを L_S/L_R 倍したものを基準値に用いる。さらに、基準値から U を拡大・縮小した lbp を複数用意し、各々で得られた $e^2(x, y)$ の中で最小となる値を (x, y) での MSE とすれば、 L_R, L_S の推定誤りや領域長の局所的変動に対応することも可能になると考えられる。

4.3 lbp の選択

N 枚の参照画像 I_{Rn} ($n = 1, 2, \dots, N$) から参照バンドパターンを作成する際は、まず、各 I_{Rn} 中で個々の染色体領域を抽出し、図 3 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、以下の手順により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と参照バンドパターンを作成する。

- (1) 各 I_{Rn} で推定された染色体領域の長さの総和 L_{Rn} の最大値 L_R^* を決定。
- (2) 各 I_{Rn} で得られたバンドパターンの長さを L_R^*/L_{Rn} 倍する。
- (3) i 番染色体のバンドパターンの長さの平均 $\bar{l}_R^{(i)}$ を求め、各々の長さを $\bar{l}_R^{(i)}$ に揃える。
- (4) 染色体番号毎に、バンドパターンの輝度値の平均を各点で計算。

以上の手順で参照バンドパターンを作成した場合、 I_S 中の染色体領域の長さへ lbp の長さを対応させるためには、 lbp の元の長さを L_S/L_R^* 倍することになる。

参照バンドパターンを分割し作成した lbp は、その箇所・長さにより、部分領域探索に用いた場合の効果が異なる。参照バンドパターン中で特徴的（対応箇所の決定に効果的）なバンドパターンを決定する手法としては、Lockwood らが、対応箇所以外との相関の最大値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¹⁵⁾。しかし、この手法は、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）を対象にしたものであり、さらに、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} で得られたバンドパターンを対象に、各 lbp をテンプレートとして用いた場合に正解箇所（染色体領域中で lbp に対応する箇所）の決定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、全バンドパターンの全箇所 で計算し、正解箇所での MSE の頻度分布 $f_c(e^2)$ と正解箇所以外での頻度分布 $f_i(e^2)$ を求める（MSE の分布を規分布で近似する）。次に、図 6 (a) に示すように、 $f_c(e_b^2) = f_i(e_b^2)$ となる MSE e_b^2 を決定し、次式で得られる F_e を、正解箇所の決定を誤る頻度として用いる。

$$F_e = \int_{e_b^2}^{\infty} f_c(x) dx + \int_{-\infty}^{e_b^2} f_i(x) dx \quad (8)$$

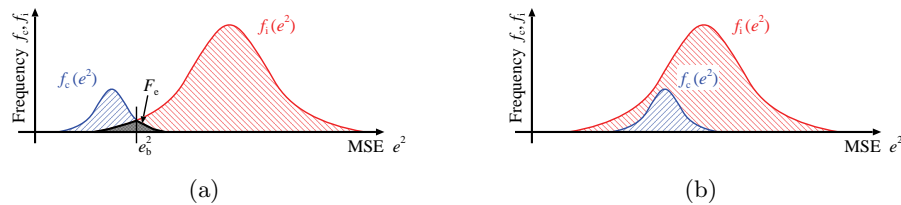


図 6 MSE e^2 の頻度分布

ただし、 $f_c(e^2), f_i(e^2)$ の分布によっては、図 6 (b) に示すように、 e_b^2 が得られないため F_e を決定できない可能性があり、そのような lbp は選択候補から除くことにする。

5. 実験

提案手法に基づき、染色体画像中で部分領域を探索する実験を行った。

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³⁾。このサイトでは、染色体の原画像と共に、その抽出・同定結果の画像も公開しており、今回は、部分領域探索精度の評価（画像中での正解箇所の判定）が

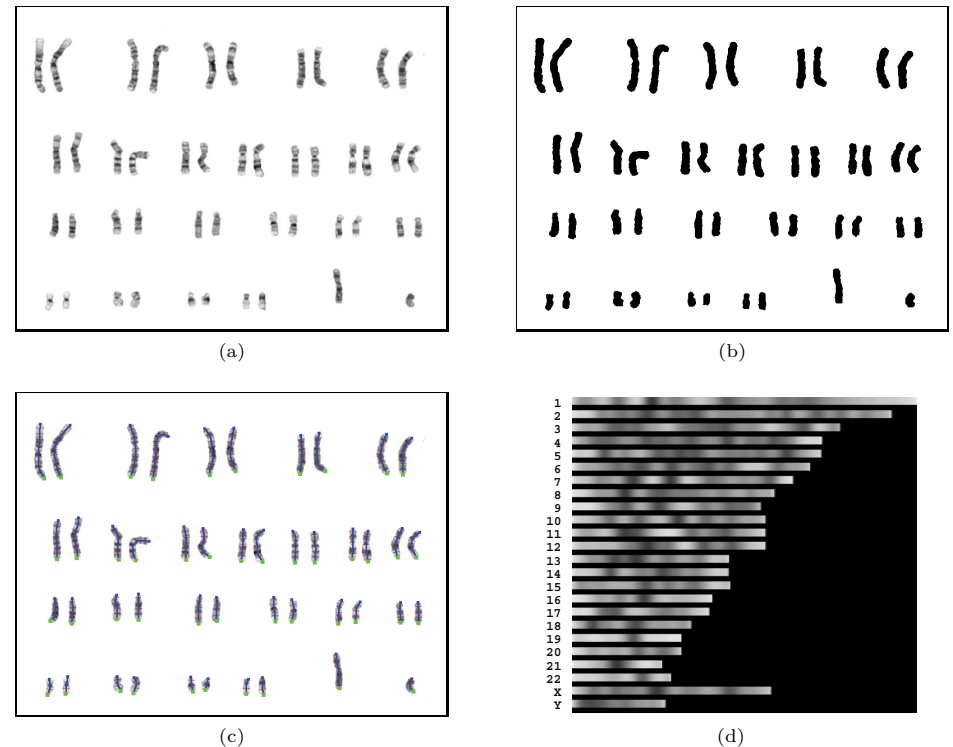


図 7 (a) 実験に用いた染色体画像の例、(b) 2 値化した画像、(c) 各染色体領域で決定された中心線、(d) 実験に用いた参照バンドパターンの例

容易なため、抽出・同定結果の画像 38 枚（数的異常、構造異常を生じている例も含む）を対象に実験を行った。各画像のサイズは 768×576 画素であり、画像中の記号は予め削除した。また、各画像を輝度値で 2 値化した後に孤立点を除去したものを作成し、画像中の領域箇所決定、および、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。さらに、画像中の全染色体領域で中心線を（一部手動で）決定し、参照バンドパターン作成と探索精度評価の際に使用した。

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作成と部分領域探索に交互に用いて実験を行った。なお、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。また、実験に際しては、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の値（参照バンドパターン作成に用いた画像中、領域箇所での輝度値の平均と分散の平均）となるよう、全ての画像に対し、領域箇所の輝度値の正規化を行った。

図 7 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

部分領域探索結果の正誤判定を行う際には、まず、テンプレートとして探索に用いた lbp に対応する正解箇所を染色体領域の中心線上で決定する。図 8 に示すように、 lbp に対応する中心線の始点と終点を結ぶベクトルを $v_m = pt_m - pi_m$ 、探索結果（部分領域）の中心線の始点と終点を結ぶベクトルを $v_i = pt_i - pi_i$ とすると、以下の条件を満たす探索結果を正解（探索成功）とし、それ以外の場合を誤り（探索失敗）と見なす。

- 対応する中心線が 60%以上 lbp で覆われ、かつ、
- $v_i v_m > 0$ (v_i, v_m のなす角が 90° 以下)

また、部分領域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

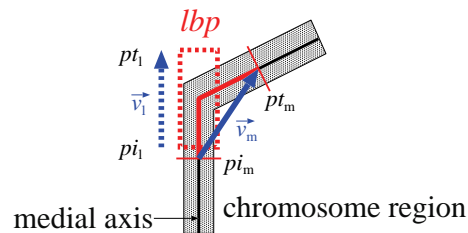


図 8 部分領域探索結果の正誤判定

(Precision) を用いた。

$$R = |C \cap D| / |C| \quad (9)$$

$$P = |C \cap D| / |D| \quad (10)$$

ここで、 C は、探索対象画像 I_S 中での正解箇所の集合を、 D は、探索結果として I_S 中で決定された箇所の集合を、 $D \cap C$ 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。従って、 R は、全正解箇所に対する探索成功箇所の割合を、 P は、全決定箇所に対する探索成功箇所の割合を表すことになる。

5.3 実験結果

4.1 節で述べた提案手法と比較するため、対象画像 I_S 中での部分領域探索には以下の値を用いた。

- I_S 中の各箇所と lbp との相関
- I_S 中の各箇所と lbp との MSE (輝度値補正なし)
- I_S 中の各箇所と lbp との MSE (O_c, O_b で輝度値補正)
- I_S 中の各箇所と lbp との MSE (O_c でのみ輝度値補正, 提案手法)

実験では、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、テンプレートとして探索に用いる lbp を以下の手順で選択した。

- 各 lbp について、全参照バンドパターン中の全箇所に対する MSE を求め、
- 各 lbp について、対応箇所以外での MSE の最小値を決定し、
- 各染色体番号に対し、決定された最小値が大きい順に 3 本の lbp を選び実験に使用。参照バンドパターンと lbp との MSE を求める際には、(b)~(d) で述べた手法を用いた ((c), (d) では、同じ lbp が選択されることになる)。また、各 I_S での探索結果を lbp 毎に MSE の昇順に並べ、上位 k 以内の決定箇所を式 (9), (10) 中の D に、上位 k 以内の正解箇所を $C \cap D$ に用いた。各 I_S に対して、染色体番号毎に 3 本の lbp を用い探索を行うため、各手法での $|C|$ は、延べ $(878 + 889) \times 3 = 5301$ となり、 $|D|$ は、延べ $38 \times 24 \times 3 \times k = 2736k$ となる。なお、(a) では、 lbp と参照バンドパターンとの相関を求め、対応箇所以外での相関の最大値が小さい順に 3 本の lbp を選択し、探索結果は相関の降順に並べた。

順位 k を変えて D を変化させ、 R の平均と P の平均を横軸と縦軸に各々プロットしたグラフを図 9 に示す。このグラフから、(a) に比べ、(b), (c), (d) では、 R, P が共に向上していることが分る。これは、(a) では、 lbp が染色体領域から大きくはみ出して重なっている箇所でも相関が大きくなり、対応箇所と誤って決定される例が多かったためである。また、(b) に比べ、(c), (d) で R, P が共に向上していることから、 I_S 中の輝度値を部分領域単位

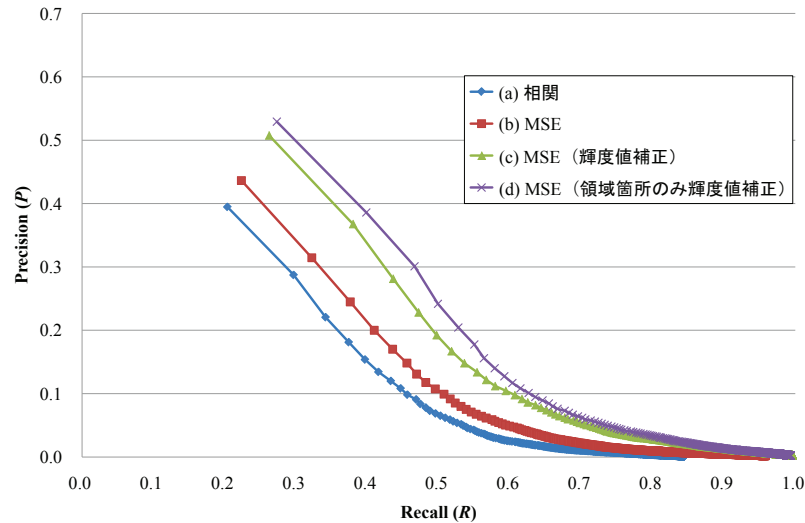


図 9 順位 k を変化した場合の再現率 R と適合率 P (部分領域探索手法の比較)

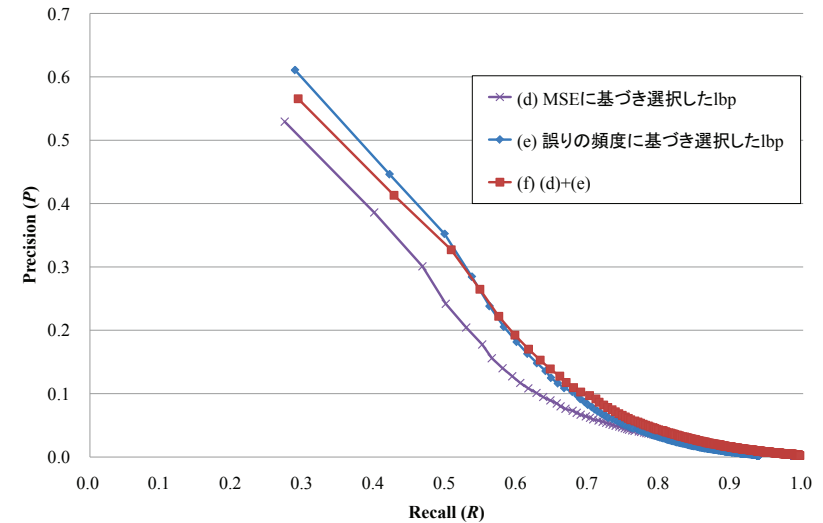


図 10 順位 k を変化した場合の再現率 R と適合率 P (lbp 選択手法の比較)

で補正した効果が確認できる。さらに、(c) よりも高い R, P が (d) で得られており、このことは、 I_S 中の領域箇所でのみ輝度値の補正を行う提案手法の有効性を示すと考えられる。

次に、4.3 で述べた提案手法を用い、誤りの頻度 F_e に基づき選択された lbp により部分領域の探索を行った結果を図 10 に示す。図 10 中の (d) は、先の実験での (d) の結果と同じものであり、(e) は、各染色体番号に対し F_e が小さい順に 3 本の lbp を選択した場合の結果である。本来は、各 I_S に対して、染色体番号毎に 3 本の lbp を用いるため、延べ $38 \times 24 \times 3 = 2736$ 本の lbp が必要となる。しかし、短い染色体 (20, 21, 22, Y) に対しては、4.3 で述べた理由から F_e を決定できず、3 本の lbp を得られない場合があった (参照画像にセット A を用いた場合は 1297 本、セット B を用いた場合は 1199 本の lbp が得られた)。従って、(e) は、延べ $1297 + 1199 = 2496$ 本の lbp のみを用い探索を行った結果である ($|D|$ は延べ $2496k$ となる)。図 10 中の (f) は、(e) で不足した $2736 - 2496 = 240$ 本の lbp を、(d) で決定された lbp により重複しないよう補完した場合の結果である。

この結果から、(e), (f) では、(d) に比べ R, P が共に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と考えられる。また、(e) と

(f) を比較すると、 $R < 0.55$ では、(e) の方で大きな P の値が得られており、 $R > 0.55$ では、(f) の方が大きな P の値を示している。これは、(e) では、用いた lbp の延べ本数が (f) に比べ少ないため、 k を大きく設定した場合、 R は余り増加せずに P のみが低下するためである。なお、(e) で得られた部分領域探索結果の例を図 11 に示す。

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、および、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、および、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 2 (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

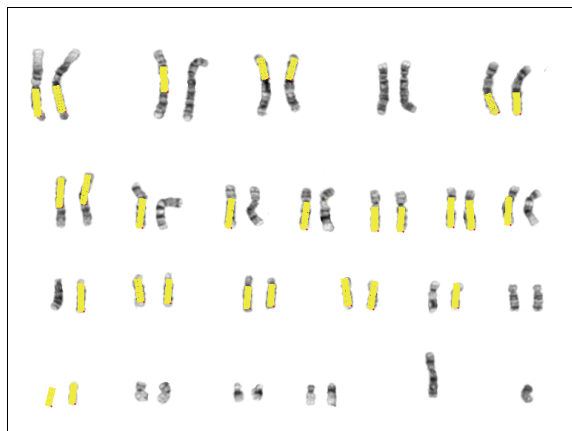


図 11 部分領域探索結果の例 (1,2 位として得られた正解箇所)

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。

謝辞 本研究を進めるにあたり、御討論頂きました(株)日本遺伝子研究所 島田 裕充 氏、日本アイ・ピー・エム(株)大庭 信之 博士に感謝致します。

参 考 文 献

- 1) Gersen, S.L. and Keagle, M.B.(eds.): *The principles of clinical cytogenetics*, Humana Press, 2nd edition (2004).
- 2) Graham, J. and Piper, J.: Automatic karyotype analysis, *Meth. Mol. Biol.*, Vol.29, pp.141–185 (1994).
- 3) ZooWeb — Karyotypes home page, <http://worms.zoology.wisc.edu/zooweb/Phelps/karyotype.html>.
- 4) Harnden, D.G. and Klinger, H.P.(eds.): *ISCN1985: An international system for human cytogenetic nomenclature (1985)*, S. Karger AG. (1985).
- 5) Groen, F. C.A., Kate, T.K., Smeulders, A. W.M. and Young, I.T.: Human chromosome classification based on local band descriptors, *Pattern Recognit. Lett.*, Vol.9, No.3, pp.211–222 (1989).
- 6) Ritter, G. and Pesch, C.: Polarity-free automatic classification of chromosomes, *Computational Statistics Data Analysis*, Vol.35, pp.351–372 (2001).
- 7) Wu, Q., Liu, Z., Chen, T., Xiong, Z. and Castleman, K.R.: Subspace-based prototyping and classification of chromosome images, *IEEE Trans. Pattern Anal. Machine Intell.*, Vol.14, No.9, pp.1277–1287 (2005).
- 8) Moradi, M. and Setarehdan, S.K.: New features for automatic classification of human chromosomes: A feasibility study, *Pattern Recognit. Lett.*, Vol.27, No.1, pp. 19–28 (2006).
- 9) hong Kao, J., hui Chuang, J. and Wang, T.: Chromosome classification based on the band profile similarity along approximate medial axis, *Pattern Recognit.*, Vol.41, No.1, pp.77–89 (2008).
- 10) Legrand, B., Chang, C., Ong, S., Neo, S.-Y. and Palanisamy, N.: Chromosome classification using dynamic time warping, *Pattern Recognit. Lett.*, Vol.29, No.3, pp.215–222 (2008).
- 11) Charters, G.C. and Graham, J.: Trainable grey-level models for disentangling overlapping chromosomes, *Pattern Recognit.*, Vol.32, No.8, pp.1335–1349 (1999).
- 12) Schwartzkopf, W., Bovik, A. and Evans, B.: Maximum-likelihood techniques for joint segmentation-classification of multispectral chromosome images, *IEEE Trans. Med. Imag.*, Vol.24, No.12, pp.1593–1610 (2005).
- 13) Piper, J. and Granum, E.: On fully automatic feature measurement for banded chromosome classification, *Cytometry*, Vol.10, pp.242–255 (1989).
- 14) 阿部 亨, 濱田智栄子, 木下哲男: 染色体画像解析のための領域長正規化手法, 信学論 D, Vol.J91-D, pp.2983–2987 (2008).
- 15) Lockwood, D.H., Johnston, D.A., Riccard, V.M. and Zimmerman, S.O.: The use of subchromosome-length unique band sequences in the analysis of prophase chromosomes, *Amer. J. Hum. Genetics*, Vol.43, pp.934–947 (1988).