Chemometrics and Intelligent Laboratory Systems 111 (2012) 46–52

Contents lists available at SciVerse ScienceDirect

Chemometrics and Intelligent Laboratory Systems

journal homepage: www.elsevier.com/locate/chemolab

Automated pairing of human chromosomes applying gradient profile and similarity matching algorithm

Xinwei Feng, Peisheng Cong *, Zhongliang Zhu, Xinyuan Du

Department of Chemistry, Tongji University, Siping Road 1239#, Shanghai, 200092, PR China

ARTICLE INFO

Article history:
Received 14 June 2011
Received in revised form 26 September 2011
Accepted 17 November 2011
Available online 25 November 2011

Keywords:
Chromosome pairing
Image processing
Gradient profile
Cosine value
Similarity matching

ABSTRACT

One of the most important tasks in cytogenetics is chromosome analysis, with which cytogeneticists can diagnose whether there are structural or numerical chromosome aberrations. In this paper, a feasible way to deal with computer-aided pairing of metaphase chromosomes is presented. From pretreating images to pairing chromosomes, feature extraction and pairing methods are described in detail. As an important feature, density profiles are extracted from banding patterns and smoothed by both median and mean filters. In these cases, density profiles are adjusted by the method proposed in this paper. Gradient profiling, a new feature attained by transforming the density profile along with the length and centromeric index, is used in our chromosome pairing procedure. Chromosomes are divided into 23 pairs by comparing their similarity, which is numeralized as the cosine value of gradient profiles. Compared with density profiling, gradient profiling improves matching accuracy by 12.73%. Furthermore, classification accuracy of chromosomes in male and female cells is 91.3% and 100% respectively. Experimental results indicate that the proposed gradient profiling method can serve for chromosomes pairing.

1. Introduction

Chromosome image identification can help cytogeneticists with clinical diagnostic procedures [1]. Since Tjio and Levan [2] discovered that a somatic cell contains 46 chromosomes, it has been found that many genetic abnormalities are directly linked to structural or numerical aberrations chromosomes [3]. In conventional cytogenetics analysis and diagnostics, chromosome identification is usually accomplished manually, which is tedious, time-consuming and inefficient [4,5]. Consequently, computer-aided human chromosome image analysis systems were developed to facilitate clinical diagnosis [6].

Chromosome identification is usually conducted on a chromosome spread photomicrograph. In the four stages of the cell cycle (prophase, metaphase, anaphase and telophase), only in late prophase or metaphase are chromosomal structures visible under a light microscope after Giemsa staining. For clinical purposes, chromosomes are divided into 24 classes (22 autosomes and 2 sex chromosomes). Chromosomes belonging to the same class, known as homologues, are nearly identical. These 24 classes can be further subdivided into seven groups, A~G. Chromosomes in the same group have similar length and centromeric index (the ratio between short arm length and the whole length of a chromosome). After grouping, chromosomes are arranged by karyotype, in which a systematized array of metaphase chromosomes from a photomicrograph of a single cell nucleus is presented in descending order by length. A typical Giemsa-stained metaphase cell spread image and a karyotype is shown in Fig. 1.

The primary objective of the new method introduced by this paper is chromosome pairing. As an important aspect of chromosome identification, chromosome pairing methods are being investigated. Our goal is to develop a computer-aided way to help professionals with cytogenetics analysis and diagnostics. Like other pattern recognition systems, the success of computer-aided chromosome analysis depends on robust feature extraction [7,8]. Because they are effective in discriminating different groups of chromosomes, length and centromeric index classification are common in the literature [9]. In the past few years, banding patterns have been proposed as another effective characteristic by which to identify chromosomes. Giemsa-stained chromosomes can be classified by the number of bands, band position and band intensity distribution along the medial axis. Therefore, theoretically, it may be possible to use banding patterns to pair chromosomes. However, in actual chromosome image analysis, the variability of chromosome morphology, e.g., various kinds of bending [10] or overlapping, makes the usage of banding pattern elusive. Feature extraction methods for banding patterns is an active area of research [11]. Density profiling, which was introduced by Piper and Granum [12], integrates the intensities along sections perpendicular to the medial axis. Density profiles have been used by many people to classify chromosomes [6,13–16]. The shape profile is another commonly used feature. Garcia et al. [17] extracted characteristic vectors from the curvature of each chromosome. Wang et al. [18] investigated a multi-stage rule-based computer scheme to

* Corresponding author.
E-mail addresses: pshcong@tongji.edu.cn, mysteryfxw@126.com (P. Cong).

0169-7439/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.chemolab.2011.11.009

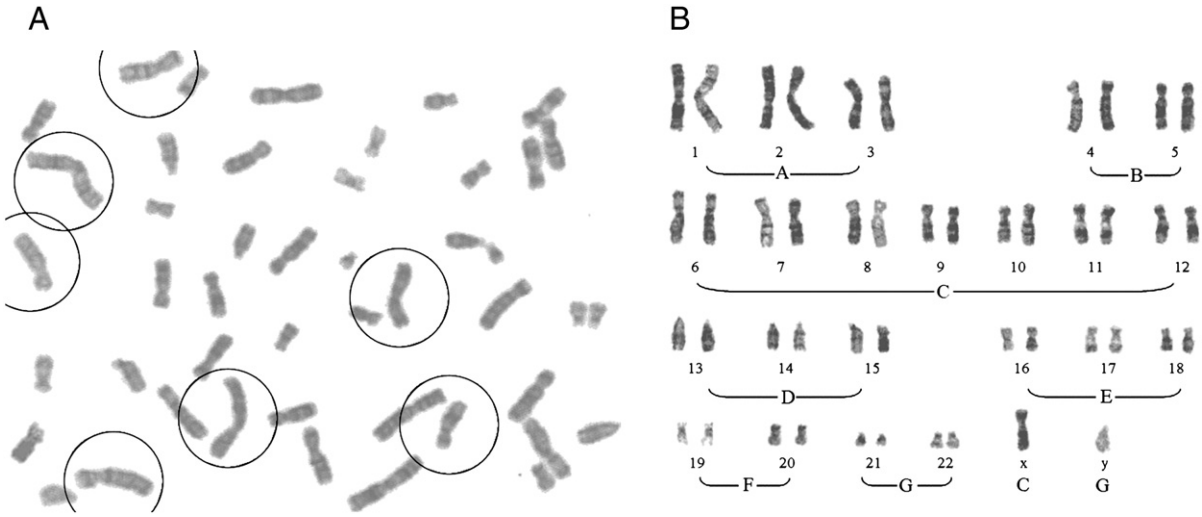

Fig. 1. (A) A Giemsa-stained metaphase cell spread image. (B) A karyotype.

automatically detect centromeres and determine polarities for both abnormal and normal metaphase chromosomes. Kao et al. [19] proposed efficient approaches for medial axis determination and profile matching of human chromosomes without identifying centromeres. In our study, we collected and adjusted density profiles to determine a gradient profile. We found that gradient profiles are more sensitive to the dark and bright patterns of chromosomes, which is a key in discriminating different classes.

In chromosome pairing, Khmelinskii et al. [20] used a supervised linear classifier together with a combinatorial optimization algorithm to compute the pairing of chromosome images. A kernel nearest-neighbor (K-NN) algorithm was proposed by Yu et al. [21] to accomplish automatic karyotyping. This method offers an alternative solution for Support Vector Machines (SVM) based classification [14] by mapping data onto a high dimensional feature space. Wu et al. [22] developed globally optimal algorithms for automated classification and pairing of human chromosomes. Sampat [23] described a fully automatic chromosome classification algorithm for multiplex fluorescence in situ hybridization (M-FISH) images using supervised parametric and non-parametric techniques. Martínez [24] presented a method for chromosome classification based on an almost unexplored neural network technique. Wang et al. [25,26] proposed a two-layer classification platform within artificial neural networks (ANN) which has higher and more robust performance. Different from these methods, we use similarity comparison because it is simpler than the aforementioned methods and suitable for small datasets.

In this paper, we discuss our research on the chromosome pairing method. We focus on images taken by ordinary CCD cameras. In Section 2, we discuss the theory of behind digital chromosome images. Density profiles, together with lengths and centromeric indices, are extracted from preprocessed images. Profiles of chromosomes are critical features in chromosome identification, but they appear coarse when first obtained. If the profiles are not well treated, they could become obstacles for later processing. Therefore, we employ both median and mean filters to smooth the profiles. Furthermore, some situations met in the process (e.g., drifting of peaks or valleys in profiles from the same class, which are caused mainly by bent chromosomes) would be resolved, and profiles would be adjusted by the proposed methods. After that, density profiles are transformed into gradient profiles, which is a better representation of banding patterns in chromosomes when calculating similarity. Next, we turn to the problem of homologue pairing. The cosine values of profiles are utilized to express the similarity of chromosomes. Finally, the experimental results and analyses are discussed.

2. Methodology

2.1. Image preprocessing

Chromosome images must be preprocessed before feature extraction. Chromosomes distribute to different directions on the image; sometimes, they can be bent, overlapped or touching (Fig. 1 shows bent chromosomes, which are circled). To optimize our results, metaphase spreads with a total of 46 isolatable chromosomes were investigated in this study.

First, we segmented chromosomes from the image background [27–29] (Fig. 2). Touching chromosomes were clipped after segmentation, and then, medial axes were extracted from all chromosomes. The medial axis of a chromosome was attained by applying a medial axis transformation (MAT) [30]. Using this transformation and a downsampling algorithm, boundary points of a region could be iteratively deleted to form a skeleton line. The medial axis was then computed by extending the skeleton line to the boundary of a chromosome. According to the medial axis, the chromosomes were straightened (Fig. 2) and normalized according to the following steps. First, the longest chromosome (usually Class 1) was normalized to 109 pixels; the other chromosomes were compressed concomitantly. We chose 109-pixels as the criterion because it is the average length of

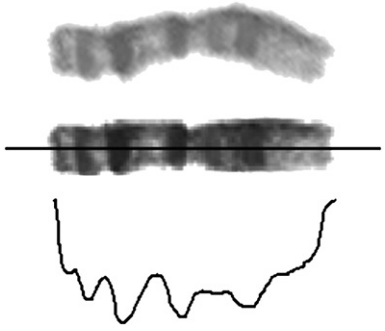

Fig. 2. A chromosome with its medial axis and density profile.