Missouri University of Science and Technology
Scholars' Mine

Electrical and Computer Engineering Faculty Research & Creative Works
Electrical and Computer Engineering

01 Jun 1998

Data-Driven Homologue Matching for Chromosome Identification

R. Joe Stanley
Missouri University of Science and Technology, stanleyj@mst.edu

James M. Keller

Paul D. Gader

Charles William Caldwell

Follow this and additional works at: https://scholarsmine.mst.edu/ele_comeng_facwork

Part of the Electrical and Computer Engineering Commons

Recommended Citation

R. J. Stanley et al., "Data-Driven Homologue Matching for Chromosome Identification," *IEEE Transactions on Medical Imaging*, vol. 17, no. 3, pp. 451-462, Institute of Electrical and Electronics Engineers (IEEE), Jun 1998.
The definitive version is available at https://doi.org/10.1109/42.712134

This Article - Journal is brought to you for free and open access by Scholars' Mine. It has been accepted for inclusion in Electrical and Computer Engineering Faculty Research & Creative Works by an authorized administrator of Scholars' Mine. This work is protected by U. S. Copyright Law. Unauthorized use including reproduction for redistribution requires the permission of the copyright holder. For more information, please contact scholarsmine@mst.edu.

Data-Driven Homologue Matching for Chromosome Identification

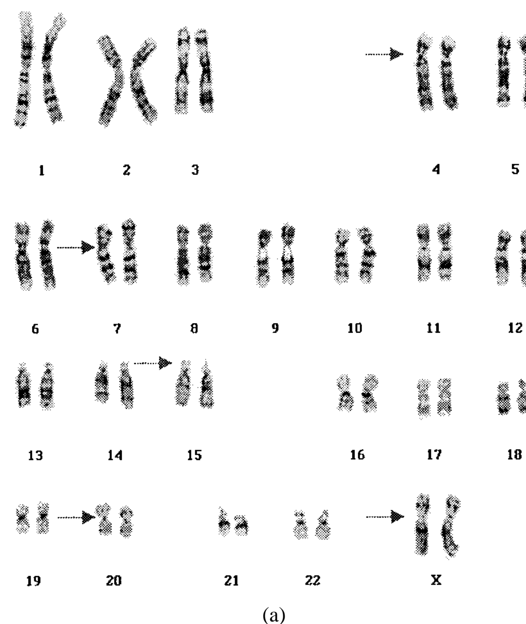
Ronald J. Stanley, James M. Keller,* *Senior Member, IEEE*, Paul Gader, *Member, IEEE*, and Charles W. Caldwell

***Abstract*— Karyotyping involves the visualization and classification of chromosomes into standard classes. In "normal" human metaphase spreads, chromosomes occur in homologous pairs for the autosomal classes 1–22, and X chromosome for females. Many existing approaches for performing automated human chromosome image analysis presuppose cell normalcy, containing 46 chromosomes within a metaphase spread with two chromosomes per class. This is an acceptable assumption for routine automated chromosome image analysis. However, many genetic abnormalities are directly linked to structural or numerical aberrations of chromosomes within the metaphase spread. Thus, two chromosomes per class cannot be assumed for anomaly analysis. This paper presents the development of image analysis techniques which are extendible to detecting numerical aberrations evolving from structural abnormalities. Specifically, an approach to identifying "normal" chromosomes from selected class(es) within a metaphase spread is presented. Chromosome assignment to a specific class is initially based on neural networks, followed by banding pattern and centromeric index criteria checking, and concluding with homologue matching. Experimental results are presented comparing neural networks as the sole classifier to our homologue matcher for identifying class 17 within normal and abnormal metaphase spreads.**

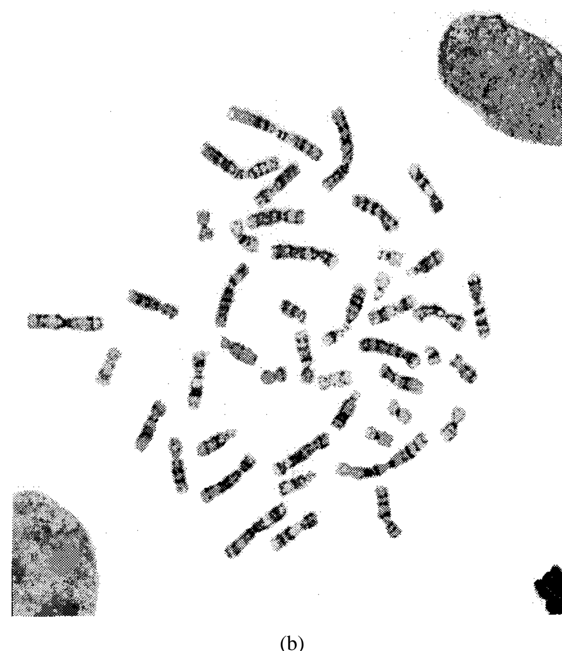
***Index Terms*— Chromosomes, dynamic programming, homologue, karyotyping, neural networks.**

I. Introduction

KARYOTYPING refers to the visualization and classification of chromosomes found in metaphase spreads. Fig. 1 contains a normal female metaphase spread and its corresponding karyotype. Visually, the chromosome pairs, known as homologues, appear similar. Normal cells contain homologues for chromosome classes 1–22, the autosomes, and the sex chromosome X for a female or paired X and Y sex chromosomes for a male. In the karyotype above, the homologues visually appear similar enough to one another as well as close enough to "typically" normal for the cell to be considered normal. Thus, there are two criteria in evaluating homologues in assessing normalcy. First, homologues must be similar to one another within the context of the cell. Second, homologues must be within the expected range of variation for normal chromosomes of the homologue class.



(a)



(b)

Fig. 1. Metaphase spread/karyotype image pair: (a) metaphase spread image and (b) corresponding karyotype image. The arrows indicate the centromere location for selected chromosomes.

Manuscript received June 11, 1997; revised January 8, 1998. This work was supported by the National Library of Medicine-Cancer Affiliate under training Grant 5 T15 LM/CA07089-04 and by a grant from the University of Missouri Research Board. The Associate Editor responsible for coordinating the review of this paper and recommending its publication was M. W. Vannier. *Asterisk indicates corresponding author.*

R. J. Stanley is with the University of Missouri, Department of Health Management and Informatics, Department of Computer Engineering and Computer Science, Columbia, MO 65211 USA.

*J. M. Keller is with the University of Missouri, Department of Computer Engineering and Computer Science, 217 Engineering Building West, Columbia, MO 65211 USA (e-mail: keller@cecs.missouri.edu).

P. Gader is with the University of Missouri, Department of Computer Engineering and Computer Science, Columbia, MO 65211 USA.

C. W. Caldwell is with the Ellis Fischel Cancer Center, Department of Pathology and Anatomical Sciences, Columbia, MO 65203 USA.

Publisher Item Identifier S 0278-0062(98)06447-7.

0278–0062/98$10.00 © 1998 IEEE

Many existing approaches for performing automated chromosome image analysis presuppose a fixed number of chromosomes per class, two, and 46 chromosomes within a metaphase spread for achieving better classification [1]–[5], which is true for normal cells. For routine automated chromosome image analysis, this may be a reasonable assumption. However, two chromosomes per class cannot be assumed for abnormality analysis. Many genetic anomalies are directly linked to structural or numerical aberrations in the chromosome complement. The focus of this paper is the development of image analysis techniques that are directly applicable to evaluating numerical aberrations evolving from structural abnormalities. Specifically, a method is presented to identify chromosomes from selected class(es) within a metaphase spread. The method capitalizes on the principle of least commitment [6], [7] and avoids the two-chromosome-per-class assumption. Chromosome assignment incorporates neural networks, banding pattern and centromeric index criteria checking, and homologue matching. The basic technique is to use as much information as available to find homologous pairs. For the selected class, the best representative or primary chromosome is found within the metaphase spread. Homologue candidates are obtained using simple criteria. The candidates are matched to the primary chromosome for homologue determination. The homologue found is rematched using the same process. With the purpose of aiding a cytogenetic expert, making no decision for chromosome assignment is better than an incorrect assignment.

II. Background

Chromosome misassignment limits the clinical application of automated karyotyping. There is an explicit need to avoid false-positive's (FP's), assigning chromosomes to a specific class which do not belong to that class. Minimizing the FP rate directly contributes to reducing the false-negative (FN) rate for abnormality assessment. Here, a FN refers to a patient determined free of a particular condition when the patient has that condition.

Several genetic conditions are directly linked to specific chromosomal aberrations. For example, acute promyelocytic leukemia has been associated with distortions in one chromosome 15 and one chromosome 17, leaving one "normal" chromosome 15 and one "normal" chromosome 17. Incorrectly classifying chromosomes as "normal" class 15, class 17, or not assigning "normal" chromosomes to those classes that belong to those classes is the difference between appropriately diagnosing a patient as potentially possessing acute promyelocytic leukemia or missing the diagnosis. Thus, the technique employed to identify chromosomes from the selected class must have high confidence that the chromosomes assigned to that selected class belong to that class. Furthermore, not only do the chromosomes have to be correctly assigned, but the number of chromosomes determined for a particular class within a metaphase spread must be correct.

A. Neural Networks and the Transportation Algorithm

Neural networks have been shown to classify individual chromosomes with high recognition rates [3], [8]–[10]. Automated chromosome classification using neural networks has achieved isolated chromosome-recognition rates as high as 95.6% for the Copenhagen dataset, which is considered one of the benchmark datasets for chromosome analysis [3], [10]. However, all 46 chromosomes within a single cell are correctly classified in only 12.6% of the cells. The features used in those studies include: normalized area, normalized length, normalized convex hull perimeter, centromeric index, weighted-density distribution's (WDD's), standardized density profiles, and others [3], [9]. The strength of using neural networks is that they are capable of generalizing the classification function from the data quite well. Neural networks used for chromosome classification treat each chromosome encountered as an independent event. neural-network misclassification at the cell level leads to incorrect chromosome assignment or to an incorrect number of chromosomes assigned for the respective class.

Our technique addresses the problems associated with neural-network misclassification by using the chromosome with the greatest margin of neural-network confidence victory in the selected class; satisfying additional banding pattern and centromeric index criteria as the prototype for the selected class. The prototype chromosome is matched to candidate chromosomes within the metaphase spread for homologue assignment. Finding other chromosome(s) from the selected class uses the context of the cell and the fact that chromosomes usually occur in similar pairs.

In addition to the misclassification case, the two-chromosomes-per-class assumption may lead to an incorrect number of chromosomes assigned to a selected class. Suppose that a metaphase spread has one "normal" chromosome from a selected class and that the neural network finds two winners for the selected class. Assume that misclassification has occurred with the assignment of the second chromosome to the selected class. Further, suppose that the patient for which the metaphase spread is analyzed has a type of leukemia that results in a wrong number of "normal" chromosomes from the selected class. In this case, the incorrect chromosome assignment results in improperly assessing the patient's condition. Our method attempts to overcome this case with homologue matching to the best representative of the chromosome for the selected class found within the metaphase spread. Thus, this technique has the capability to analyze multiple neural-network winners for potential homologue matching.

The transportation algorithm has demonstrated chromosome classification optimization in other studies [11], [12]. The transportation algorithm typically utilizes the two-chromosomes-per-class constraint for classes 1–22, two class-X chromosomes, and one class-Y chromosome. Utilizing surplus and slack variables, the transportation algorithm can be used to accommodate for a variable number of chromosomes for a given class. The transportation algorithm was used as the benchmark for comparison in this study based on optimizing feedforward neural-network confidence values for chromosome assignment.

B. Chromosome Feature Delimiters

There are many chromosome features that have shown capability of identifying and delimiting chromosomes. Most

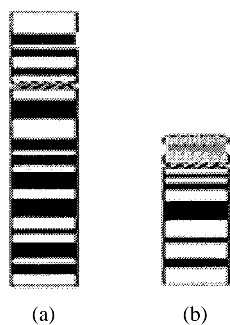


Fig. 2. Ideograms for classes 4 and 15 at the 400-band level: (a) class 4 and (b) class 15.

chromosomes can be classified based on their relative length, centromeric index, and banding pattern in the context of the cell. From Fig. 1 notice that chromosome length gradually declines following from class 1–22. With subtle variations among cells, such as preparation technique, size, and chromosome morphology, many classifiers use that information via normalized-chromosome length for the cell.

Besides length, the centromeric index provides a significant amount of chromosome delimiting capability. The centromeric index is commonly defined as the ratio of the chromosome's short arm (p-arm) to the total chromosome length. The centromere is usually located in a constricted region along the chromosome contour. Fig. 1 illustrates the unique positions of centromeres for chromosome classes 4, 7, 15, 20, and X (black arrows). Classes 13, 14, 15, 18, 21, 22, and Y contain acrocentric centromeres near the top of the chromosome's short arm. The other chromosome classes, where the centromere is closer to the middle of the chromosome, are referred to as metacentric. The centromere locations, which may occur at unique points along the chromosome, also provides delimiting capability. Thus, knowing the centromeric index delimits the set of classes to which the chromosome belongs.

The banding pattern provides important information for chromosome classification. Each of the 24 classes possesses unique banding patterns. Fig. 2 presents an idealized banding pattern for two different classes at the 400-band level, which refers to the approximate composite number of bands found over all chromosomes within a single metaphase spread. Common band levels include 300, 400, 550, 800, and 2000. The band level often indicates the degree of condensation and length of chromosomes found within a metaphase spread and, therefore, affects certain features, such as length and banding pattern.

Unique banding patterns for each class provide for several band features which are useful for chromosome-classification purposes. One of the simplest is the density profile, the mean grey level along perpendicular lines to the medial axis of the chromosome, providing a grey-level representation of the banding pattern. Due to variations in preparation technique, sources of the cells, and image enhancement techniques accompanying image acquisition, it is difficult to consistently obtain a generalizable banding pattern for grey-level chromosome images. Fig. 3 shows several chromosome

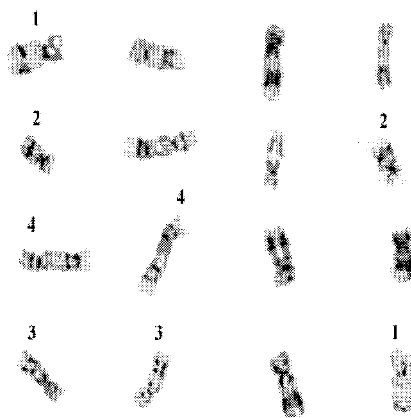


Fig. 3. Image of chromosome 17's from various cells. Chromosomes labeled 1, 2, 3, and 4 are homologues.

17's in a composite image, illustrating the variation in size and banding pattern for those chromosomes. Incorporating this knowledge, Zimmerman *et al.* performed chromosome classification primarily relying on the number of bands [13].

C. Homologue Matching and Markov Models

Our approach to chromosome classification incorporates the knowledge that there should be 22 pairs of homologues and the sex chromosomes. Instead of 46 chromosomes, classification is simplified to assigning 22 pairs plus the X and Y chromosomes to the appropriate classes. Zimmerman *et al.* established the premise of matching chromosomes as homologous pairs for classification purposes [13]. Their procedure does not directly extend to classification for numerical aberrations because chromosomes are homologously paired for karyotyping.

Performing chromosome classification in the context of a single cell is important for analyzing many chromosome features. Chromosome banding pattern comparison within a cell is an approach to compensate for the variability between cells. Gregor and Granum investigated chromosome analysis incorporating Markov networks [14]. Dynamic programming in the framework of the Markov model has been applied to chromosome analysis [15].

Although Markov models have been used with relative success, there are several difficulties. First, Markov models are complex. There are many parameters which must be incorporated into the model, such as transition probabilities and exclusionary and inclusionary substring probabilities to specific classes. These parameters detract from the direct comparison, which is made between chromosomes within the same cell. Second, substrings are sought out through comparing strings from the same class. Again, the Markov process circumvents direct cell-specific analysis of chromosomes, which is the context wherein anomalies should be assessed.

Our approach for incorporating dynamic programming uses chromosomes with high confidence for belonging to the specific class within the cell of interest as the benchmark for further comparisons. This accounts for the natural variation of chromosomes from cell to cell that are considered normal.

```
Find chromosomes within metaphase spread through segmentation process
Find medial axis representation for all chromosomes within metaphase spread
Compute features for all isolated chromosomes within metaphase spread
Determine candidate pool for selected class
Eliminate candidates based on banding pattern and centromeric index criteria
IF candidates remain
        THEN  Determine primary chromosome for selected class
              Scale profile(s) of primary and candidate chromosomes
              FOR remaining candidates
                      Use dynamic programming to match the primary chromosome to the candidate chromosomes
              Form confidence distribution for candidates from dynamic programming matching scores
              FOR remaining candidates
                      Compute distance measure for size features and centromeric index between primary chromosome and
                      candidate
              Form confidence distribution based on distance measure for candidates
              FOR remaining candidates
                      Compute the final confidence
              Take the highest confidence candidate as the homologue
              Find neural network winning class for homologue
              Determine candidates based on homologue winning class and primary chromosome
              Eliminate candidates based on size feature and centromeric index criteria
              FOR remaining homologue candidates
                      Use dynamic programming to match the homologue chromosome to the candidate chromosomes
              Form confidence distribution for candidates from dynamic programming matching scores
              FOR remaining homologue candidates
                      Compute distance measure for size features and centromeric index between homologue
                      chromosome and candidate
              Form confidence distribution based on distance measure for candidates
              FOR remaining homologue candidates
                    Compute the final confidence
              Take the highest confidence candidate as the matching chromosome
              IF matching chromosome is primary chromosome
                      THEN assign primary and homologue chromosomes to selected class
                      ELSE
                              assign primary chromosome to selected class
        END IF
ELSE
        no chromosomes assigned to selected class
```

Fig. 4. Algorithm for homologue matching.

III. Homologue-Matching Algorithm

The goal for the homologue-matching algorithm presented is to make chromosome assignments with high confidence of correctness. No assignment is considered better than an incorrect assignment because FP classifications can lead directly to FN diagnostic assessments for genetic conditions relating to specific chromosomal anomalies. The algorithm for determining homologues from metaphase spread images relating to isolated chromosomes is contained in Fig. 4 and is described in the following sections.

A. Image Segmentation and Skeletonization

The feature extraction process utilizes three program inputs. The inputs are: 1) the original metaphase spread image, 2) the segmented image of the metaphase spread image, and 3) the skeletons determined from the segmented image. Fig. 5 shows the feature extractor inputs.

The segmented images were generated from a three step process [16]. The initial step automatically generated and applied a global threshold to the entire input grey-level image, providing an under segmentation of the image. Second, for each connected component of the globally thresholded image, an automated local thresholding procedure was used to facilitate object separation. Finally, all objects were labeled using connected components analysis [17]. Each segmented image served as the outer boundary for orthogonal line construction from the skeleton. Only isolated chromosomes within metaphase spreads were of interest for this study.

B. Feature Extraction

Following segmentation and skeletonization, feature extraction was performed for each chromosome found within the metaphase spread image. The features computed and used for analysis include: 1) chromosome size (length and area), 2) centromeric index, 3) banding pattern features including total number of bands, p-arm bands, and q-arm bands, 4) profiles along the medial axis based on orthogonal lines to that axis, and 5) WDD function values [2], [19] from profiles including density and shape.

The algorithm for extracting features from chromosomes within metaphase spreads was largely based on the algorithm described in [18]. The actual implementation of the mean grey level or density profile, shape profile, chromosome width

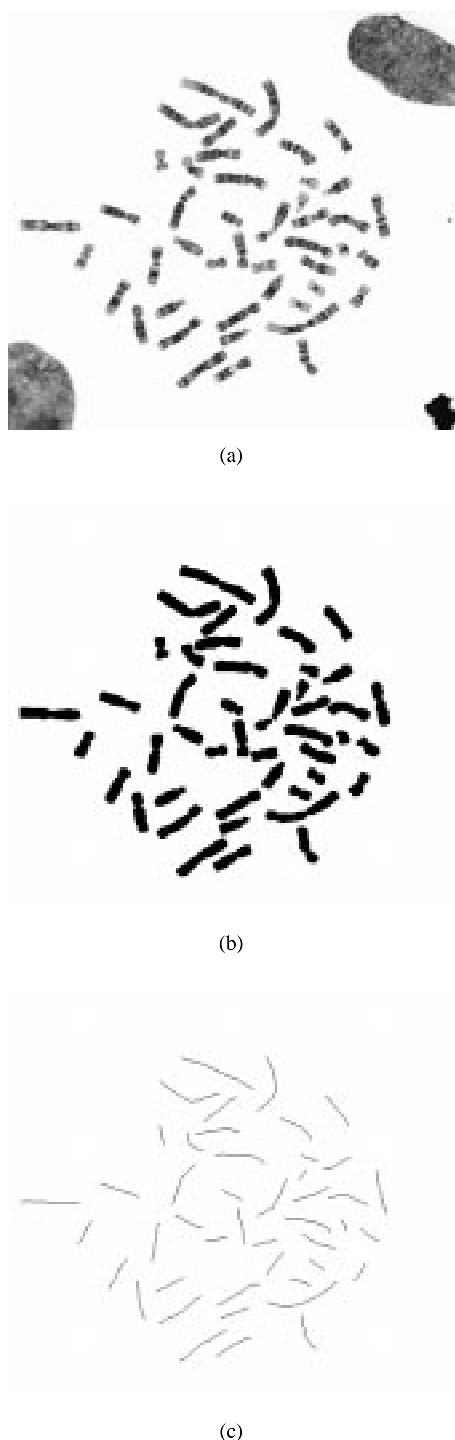


Fig. 5. Inputs to the feature extractor: (a) original metaphase spread image, (b) segmented image, and (c) skeletonized image.

profile, and WDD features is extensively described in [19]. The shape profile was computed at each axis point as the sum of the grey values at each orthogonal line point multiplied by its corresponding squared Euclidean distance from the axis point divided by the sum of the grey values along the perpendicular line [18], [20]. From those features, other attributes including the centromere, polarity assignment, and banding pattern representation were computed. The WDD features, derived by Granum *et al.* [2], [19], were computed over all isolated chromosomes and used as inputs to a feedforward network for assigning confidence values to the desired class.

C. Candidate Determination

Once the features were computed for the chromosomes found within the metaphase spread, candidate chromosomes from a selected class were determined using confidence values obtained from a feedforward neural network with scaled WDD features as inputs. The WDD features have provided high classification capability in the automated karyotyping system development at the University of Missouri. Other studies have shown that the WDD features provided significant discriminating capability for identifying chromosomes from specific classes [3], [19]. The weights used in the neural network for assigning confidence values to chromosomes within the metaphase spread were obtained from an "optimal" network. The "optimal" network was obtained using the procedure in Fig. 6.

The WDD features over the 4200 chromosomes comprising the training set were analyzed. The maximum and minimum WDD value for each of the 18 features were determined and stored in a look-up table. The WDD features computed in the test metaphase spreads were linearly scaled using the look-up table created from the training set. This mapped the feature values into the range -1 to 1.

The WDD features were computed and scaled for each chromosome within the metaphase spread image and input to this "optimal" feedforward network. Confidence values for each chromosome belonging to the class of interest were determined. The initial candidates chosen were the chromosomes with confidence values greater than zero in the desired class.

D. Candidate Elimination Using Band Features and Centromeric Index

Other studies suggest that the number of bands and centromere location provide chromosome classification discriminating capability [22], [23]. Upon finding the initial candidates, our research focused on eliminating candidates not possessing banding pattern representations and centromeres characteristic of chromosomes from the selected class. The banding pattern representation was based on finding the light-to-dark transitions in the median filtered density profile. Bands were segmented by finding the light-to-dark and corresponding dark-to-light transitions. Light-to-dark transitions are identified as inflection points (second derivative equal to zero) with first derivative greater than zero. Dark-to-light transitions are labeled as inflection points with first derivative less than zero. Dark bands consisted of the profile samples from the light-to-dark transition point up to but excluding the corresponding dark-to-light transition point. Light bands were defined similarly.

Based on the band segmentation, four band features were determined. The four band features were: 1) total bands, 2) p-arm bands, 3) q-arm bands, and 4) binary band profile. Total bands were computed as the total number of light and dark bands detected. In relation to centromere locations, p-arm and q-arm bands were calculated. The binary band profile is the

Obtain 4200 chromosomes with 175 per class from G-banded chromosome image library at the University of Missouri
Compute WDD features based on density and shape profiles for the 4200 chromosomes
Determine maximum and minimum value for each WDD feature
Linearly scale each WDD feature based on maximum and minimum found
Perform an 80-20 jackknife to obtain 133 per class for training set and 42 per class for testing set
Train and test network every 100 epochs until testing classification rate decreased
Obtain "optimal" network training/testing from previous step for variety of network configurations with one hidden layer
Retrain entire 4200 chromosome image features for "optimal" network configuration and number of epochs to obtain weights

Fig. 6. Procedure for finding "optimal" neural-network weights.

TABLE I
FEATURE VARIATION FOUND WITHIN CHROMOSOME-17 DATASET FOR 400-BAND LEVEL

	Centrometric index	Total bands	p-arm bands	q-arm bands
Maximum found	0.44	13	6	9
Minimum found	0.18	5	1	3

binary representation of the median filtered density profile as light and dark bands with dark bands as ones and light bands as zeros.

The centromeric index is defined as the ratio of the chromosome p-arm or short arm to the total chromosome length. The short and long arms are determined in relation to the centromere based on a centromere attribute integration approach. Based on the shape and width profiles, there are three centromere attributes integrated for centromere identification. The centromere should be located in a region of: 1) minimum width, 2) extreme concavity along the chromosome contour, and 3) uniform grey level. A formal discussion of the centromere identification algorithm is presented in [24].

In [22], an approximate linear relationship for the number of bands found with the chromosomes within a given cell was demonstrated. In order to account for the cell variability and band relationship, the range of band features and centromeric index were found over a chromosome image database for the selected class. Based on the band feature and centromeric index ranges found for the selected class such as chromosome 17 in Table I (Section II-E), candidate inclusionary rules were determined from the range of variation for those features.

E. Candidate Cross Validation

The candidates found from the neural network were cross validated with the band features and centromeric index ranges. Candidates not possessing band features and centromeric index within the acceptable ranges found from the database were eliminated from consideration. The ranges used were maximum and minimum values obtained from the training data, as shown in Table I. The band feature and centromeric index range constraints were utilized for two reasons: 1) they incorporated the chromosome classification delimiting capability of the number of bands [22] and centromeric index as shown in Fig. 1 and 2) they accommodated for variability in normal cells in establishing a baseline for normalcy. As an additional user-imposed constraint, candidates found to be strong winners by a significant margin, 0.90, in another class were also eliminated from further consideration. The homologue-matching algorithm utilized the strongest neural-network winner in the selected class as the prototype. Strong neural-network winners in other classes appeared to be potential prototypes for those classes, and, thus, rejected as candidates for the selected class. Table I presents the ranges found within the database used for validating candidates from class 17.

In order to facilitate banding pattern comparisons performed in the dynamic programming matching process, polarity must be determined for all chromosomes. Polarity refers to finding the top of the chromosome through identification of the p- and q-arms, designating the top of the p-arm as the top of the chromosome. The algorithm implemented for assigning polarity is discussed in detail in [25].

F. Primary Chromosome Selection and Matching to Candidate Process

From the remaining candidates, the chromosome with the greatest margin of victory in neural-network confidence was chosen as the reference, prototype, or primary chromosome. If no candidates remain, no chromosome was assigned to that class for the cell under consideration. After determining a primary chromosome and a set of candidate chromosomes, the remaining candidates were automatically inspected to determine the matching homologue using dynamic programming. Recalling the profile determinations from Section III-B, each isolated chromosome is characterized by profile values found sequentially at every medial axis point (sample). The designated starting point for every chromosome was the sample corresponding to the top of the chromosome, as found using the polarity assignment algorithm in Section III-E. Every sample point was characterized with a two-dimensional (2-D) vector including the scaled profile value and the corresponding chromosome incremental length value. The total chromosome Euclidean-distance length was broken into increments between consecutive samples, with incremental lengths between medial axis points of one or $\sqrt{2}$. The sum of those increments yielded the chromosome length. Fig. 7 illustrates the breaking of the medial axis lengths into incremental lengths with the

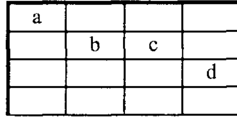


Fig. 7. Digital image grid containing medial axis points a, b, c, and d. The eight-connected Euclidean distance assigned to a and c is the square root of two, and Euclidean distance of one is reserved for b. Point d is a reference point and is excluded from the 2-D vector.

corresponding scaled profile values. The grid shown in Fig. 7 represents a digital image grid. Points a and c have incremental lengths of $\sqrt{2}$, and point b has incremental length of one. Point d is used as a reference position and is excluded from the 2-D vector over the chromosome length. The total length for the axis shown in Fig. 7 is two $\sqrt{2}+1$. For each axis point a, b, and c, there is a corresponding scaled profile value to make the 2-D vector along the chromosome length.

A scaled profile value corresponded to each profile length increment. With the length increments of one or $\sqrt{2}$, the profile was scaled mapping the maximum primary profile value to $\sqrt{2}$ and the minimum profile value to zero. Linear scaling was used to adjust the remaining primary profile values and also utilized to adjust the candidate profile values based on the same linear factors used in the primary profile. The profiles examined here were the median filtered density profile, the median filtered shape profile, and the binary band segmentation profile.

The matching process was performed between 2-D vectors of the primary and each of the candidate chromosomes. For our chromosome sequence matcher implemented, the cost function incorporated deletions and insertions but no substitutions [26]. The shorter of the primary sequence and the candidate sequence in samples was used as the reference sequence with the other as the comparator sequence. All 2-D elements or samples of the reference and comparator sequences were used without deletion. Deletions, in the context of this algorithm, were insertions of comparator sequence elements into the reference sequence. Our method for scoring the matching process was as follows:

1) Let the reference sequence be given as $\mathbf{Y} = (y_1, y_2, \cdots, y_m)$ and the comparator sequence be given as $\mathbf{X} = (x_1, x_2, \cdots, x_n)$, where $n \geq m$, $j = 1, \cdots, m$, $i = 1, \cdots, n$.

Here, the first component x_{i_1}, y_{j_1} represents the length increment, and the second x_{i_2}, y_{j_2} corresponds to the scaled profile value.

2) Define x_j^* as the current comparator element and $s_{r,s}$ as the minimum cost of matching the sequence $y_1, y_2, \cdots, y_r$ to $x_1, x_2, \cdots, x_s$.

For $j = 1$:
For $i = j, \cdots, j+n-m$: compute

$$x_{i_1}^* = \sum_{c=1}^{i} x_{c_1} \quad x_{i_2}^* = \frac{\sum_{c=1}^{i} x_{c_2}}{i},$$
$$s_{1,i} = |x_{i_1}^* - y_{j_1}| + |x_{i_2}^* - y_{j_2}|.$$

For $j = 2, \cdots, m-1$:
For $i = j, \cdots, j+n-m$:
For $k = j, \cdots, i$: compute

$$x_{i_1}^* = \sum_{c=k}^{i} x_{c_1}, \quad x_{i_2}^* = \frac{\sum_{c=k}^{i} x_{c_2}}{i-k+1},$$
$$s_{j,i} = \min_k (s_{j-1,k-1} + |x_{i_1}^* - y_{j_1}| + |x_{i_2}^* - y_{j_2}|).$$

For $j = m$:
$i = n$:
For $k = m, \cdots, n$: compute

$$x_{i_1}^* = \sum_{c=k}^{i} x_{c_1}, \quad x_{i_2}^* = \frac{\sum_{c=k}^{i} x_{c_2}}{n-k+1},$$
$$s_{m,n} = \min_k (s_{m-1,k-1} + |x_{i_1}^* - y_{j_1}| + |x_{i_2}^* - y_{j_2}|).$$

Here, $s_{m,n}$ is the final cost for matching X to Y. The dynamic programming matcher was used to score the primary chromosome to all other candidate chromosomes.

G. Confidence Distribution Formations

After completing the dynamic programming matching on all candidates, a confidence distribution was formed over all candidate to primary matching scores. The dynamic confidence distribution (DCD) was formed by the expression: 1—(*candidate score/highest score*), where the highest score was associated with the worst match case.

In addition to the confidence distribution formed by the dynamic programming scores, a confidence distribution was formed from a distance measure relating to the three-dimensional (3-D) chromosome feature vector containing scaled length, scaled area, and scaled centromeric index. For each isolated chromosome within the metaphase spread, the length, area, and centromeric index were computed. The mean and standard deviation were found for the length, area, and centromeric index over all isolated chromosomes within the metaphase spread. The length, area, and centromeric index features for each chromosome were scaled so that the mean value for each feature was zero and the standard deviation was one.

Based on the scaled length, area, and centromeric indexes, the city block distance was computed over the 3-D feature vectors between the primary chromosome and the candidate chromosomes. Using the distance measures between the primary and candidate chromosomes, a confidence distribution was formed. The distance measure confidence distribution (DMCD) was calculated as: *1—(candidate distance/maximum distance)*.

The final set of confidence values was determined from the confidence distributions of the dynamic programming scores and the size and centromere distance values as the product of the corresponding confidence values from the DCD and the DMCD. The candidate with the highest final confidence value was deemed homologue to the primary chromosome.

Retain original primary chromosome as a candidate
Find all non-zero neural network confidence chromosome corresponding to winning class of homologue
Apply band feature, centromeric index, and winning neural network margin rules to newly determined candidates
Perform dynamic programming matching between homologue chromosome and its candidates
Find dynamic confidence distribution
Compute city block distance between homologue and its candidates over scaled size and centromeric index features
Find distance measure confidence distribution
Compute final confidence distribution using confidence values from dynamic programming scores and distance measures
Identify best match to homologue as candidate with highest final confidence value

Fig. 8. Algorithm for matching homologue to candidates.

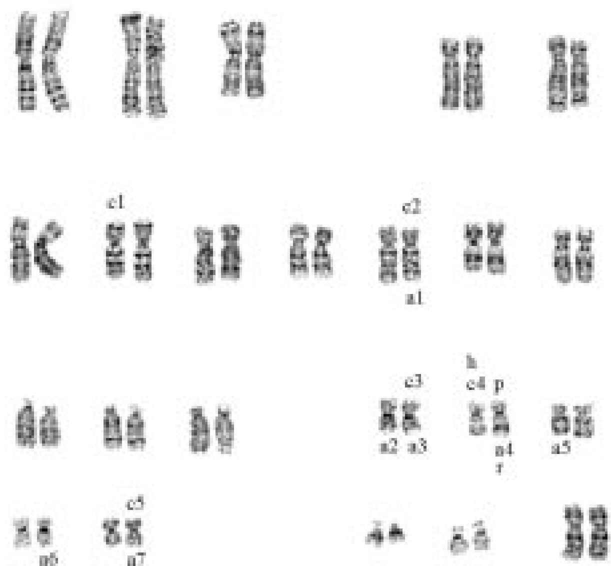


Fig. 9. Sample image analyzed by homologue matcher. $c1, \cdots, c5$: candidates for finding homologue; p: primary chromosome for finding homologue; h: homologue (best match to primary); $a1, \cdots, a7$: candidates for rematching homologue; r: rematch to homologue.

H. Homologue-Rematching Algorithm

The homologue was matched to find its closest fit using the algorithm in Fig. 8. If the best match was with the original primary chromosome, two chromosomes from the selected class were assigned for the metaphase spread. If the best match was not the original primary, the homologues were not matching. For this case, only the primary chromosome of the selected class was assigned for the metaphase spread. If no primary chromosome was found, no chromosome of the selected class was assigned for the metaphase spread.

I. Homologue-Matching Example

In order to illustrate the homologue-matching algorithm, an image example is presented. Fig. 9 contains a karyotype image analyzed by the homologue matcher. The chromosomes denoted with "c" are the initial candidates found satisfying the banding pattern and centromeric constraints from Table I, and label "p" corresponds to the primary chromosome. From the candidates, "h" denotes the homologue or best match to the primary chromosome using the density profile in the matching process. The homologue candidates are labeled "a," keeping the primary chromosome as one of the candidates, and "r" corresponds to the homologue's best match. Note that the homologues found are matching chromosome 17's. The winning neural-network classes for the primary and homologue chromosomes, 17 and 16, respectively, cause the candidate differences in finding the homologue and in rematching the homologue. Table VI presents the primary to homologue-matching confidence values for the dynamic programming confidence distribution and the DMCD in forming the final confidence values. The homologue-rematch candidates and confidence values are also shown in Table VI. The candidates are labeled in Table VI in accordance with Fig. 9.

IV. Experimental Results and Discussion

Three types of experiments were performed to assess the homologue-matching algorithm: 1) homologue matching over an entire image containing the same class of chromosomes, 2) homologue matching in the context of metaphase spreads, and 3) homologues chosen as top two neural-network winners.

A. Homologue Matching Over Entire Image Containing Same Class

The testing conditions were as follows: 1) obtain matching homologues from four cells, 2) imbed those matching homologues into an image of additional chromosomes which were from the selected class but were not matching homologues, 3) apply the homologue-matching algorithm using one chromosome from each pair as the primary chromosome, and 4) score the algorithm's ability to find the matching homologue for the four cases.

For the experiment performed, the composite image consisted of 16 chromosome 17's from various cells, possessing band levels from 400 to 550. There were four homologous pairs of chromosomes within the composite image. Fig. 3 contains the image of 16 chromosome 17's. The homologous pairs of chromosomes are labeled as pairs 1, 2, 3, and 4 on the image. The experimental results using each profile for finding the homologues are displayed in Table II, which shows that all homologues were matched. This is an important result for several reasons. First, matching homologues within the composite image of chromosomes from the same class manifests some of the differences between cells in preparation technique, banding pattern, and size. Second, the matching process shows the capability to find similar chromosomes. For the 16 chromosome 17's contained within the composite image, 14 had neural-

TABLE II
HOMOLOGUE MATCHING FOR CLASS 17 FROM COMPOSITE IMAGE

Homolg pair	DM	SM	BM
1	y	y	y
2	y	y	y
3	y	y	y
4	y	y	y

DM: Scaled density profile with the size distance measure.
SM: Scaled shape profile with the size distance measure.
BM: Scaled binary band segmentation profile with the size distance measure.
y: Homologues correctly found.

TABLE III
HOMOLOGUE MATCHING RESULTS FOR 55 METAPHASE SPREADS

Cell	DM	SM	BM	Cell	DM	SM	BM
1	y	y	y	29	y	y	m
2	y	y	y	30	y	i	y
3	y	i	y	31	i	i	i
4	y	y	i	32	y	y	y
5	y	y	y	33	y	i	y
6	y	y	y	34	y	y	y
7	y	i	i	35	y	y	y
8	y	y	y	36	y	y	i
9	y	y	y	37	y	y	y
10	i	i	y	38	y	y	y
11	y	i	y	39	y	m	y
12	y	y	y	40	y	y	y
13	y	i	y	41	y	y	y
14	y	y	y	42	y	y	y
15	y	y	y	43	y	y	y
16	y	y	y	44	y	y	y
17	y	i	y	45	y	y	y
18	y	y	y	46	y	i	y
19	y	y	y	47	y	y	y
20	i	i	y	48	y	y	y
21	y	y	y	49	y	y	y
22	y	y	y	50	y	y	y
23	i	i	i	51	y	y	y
24	y	y	y	52	y	y	y
25	i	y	i	53	i	i	i
26	y	y	y	54	i	i	i
27	y	y	y	55	i	i	i
28	y	y	y				

DM: Scaled density profile with the size distance measure.
SM: Scaled shape profile with the size distance measure.
BM: Scaled binary band segmentation profile with the size distance measure.
y: Correct homologues for chromosome 17 found (two chromosomes assigned).
i: No matching homologues found for chromosome 17 (one chromosome assigned).
m: Incorrect homologues for chromosome 17 found (two chromosomes assigned).

network confidence values which were the winner for class 17. The ability to find chromosomes coming from a specific class is important for karyotyping and anomaly detection. For analyzing chromosomal anomalies, a baseline for comparison must be established. Because chromosomes in a normal cell occur in pairs or homologues, the ability to disseminate the degree of similarity is important for evaluating chromosomes within the context of a given cell. For abnormality analysis, the higher the confidence in the degree of similarity between chromosomes (pairs), the higher the confidence of predicting that chromosomes found for a specific class within a given cell are normal. Using a vigorous matching process for homologue assessment enables prediction with high confidence that chromosomes for a selected class indeed belong to that class and are similar, which is the process of evaluating normalcy. This experiment demonstrates that the baseline is not achieved using only neural-network confidences.

B. Homologue Matching Using Metaphase Spreads

The second set of experiments analyzed the ability of our system to find chromosomes of the selected class from metaphase spreads. Chromosome 17 was the selected class. Using an independent test set, the experimental procedure followed for 55 metaphase spreads obtained from the University of Missouri library was: 1) input each metaphase spread using the algorithm for homologue matching previously described and 2) score the homologue-matching results for the density profile, the shape profile, and the binary band segmentation profile. The experimental results for the 55 metaphase spreads tested are shown in Table III.

Of the 55 metaphase spreads analyzed for this experiment, 53 contained two isolated chromosome 17's. For those metaphase spreads, a "y" was obtained where the primary chromosome was properly determined, the correct homologue was found, and the homologue matched to the primary chromosome. An "i" was assigned where only the correct primary chromosome was found, and the homologue did not match the primary chromosome, i.e., the homologue matched to a chromosome other than the primary. For several cases, the primary chromosome matched to the correct homologue but the homologue matched to another chromosome, thereby making the homologue match incomplete. An "m" was assigned for cases where the primary chromosome was correctly found but matched to an incorrect homologue, i.e., chromosome from a class other than 17. Also, an "m" was assigned for cases where the primary chromosome was incorrectly determined. For the dataset tested, the primary chromosome was always found from the correct class, chromosome 17.

Besides the 53 cases where there were two isolated 17's within the metaphase spread, there were two cases with only one normal chromosome 17 within the cell. These two cases instantiated the proof of concept process for abnormality assessment using homologues. In those cells there was a translocation between chromosomes 15 and 17, distorting one of the chromosome 17's. This translocation has been associated with a form of acute promyelocytic leukemia. Possessing only one "normal" chromosome 17, primary chromosome determination is expected. However, the primary chromosome should have no matching homologue for proper analysis of the cell, i.e., there is only one chromosome 17 assigned from the metaphase spread. Correspondingly, a correct analysis of the metaphase spread is scored with an "i." Metaphase spreads (cells) 54 and 55 in Table III have the single normal chromosome 17.

As part of the algorithm for obtaining the primary chromosome and candidate chromosomes, the confidence values from the feedforward network were analyzed. In order to have an

effective, reliable diagnostic tool, the FP rate must be minimal. FP's for this experiment are expressed as m's, meaning incorrect homologues are matched. This becomes important in abnormality assessment for cases where the anomaly is detected based on numerical aberrations for a specific class. An "m" corresponds to asserting that a patient does not have a specific condition in the case where the condition is present. Future studies will extend the example for cells 54 and 55 concerning numerical aberrations involving normal chromosomes 9 and 22 in relation to leukemias associated with the Philadelphia chromosome. The homologue-matching algorithm will be used to detect a single "normal" chromosome 9 and a single "normal" chromosome 22 within a metaphase spread.

C. Transportation Algorithm for Assigning Homologues

As previous discussed, the transportation algorithm was used as the benchmark for comparison to the homologue-matching approach. The transportation algorithm was used to optimize the feedforward neural-network confidence values for chromosome assignment. The transportation-algorithm implementation for this study utilized the two-chromosome-per-class constraint for classes 1–22, two class-X chromosomes, and one class-Y chromosome. Because the homologue-matching algorithm had no *a priori* knowledge concerning the sex associated with the metaphase spreads tested, the transportation algorithm made the sex determination in the chromosome assignment process. The feedforward neural network used in the homologue-matching algorithm was based on a 24-class problem. The cost values incorporated into the transportation algorithm simply took one minus the neural-network confidence value over the 24 classes for each isolated chromosome. The transportation algorithm, then, made the best possible assignments for the isolated chromosomes within the metaphase spreads.

The results from applying the transportation algorithm over the same dataset are shown in the Table IV. For each cell, the transportation algorithm found the correct homologues, denoted "y," or did not find the correct homologues, denoted "n." For abnormal cells 54 and 55, finding the correct single chromosome 17 is denoted "y," and any other number of chromosome 17's found is denoted "n." From the tabular results, the transportation-algorithm approach found the correct homologues in 44 of the 55 metaphase spreads, 80.0% of the cases. For comparative purposes, the summary results for Table III are shown in Table V.

Although the transportation-algorithm approach has been successfully applied to chromosome recognition, there are several pitfalls. First, the transportation algorithm assumes the chromosomes within the cell are normal for classification purposes. Cells 54 and 55 contained one "normal" chromosome 17 and one "abnormal" chromosome 17 (also one "normal" chromosome 15 and one "abnormal" chromosome 15), but the transportation algorithm assigned the normal and abnormal chromosome(s) based on their cost values to some class. The chromosome assignment was made without regard to the normalcy of the chromosomes within the metaphase spreads. Using the principle of least commitment and the indeterminate case facilitates omitting chromosome assignment to any class within the metaphase spread. From Tables III and IV, two chromosomes were assigned to class 17 for all 55 metaphase spreads tested. Thus, the automated analysis of the cells provided for normal class-17 homologues. The homologue-matching algorithm provides a cueing capability with the indeterminate case that further analysis of chromosomes from the selected class is appropriate. This further analysis can be performed within the framework of the karyotyping process. Second, there is a difference between numerical aberrations within a cell and abnormal chromosomes within a cell. Down's Syndrome involves a trisomy 21, three "normal" chromosome 21's within a metaphase spread. Through the use of surplus and slack variables, the transportation algorithm has the capability

TABLE IV
TRANSPORTATION-ALGORITHM HOMOLOGUE IDENTIFICATION RESULTS FOR 55 METAPHASE SPREADS

Cell	Transport Homologues	Cell	Transport Homologues
1	n	29	y
2	y	30	n
3	y	31	y
4	n	32	y
5	y	33	y
6	y	34	y
7	n	35	y
8	y	36	y
9	y	37	y
10	y	38	y
11	y	39	n
12	y	40	y
13	y	41	y
14	y	42	y
15	y	43	y
16	n	44	y
17	y	45	y
18	y	46	y
19	y	47	y
20	y	48	y
21	y	49	y
22	y	50	y
23	y	51	y
24	y	52	n
25	n	53	n
26	y	54	n
27	y	55	n
28	y		

y: Transportation algorithm homologues found were correct. (two chromosomes assigned).
n: Transportation algorithm homologues found were incorrect or incorrect for number of chromosomes 17's assigned for cell.

TABLE V
CORRECT HOMOLOGUE MATCHING RESULTS FOR 55 METAPHASE SPREADS

	DM	Sm	BM
Number correcy	49	42	47
% correct	89.1	76.4	85.5

DM: Scaled density profile with the size distance measure.
SM: Scaled shape profile with the size distance measure.
BM: Scaled binary band segmentation profile with the size distance measure.

TABLE VI
CONFIDENCE SCORES FOR DYNAMIC PROGRAMMING AND DISTANCE MEASURE FOR PRIMARY TO HOMOLOGUE MATCH AND REMATCH FOR FIG. 9. COLUMNS 2 AND 6 CONTAIN THE DCD OVER THE REFERENCE CHROMOSOME-TO-CANDIDATE MATCH. COLUMNS 3 AND 7 CONTAIN THE DMCD OVER THE REFERENCE CHROMOSOME-TO-CANDIDATE MATCH. COLUMNS 4 AND 8 CONTAIN THE FINAL CONFIDENCE VALUES (DCD $\times$ DMCD) IN THE MATCHING PROCESS. * DENOTES THE WINNING CANDIDATE

Primary to Homologue Match				Homologue Rematch			
	DCD	DMCD	Final Confidence		DCD	DMCD	Final Confidence
c1	0	0	0	a1	0	0	0
c2	0.35	0.266	0.093	a2	0.657	0.861	0.565
c3	0.583	0.507	0.296	a3	0.557	0.378	0.21
*c4	0.801	0.913	0.731	*a4	0.723	0.884	0.639
c5	0.789	0.816	0.644	a5	0.661	0.9	0.595
				a6	0.738	0.713	0.526
				a7	0.719	0.787	0.566

to accommodate for the numerical aberration in class 21. However, the surplus and slack variables cannot be used for discerning the normalcy of a particular chromosome, i.e., structural abnormalities in relation to the corresponding homologue. Using a prototype of a selected class within the metaphase spread, a benchmark for normalcy is generated and applied to metaphase spread with respect to the selected class. The database of normal chromosome features is utilized for establishing the benchmark in finding the prototype and its homologue. Because of chromosome variations due to preparation technique and natural variations within and between people, ideal prototype generation is difficult. The data-driven approach accommodates for the variability and allows chromosome assignment to be performed within the context of the cell. Third, the primary reason for implementing the matching procedure is to accommodate a variable number of chromosomes per class, particularly in relation to anomaly analysis. Although the transportation algorithm has the capability for accommodating for a variable number of chromosomes per class, the transportation algorithm in this study assigned two chromosomes 17's for all 55 metaphase spreads tested. Using the feedforward neural network in tandem with a matching process places tighter constraints on the classification of chromosomes from a selected class. This is evident in the homologue-matching results from Table V for the density profile (DM) and binary band segmentation profile (BM) with 89.1% and 85.5% correct recognition rates, respectively, in comparison to the transportation-algorithm approach of 80.0%. The homologue-matching algorithm has the capability to account for a variable number of chromosomes and to omit chromosomes outside the expected range for a normal chromosome from class assignment within a metaphase spread. Finally, the homologue-matching scheme addresses concerns beyond correct classification. The principle of least-commitment extends the decision making process, making decisions in extremely confident cases or delaying decision in nebulous cases. For the homologue-matching algorithm presented, in the context of diagnosing an abnormality on the basis of chromosomal numerical aberrations, not assigning a chromosome to a specific class necessitating further analysis for the metaphase spread is better than making an incorrect assignment and missing the diagnosis completely. In terms of karyotyping, this corresponds to avoid matching homologues that do not match and using supplemental analysis of the metaphase spread to find the ultimate chromosome classification.

V. CONCLUSION

The experimental results from the homologue-matching algorithm were better for the density and binary band representation profiles than for the transportation-algorithm approach. Not only did the homologue-matching technique achieve better results for those profiles than the transportation-algorithm approach, many of the mistakes made by the homologue method were recoverable. From Table III, "i" refers to an incomplete homologue match, where only the primary chromosome is assigned to the selected class for cells containing two normal chromosomes, honoring the principle of least commitment. In the case of normal cells, no assignment is made for the second chromosome of the selected class, allowing further analysis for more informed assignment. For abnormality analysis, an "i" means that the algorithm was unable to detect an expected homologous pair for the selected class, necessitating further metaphase spread analysis. In other words, an "i" sets a flag for potential abnormality presence.

In addition to recoverable mistakes, an "m" from Table III is an unrecoverable error. This means that at least one chromosome is incorrectly assigned to the selected class. From an abnormality perspective, an unrecoverable error results in a FN classification, leading to a potential negative diagnosis that a patient possesses a specific condition. Notice that "m" is absent in the experimental results for the density profile extension of the homologue-matching algorithm. For the experiments performed, this means that the mistakes made are recoverable. Encouraging results have been obtained for the selected class, 17. Further experiments in karyotyping and anomaly assessment will test the homologue-matching algorithm robustness.

REFERENCES

[1] C. Lundsteen, A. M. Lind, and E. Granum, "Visual classification of banded human chromosomes i. Karyotyping compared with classification of isolated chromosomes," *Ann. Human Genetics,* vol. 40, no. 1, pp. 87–97, 1976.

[2] E. Granum, T. Gerdes, and C. Lundsteen, "Simple weighted density distributions, WDD's for discrimination between G-banded chromosomes," presented at *4th Eur. Chrom. Anal. Workshop,* Edinburgh, 1981.

[3] W. P. Sweeney, M. T. Musavi, and J. N. Guidi, "Classification of chromosomes using a probabilistic neural network," *Cytometry,* vol. 16, pp. 17–24, 1994.

[4] J. Piper, "Classification of chromosomes constrained by expected class size," *Pattern Recogn. Lett.,* vol. 4, pp. 391–395, 1986.

[5] J. Graham and J. Piper, "Automated karyotype analysis," in *Methods in Molecular Biology: Chromosome Analysis Protocols,* J. R. Gosden, Ed. Totowa, NJ: Humana, 1994, vol. 29.

[6] D. Marr, *Vision.* San Francisco, CA: Freeman, 1982.

[7] J. Keller, P. Gader, and C. W. Caldwell, "The principle of least commitment in the analysis of chromosome images," in *Proc. SPIE Symp. OE/Aerospace Sensing and Dual Use Photonics,* Orlando, FL, Apr. 1995, pp. 178–186.

[8] D. F. Specht, "Probabilistic neural networks," *Neur. Networks,* vol. 3, pp. 109–118, 1990.

[9] P. A. Errington and J. Graham, "Application of artificial neural networks to chromosome classification," *Cytometry,* vol. 14, pp. 627–639, 1993.

[10] B. Lerner, H. Guterman, I. Dinstein, and Y. Romem, "Medial axis transform-based features and a neural network for human chromosome classification," *Pattern Recogn.,* vol. 28, no. 11, pp. 1673–1683, 1995.

[11] M. K. Tso and J. Graham, "The transportation algorithm as an aid to chromosome classification," *Pattern Recogn. Lett.,* vol. 1, pp. 489–496, 1983.

[12] M. Tso, P. Kleinschmidt, I. Mitterreiter, and J. Graham, "An efficient transportation algorithm for automatic chromosome karyotyping," *Pattern Recogn. Lett.,* vol. 12, pp. 117–126, 1991.

[13] S. O. Zimmerman, D. A. Johnston, F. E. Arrighi, and M. E. Rupp, "Automated homologue matching of human G-banded chromosomes," *Comput. Biol. Med.,* vol. 16, pp. 223–233, 1986.

[14] J. Gregor and E. Granum, "Finding chromosome centromeres using band pattern information," *Comput. Biol. Med.,* vol. 21, nos. 1/2, pp. 55–67, 1991.

[15] M. G. Thomason and E. Granum, "Dynamic programming inference of Markov networks from finite sets of sample strings," *IEEE Trans. Pattern Anal. Machine Intell.,* vol. PAMI-8, pp. 491–501, 1986.

[16] A. Gintingsuka, "Using pale paths and cross section sequence graphs to segment chromosomes automatically," M.S.E.E. thesis, Electrical and Computer Eng., Univ. Missouri, Columbia, 1996.

[17] R. C. Gonzalez and R. E. Woods, *Digital Image Processing.* Reading, MA: Addison-Wesley, 1992.

[18] E. Granum, "Pattern recognition aspects of chromosome analysis. Computerized and visual interpretation of banded human chromosomes," Ph.D. thesis, Electronics Lab., Tech. Univ. Denmark, Lyngby, Denmark, pp. 7306–7316, 1980.

[19] J. Piper and E. Granum, "On fully automatic feature measurement for banded chromosome classification," *Cytometry,* vol. 10, pp. 242–255, 1989.

[20] R. J. Stanley, "Feature stability in automated chromosome analysis," M.S.E.E. thesis, Electrical and Computer Eng., Univ. Missouri, Columbia, 1994.

[21] J. Piper, "Finding chromosome centromeres using boundary and density information," in *Digital Image Processing,* J. C. Simon and R. M. Haralick, Eds. Dordrecht, the Netherlands: Reidel, 1981, pp. 511–518.

[22] D. A. Johnston, K. S. Tang, and S. Zimmerman, "Band features as classification measures for G-banded chromosome analysis," *Comput. Biol. Med.,* vol. 23, no. 2, pp. 115–129, 1993.

[23] K. M. Gustashaw, "Chromosome stains," in *The ACT Cytogenetics Laboratory Manual,* 2nd ed., M. J. Barch, Ed. New York: Raven, 1991, pp. 205–269.

[24] R. J. Stanley, J. Keller, C. W. Caldwell, and P. Gader, "A centromere attribute integration approach to centromere identification," *Biomed. Sci., Instrum.,* vol. 32, pp. 23–29, 1996.

[25] R. J. Stanley, J. Keller, P. Gader, and C. W. Caldwell, "Centromere attribute integration based chromosome polarity assignment," *JAMIA,* vol. 3, suppl., pp. 284–288, 1996.

[26] D. Sankoff and J. B. Kruskal, *Time Warps, String Edits, and Macromolecules: The Theory and Practice of Sequence Comparison.* Reading, MA: Addison-Wesley, 1983.