Automated metaphase chromosome centromere refinement using fuzzy inference systems.

Akila Subasinghe A[a1], Jagath Samarabandu[a2] , Joan Knoll[b1] and Peter Rogan[b2]

[a]Image Recognition & Intelligent Systems Laboratory,University of Western Ontario,ON,Canada;
[b]Schulich School of Medicine & Dentistry,University of Western Ontario,ON,Canada.

ABSTRACT

Accurate detection of the centromere is a crucial aspect in any automated chromosome analysis system. Although many approaches exist in literature, they suffer from significant inaccuracies in detection. This is especially the case when presented with highly bent chromosomes or chromosomes with high boundary noise. Therefore, we propose an algorithm which utilizes two of the most common profile features used in literature and refines the position of inaccurate centromere localizations. The proposed algorithm is based on zadeh-mamdani type fuzzy logic systems and can be applied selectively using a measure proposed in our previous works. Preliminary tests have yielded satisfactory results towards the performance of the algorithm.

Keywords: fuzzy inference systems, centromere refinement, chromosome analysis

1 INTRODUCTION

The centromere is the most condensed and constricted region of a chromosome to which the spindle fiber is attached during mitosis (cell division). The detection of centromere is important in almost all classification methods as the centromere index value is a key feature which can be used to relate a given chromosome to its group. Centromere detection processes are in general prone to significant levels of detection error [1],[2]. Most of these detection errors arise due to errors in misalignment present in the detected centerline of the chromosome [3]. These noisy data points typically cause the width profile measurements to miss the sharp width change at the constriction. This detects an incorrect position as the centromere. The majority of these deviations are of local in nature. i.e. - the detected position and the desired position are close together [4]. These slight offsets can adversely affect many measurements such as centromere index (CI = p / (p + q)), polarity assignment (p-short and q-long arm assignment). This is especially applicable to most of the meta-centric chromosomes which have a centromere index (CI) very close to 0.50.

Although false detections can alter many measurements, existing centromere detection algorithms tend to ignore this problem and move ahead with further calculations. The inability to gauge the reliability of a detected centromere location prove to be a major contributor to the said problem. In previous works, we have proposed a measurement termed the CCF (centromere confidence measurement) [5]. This provides a confidence measurement for the detected centromere location based on the width and intensity profiles. CCF has been observed to be a necessary but not sufficient condition for a false centromere location. Therefore by applying a threshold for this measurement, we can apply the refinement process to centromere locations selectively. In this paper, a refinement procedure is proposed to converge on the correct centromere location.

2 METHOD

We have proposed a novel 'centromere refinement' process which attempts to correct offsets present in detected centromere locations. The proposed refinement algorithm was designed to satisfy the following objectives,

- This process should not displace an accurate centromere localization.
- The proposed algorithm should attempt to correct a local incorrect localization due to noisy measurements in the profile.

The proposed refining stage utilizes information gathered through previous stages of our algorithm [4],[6], which is the centerline (axis of symmetry) and the segmented chromosome outline. We have also developed a measurement to gauge the confidence of the current centromere location, called the centromere confidence measure (CCF

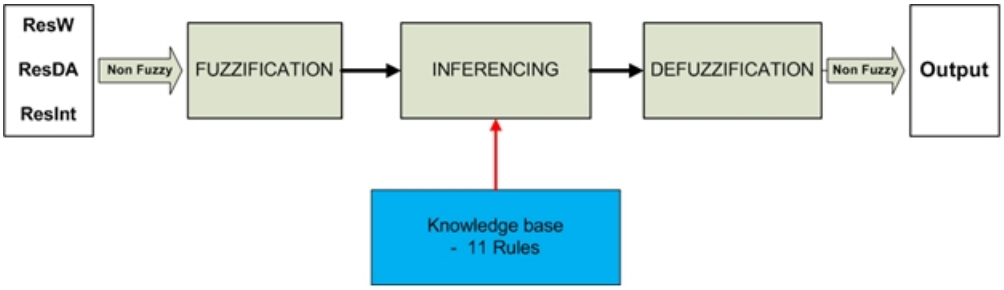

Figure 1. The basic fuzzy logic setup

value) [5]. The use of this CCF value allows us to apply the proposed refinement algorithm selectively. The proposed refinement process was based on the width and intensity profiles of the chromosome. This selection of feature profiles increases the adaptability of the proposed refinement process into many existing centromere detection algorithms [1]. Results obtained from previous work demonstrates that majority ($> 90\%$) of the centromere detections are either accurate or neighboring the actual centromere location [4]. Therefore the search space for the proposed method was set to 20% of the total length of the chromosome, in the vicinity of the originally detected centromere location.

Next, a total of 72 line segments covering all 360° were drawn through each of these extracted centerline points. The end points of these line segments were detected using the previously extracted binary object contour (using GVF active contours). Then for each line segment, the following three measurements were recorded.

- W_R-The relative width of the line segment, based on the binary object contour. The value W_R was calculated as a ratio between two measurements as -

$$\left[\frac{width\,of\,the\,line\,segment}{average\,of\,the\,width\,profile} \right]$$

  Therefore, the W_R is a relative measure that depicts the width of any line segment compared to the average width value calculated from the original width profile. A particular value below 1.0 for W_R would yield a potential candidate for the centromere location while increasing candidacy with decreasing values of W_R.

- θ_D-The angular difference or offset between the direction of the line segment and the expected direction given by the current normal to the centerline pixels. This feature provides information regarding the deviation of the current line segment from that considered in the original width profile.

- I_A-The average pixel intensity in a 5X5 neighborhood around the candidate centerline pixel. This value was normalized as follows -

$$\left[\frac{average\,pixel\,intensity\,(local)}{maximum\,average\,intensity\,in\,the\,search\,space} \right]$$

  The I_A value gives the relative intensity (on Inverted DAPI) at a particular centerline pixel. Taking the average reduces biasing due to noisy data. Higher values of I_A (close to unity) is preferred.

The two best entries (based on 'W_R') were selected to avoid getting false negatives in the final data set while possibly containing the optimum candidate. This yields 2 x n candidate solutions, where 'n' is the size of the search space.

Detecting the centromere location of a human metaphase chromosome is a challenging task even for a human eye (untrained). Furthermore, the relationships between different levels of the input features and the centromere location had to be derived. Therefore, expert knowledge has to be incorporated into the decision making process of selecting the best candidate for the centromere. This was achieved using "Fuzzy logic systems" (zadeh - mamdani type) [7], which is a well known method formulated to embed a set of "linguistic rules" into the inference engine (figure 1).

After setting up appropriate fuzzy membership functions, a set of 11 fuzzy rules (refer table 3) were formulated to define the solution space for the system (the output variable). These rules are meant to bias the output towards lower W_R and θ_D while retaining high I_A values. This aids in finding a better constriction using both width and

Table 1. If-Then rule set used in the fuzzy logic setup. These 11 rules are meant to bias the fuzzy output towards lower W_R and θ_D values while maintaining high values for I_A

Rule Number	W_R	θ_D	I_A	Output Level
01	Very Low	Low	-	Ultra High
02	Very Low	Medium	-	Ultra High
03	Low	Low	-	Very High
04	Low	Medium	-	Medium High
05	Low	High	-	Medium Low
06	Medium	Low	High	Medium High
07	Medium	Medium	High	Medium Low
08	Medium	High	-	Very Low
09	Medium	Low	Low	Medium Low
10	Medium	Medium	Low	Very Low
11	High	-	-	Ultra Low

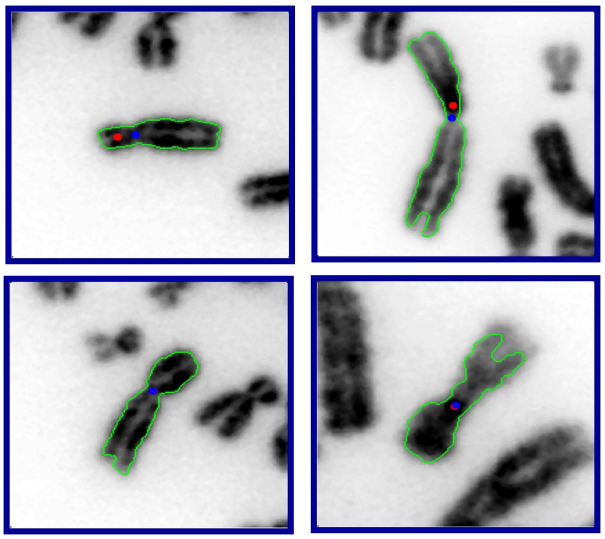

Figure 2. Sample results of the proposed centromere refinement algorithm. 'RED'-Initial centromere location, 'BLUE'-Refined centromere location

intensity while allowing some flexibility for the angular difference. This flexibility will alow the refining process to capture more suitable candidates and thereby compensate for the missed constrictions due to noisy centerline data. Finally, the refined centromere location was calculated by selecting the optimum fuzzy logic output value out of the 2*n fuzzy outputs. Due to low granularity in some regions of the solution space, identical optimum output values can occur. In these special cases, the final decision is derived to support the candidate point with the lowest offset from the original centromere location.

3 RESULTS

Preliminary tests were carried out to validate the performance of the proposed algorithm on metaphase chromosomes of human patients. 26 out of 30 chromosomes (87%) tested with the proposed algorithm, either improved or kept the current centromere location while 13% cases showed a significant error or deviation from the actual centromere. Within the 26 cases, 12 (46%) showed a significant improvement of the current centromere location towards the actual centromere, while 14 (54%) retained the previous centromere location. The proposed method performs well in retaining correct centromere locations while improving the positions of mischaracterized centromeres. Retaining a good centromere location is equally important as improving an inaccurate centromere location in practice. Figure 2 depicts some results of the proposed centromere refinement algorithm.

4 CONCLUSION & FUTURE WORK

We have presented an algorithm for refining the location of a human metaphase chromosome with considerable accuracy. Setting a threshold for the CCF value at 90% allows selecting centromere locations that have a higher probability of been misplaced. Correcting or refining the centromere location will in return improve the accuracy of any other calculations such as the Centromere Index (CI), polarity assignment. The rules and the membership functions can be further adjusted and tuned for better performance in the future. Studying more test cases would further justify the algorithm while allowing for better adjustments of the parameters.

References

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

[2] M. Moradi and K. Saterahdan, "New features for automatic classification of human chromosomes : A feasibility study," *Pattern Recognition Letters* (27), pp. 19–28, 2006.

[3] X. Wang et al., "A rule-based computer scheme for centromere identification and polarity assignment of metaphase chromosomes," *Computer Methods and Programs in Bio Medicine* **89**, pp. 33–42, 2008.

[4] A. Subasinghe A. et al, "An accurate image processing algorithm for detecting fish probe locations relative to chromosome landmarks on dapi stained metaphase chromosome images," *Computer and Robot Vision, Canadian Conference* , pp. 223–230, 2010.

[5] "Towards automated single copy fish probe and centromere identification on metaphase chromosomes," in *International Biodosimetry workshop*, May 2010.

[6] A. S. Arachchige, J. Samarabandu, J. Knoll, W. Khan, and P. Rogan, "An image processing algorithm for accurate extraction of the centerline from human metaphase chromosomes," in *ICIP*, pp. 3613–3616, 2010.

[7] L. Zadeh, "Fuzzy sets as a basis for a theory of possibility," *Fuzzy Sets and Systems* **1**(1), pp. 3 – 28, 1978.