CURVES FOR MODELLING CHROMOSOME SHAPES

C. M. Merritt
National Research Council of Canada
Ottawa, Ontario, Canada

ABSTRACT

This summary outlines the results obtained by using a number of different types of smooth curves for modelling the backbones of chromosomes, an important step in image analysis of chromosomes. Of the curves investigated, only the cubic spline provides adequate flexibility while retaining smoothness. The full paper is to appear in Analytical and Quantitative Cytology.

KEYWORDS: image processing, chromosomes, curve-fitting.

SUMMARY

In computer-aided systems for chromosome analysis, features are extracted from the image of a chromosome spread which allow the identification of individual chromosomes in the sample or the detection of fine chromosomal structures. In the case of visual analysis of human chromosomes stained to reveal a banding pattern, the banding pattern is used to uniquely identify each chromosome in a cell. The bands lie perpendicularly to the medial axis or backbone of the chromosome; thus, in a computer-aided system for chromosome analysis, the integrated optical density (i.o.d.) profile, which is formed by summing the densities on vectors perpendicular to the medial axis and centered at each point on the axis, can be used as a feature reflecting the banding pattern of a chromosome. The i.o.d. profile is also valuable in the identification of homogeneously stained (unbanded) chromosomes, by enabling the chromosome centromere to be located.

In order to calculate the i.o.d. profile, the backbone of the chromosome must first be modelled by a smooth curve. From such a curve points lying at regular intervals along the backbone can be interpolated, and the orientation of vectors centered at these points can be calculated in order to form the profile. In order for the i.o.d. profile to accurately represent the banding pattern of a chromosome, the curve must closely follow the actual backbone of the chromosome and must have smoothly varying first and second derivatives.

A number of different types of curves for modelling chromosome shapes has been investigated. Two types of curves have commonly been used in existing systems for human chromosome analysis: 1) straight line approximations, and 2) parabolic or low-order polynomial approximations. A straight line approximation is found by calculating the principal axis based on the contour points (Figure 1(a)). A parabolic approximation is obtained when this principal axis is modified to a parabolic axis (Figure 1(b)). Alternatively, points on the backbone of the chromosome may first be found by a skeletonization process and a low-order polynomial may be fitted to these points (Figures 1(c) and 1(d)).

Our studies of plant chromosomes have shown that these chromosomes tend to be long and often severely bent, and consequently require a rather flexible modelling curve. Both the straight line and the polynomial approximations have been found to be insufficiently flexible to closely follow the backbone of such bent chromosomes. The piecewise cubic polynomial known as the cubic spline provides the required flexibility, while having continuous derivatives. Fitted to knots selected from the skeleton points, this curve has been shown to closely follow the chromosome backbone. In some cases, however, extraneous inflection points between knots may cause the curvature and slope of the curve to vary too quickly; as a result, the vectors used to form the i.o.d. profile do not lie at appropriate angles relative either to the chromosome or to each other. A modification to the regular cubic spline, described by A.K. Cline, in which the curve is fit "under tension" has been found to adequately smooth the derivatives of the curve. The method involves imposing tension on the curve, as if pulling on its ends, in order to restrict the amount of curviness allowed, thereby

removing unwanted inflection points. The amount of tension applied may be varied: as it approaches zero the curve approaches the regular spline and as it becomes large, the curve becomes nearly a polygonal line. Cline suggests a standard value of 1.0 for the tension factor; this value has given good results when used for the chromosome samples (Figure 1(e)). Recent results indicate that, when the knots are well-chosen, the regular cubic spline gives the same curve as the spline fitted under a tension factor of about 1.0. The tension factor does, however, provide an improvement when the knots are not well-positioned.

Initial investigation of banded human chromosome samples, as shown in Figure 1, indicates that many of these also require the flexibility provided by the cubic spline curve. In particular, chromosomes which are abnormally long due to insertions or duplications of chromosomal material tend to exhibit multiple bends and therefore cannot be modelled by straight line or parabolic approximations. If the structures of these abnormalities are to be analysed in detail, it is particularly important that the i.o.d. profiles accurately reflect the banding patterns of the chromosomes. Figure 2 shows the i.o.d. profiles obtained from each of the curves of Figure 1; although most of the profiles exhibit a similarity in overall form, differences in details can be seen.

While the straight line or low-order polynomial approximations have been found to adequately model some chromosome shapes, in particular those simple chromosomes which are short and only slightly bent, they are not general enough to fit the wide variety of chromosome shapes encountered. The cubic spline curve, on the other hand, is sufficiently flexible to model both the simple chromosomes and those more severely bent; at the same time, the cubic spline provides the smoothness of first and second derivatives required to form an accurate i.o.d. profile.

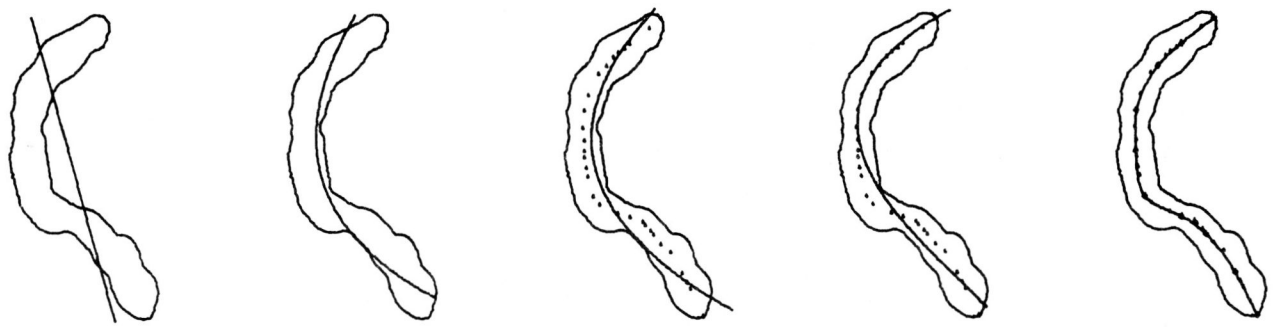

Figure 1. Contour of normal banded human chromosome with modelling curves.

(a) Straight line (b) Parabola (c) 2nd degree polynomial (d) 3rd degree polynomial (e) cubic spline under tension

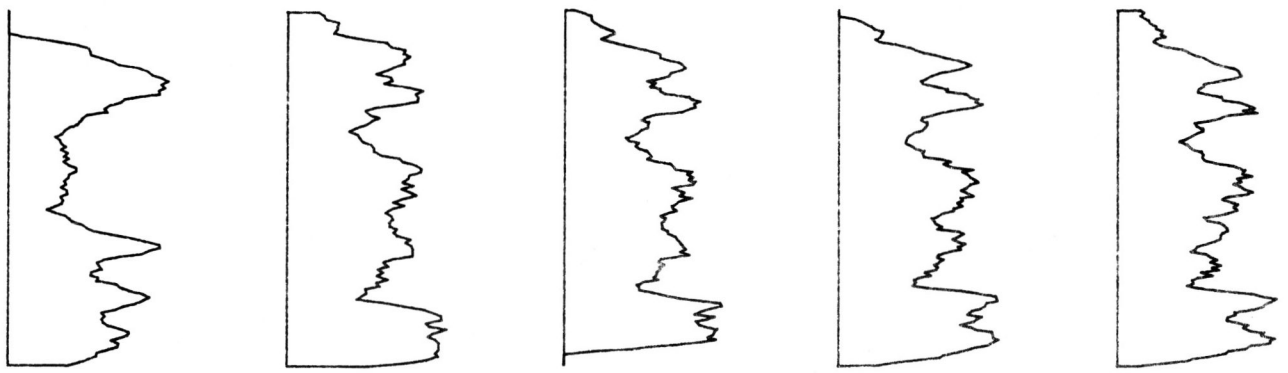

Figure 2. Integrated optical density profiles calculated using curves of Figure 1.