STEREO MATCHING OF CURVES

Andrew T. Brint & Michael Brady
Robotics Research Group,
Oxford University, OX1 3PJ.

This paper describes a stereo algorithm which matches connected chains of edgels (curves) between images. It is based on representing the curves as elastic strings and measuring the amount of deformation the strings have to undergo to transform between corresponding curves, and incorporates the ideas of the disparity gradient and the fact that matching sections of curve have to be of a similar shape. This explicit use of shape information means that a precisely known epipolar geometry is no longer crucial. Pairs of potentially corresponding curves which lead to a large deformation energy, are eliminated and the greatly reduced number of potentially matching pairs are passed on to a tree search stage. A typical result of running the algorithm on a stereo triple is presented.

In the past, edge pixels (edgels) have usually been the basic units used for matching in feature based stereo. However, this is computationally expensive because of the large number of pixels involved, and, more significantly, leads directly to the "correspondence problem" being considered the central problem in stereo, due to the number of possible matches for each pixel [1]. More recently, line segments have been used so as to reduce the number of primitives needing to be considered, thus alleviating both problems [2]. However, when dealing with curves, piecewise linear approximations are unstable so that the approximations can be significantly different between stereo images, and the number of straight line segments required to approximate a curve (to a pre-specified tolerance) grows with the curvature of the curve. The consequences of this can be seen in Figure 2 which shows the result of running a careful reimplementation of Ayache and Lustman's algorithm on a stereo triple, one of which is shown in Figure 1; this contrasts with the very good performance of the algorithm in a "linear" environment.

Also, although the correspondence problem is reduced, the low level of descriptive information contained in a straight line segment means that it is still non-trivial. Consequently, there has been growing interest in using curves as the match primitive. Some indication of the importance of curves in human stereo vision comes from the line drawing stereograms of Wheatstone [3] and his successors, where it seems that the *shapes* of the curves are crucial in determining the stereo correspondences (Figure 3).

Additionally, most stereo systems rely on a very tightly known epipolar geometry to resolve the correspondence problem. However, the calibration parameters of a mobile robot are likely to vary slightly with time and recently work has been carried out on recovering the new rotation from points matched between the images [4]. This paper describes an algorithm for matching curves between stereo images which is not heavily reliant on the calibration set up since it is based on the idea that curves in the world have very similar projections in the cameras. This similarity is exploited by measuring how much a curve in (say) the left image, has to be deformed to match a candidate curve in the right image. The algorithm can work with either binocular or trinocular sets of images, in the latter case the extra image is used to match lines coinciding with the epipolar directions in the first pair of images, and also as a check on the correctness of the binocular matches.

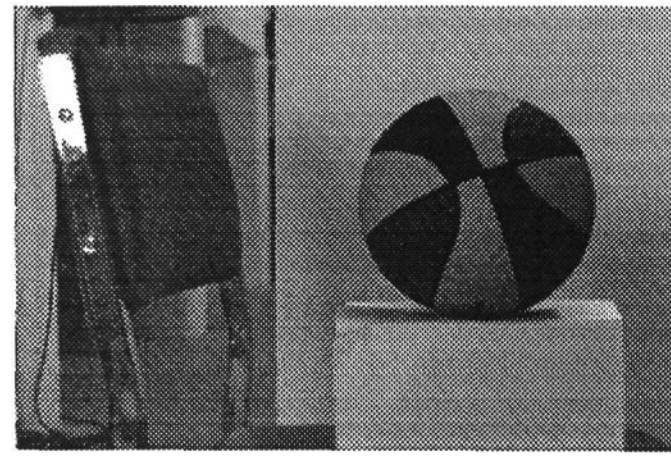

Figure 1: One image of a trinocular triple that is used for illustration. The image size is 512 by 340 pixels.

Using The Similarity Of Curves Between Images

Several algorithms for matching curves between pairs or triples of stereo images have been put forward by, amongst others, Keriven and Robert [5], Srinivasan et al. [6], Deriche and Faugeras [7], and Nasrabadi et al. [8], whilst Gazit and Medioni [9], and Schwartz and Sharir [10] and Wolfson [11] have considered, respectively, the related problems of matching curves in a motion sequence, and matching to a template. However, we feel that these algorithms do not adequately deal with the distortions in shape which occur in stereo.

The similarities in shape between the projections of the same world curve in stereo which the algorithm exploits, stem from (i) the nature of perspective projections, and (ii) the small distance apart of the cameras in the typical

AVC 1989 doi:10.5244/C.3.32

stereo system compared with the distance to the object. As an example of the former category, Verri and Yuille [12] have shown that the zeros of curvature of planar curves are preserved, and this result can be extended to show that the sign of curvature in the two images is the same unless the plane of the curve cuts the line segment joining the optical centres of the cameras. Unfortunately, maxima of curvature are not preserved but usually they do not move far between the stereo images. In addition to these pointwise properties, the relatively small baseline nearly always means that a section of the world curve undergoes a similar distortion when projected into each camera.

However, although projections of the same curve are similar in the different images, problems arise if they are approximated by say pieces of ellipses and then these pieces are matched between images, because of noise and the distortion produced by slightly different perspective views, leading to an unstable representation. Consequently, in the algorithm described here, each curve is represented as an elastic string/snake which can be locally stretched or shrunk by (respectively) decreasing or increasing the density of the curve's underlying parameter. (That is for a parameterised curve $\mathbf{r}(p)$, p varies less rapidly in "stretched" regions.) This provides a way of quantifying the difference in shape between similar curves by representing each curve as an elastic string and measuring the amount of stretching/shrinking needed to transform one curve into another.

This approach has similarities with the curve tracking work of Kass et al. [13], but, when dealing with the stereo problem, they only consider one contour in each image, and so all the emphasis is on reconstruction rather than on matching. Furthermore, we place a heavy emphasis on using shape information to bring two curves into correspondence by way of a stiffness term, whereas Kass et al. only use shape implicitly through their energy function (see equation 1 below).

Representing A Curve By An Elastic String

Each curve is represented by an energy function which is composed of a stretching energy term which measures the change in density of the curve's parameter, and a curvature energy term which encodes shape information by being defined so that points on the curve in higher curvature regions have more energy than those in lower curvature sections. The standard energy expression used by Kass et al. [13] for an elastic string/snake representing a parameterised curve $\mathbf{r}(p) = (x(p), y(p))$, is

$$E = \int_{P_0}^{P_1} \left(\left| \frac{d^2\mathbf{r}}{dp^2} \right|^2 + \lambda(p) \left| \frac{d\mathbf{r}}{dp} \right|^2 \right) dp \qquad (1)$$

where $\lambda(p)$ is a user defined variable which is usually taken to be a constant.

The density of the curve's parameter along the curve is obtained by minimising the integral of the energy over the length of the curve. The effect of this minimisation is to reparameterise the curve so that the parameterisation is more dependent on the shape of the curve than would have been the case from just using the arc length. This reparameterisation is sensible in stereo as the arc length is not the same between pairs of images but certain features of the shape of the curve are either preserved or at the least, very nearly preserved. In particular, in addition to moving their position very little between stereo images, maxima of curvature require a higher concentration of samples to describe them than areas of lower curvature where the rate of change is slower. (This greater amount of information present in high curvature regions formed the basis of Asada and Brady's Curvature Primal Sketch [14].)

The nature of equation 1 can be clarified by using the chain rule to differentiate $\mathbf{r}$ with respect to the arc length s rather than p, leading to

$$\frac{d\mathbf{r}}{dp} = \frac{d\mathbf{r}}{ds}\frac{ds}{dp} = \hat{\mathbf{t}}\frac{ds}{dp} \qquad \text{so} \qquad \frac{d^2\mathbf{r}}{dp^2} = \frac{d\hat{\mathbf{t}}}{dp}\frac{ds}{dp} + \hat{\mathbf{t}}\frac{d^2s}{dp^2},$$

where $\hat{\mathbf{t}}$ is the tangent direction. Therefore, as

$$\frac{d\hat{\mathbf{t}}}{dp} = \frac{d\hat{\mathbf{t}}}{ds}\frac{ds}{dp} = -\kappa\hat{\mathbf{n}}\frac{ds}{dp}$$

where $\hat{\mathbf{n}}$ is the normal direction and κ is the curvature, we have

$$\left| \frac{d^2\mathbf{r}}{dp^2} \right|^2 = \kappa^2 \left(\frac{ds}{dp} \right)^4 + \left(\frac{d^2s}{dp^2} \right)^2. \qquad (2)$$

The first term in equation 2 encourages points to concentrate near corners whilst the second term encourages ds/dp to vary as little as possible, thus counterbalancing the first term. Hence it is not possible to fully control how much emphasis is placed on corners by varying λ in equation 1. Several variations on equation 1 have been considered by Weiss [15] with the idea of placing extra emphasis on corners, the main one being replacing $d^2\mathbf{r}/dp^2$ by $d\hat{\mathbf{t}}/dp$. However, we place greater emphasis on maxima of curvature by eliminating the second term from equation 2 by taking the scalar product of the first term in the integrand of equation 1 with the unit normal $\hat{\mathbf{n}}$ before squaring it, that is

$$E = \int_{P_0}^{P_1} \left(\left| \hat{\mathbf{n}}.\frac{d^2\mathbf{r}}{dp^2} \right|^2 + \lambda(p) \left| \frac{d\mathbf{r}}{dp} \right|^2 \right) dp. \qquad (3)$$

The Euler-Lagrange equation (in conjunction with equation 2) for the minimisation of this integral is

$$\left(\lambda + 6\kappa^2 \left(\frac{ds}{dp} \right)^2 \right) \frac{d^2s}{dp^2} + 3\kappa\frac{d\kappa}{ds}\left(\frac{ds}{dp} \right)^3 = 0,$$

which formally shows that (assuming that λ is positive and that ds/dp is greater than zero to avoid degeneracy)

the density of the parameterisation does increase with the size of the curvature.

So far the energy formulation has been for the continuous case, however it is more convenient to work in discrete terms with a small number of points as the curves themselves are made up of a finite number of points, and as the minimisation of equations 1 and 3 becomes computationally easier. This can be done by placing points ("knots") at equal spacings of the parameter p. For notational convenience, we will assume that P_0 has the value 1 and that P_1 has the value n, then placing knots at the integer values of p means that equation 1 can be rewritten as

$$E = \sum_{i=1}^{n} E_{int}(i) \tag{4}$$

where $$E_{int}(i) = |\mathbf{r_{i+1}} - 2\mathbf{r_i} + \mathbf{r_{i-1}}|^2 + \lambda(\delta s_i)^2 \tag{5}$$

and $\mathbf{r}_i$ is the position of the i^{th} knot and (see Figure 4)

$$\delta s_i = \sqrt{(\mathbf{r}_i - \mathbf{r}_{i-1})^2}.$$

Using $\mathbf{r}_{i+1} - \mathbf{r}_{i-1}$ as an approximation for the direction of $\hat{\mathbf{t}}_i$, $\hat{\mathbf{n}}_i$ can be calculated and the discrete formulation of equation 3 takes the same form as equation 4 but with the $E_{int}(i)$ given by

$$E_{int}(i) = |\hat{\mathbf{n}}_i.(\mathbf{r}_{i+1} - 2\mathbf{r}_i + \mathbf{r}_{i-1})|^2 + \lambda(\delta s_i)^2. \tag{6}$$

Results For A Single Curve

The edge pixels marked by an edge detector are linked together into curves, and the knots are placed equidistantly along the curves. They are then repositioned on the curves using the dynamic programming approach of Amini et al. [16] so as to minimise equation 4.

The effect of the different energy formulations of equations 5 and 6 can be seen in Figure 5. The left curve displays the knots placed on a curve composed of the intersection of two straight lines when using the energy expression of equation 1, whilst the right gives the positions obtained from using equation 3 (with two knots being located in the corner).

Figure 7 shows the curves extracted by a simple feature point tracker after the basketball stereo triple have been processed by a Canny edge finder. Figure 6 shows the results of knot placement following energy minimisation for one of the curves shown in Figure 7. Figures 8 and 9 show the values at each knot of the two energy components in equation 3 for the curves in Figure 6 (where the starting knot is the topmost one). It can be seen that the stretch energy is low where the curvature energy is high and vice versa, as is to be expected from the above discussion.

Matching Of Curves

So far only individual curves have been considered, but when determining whether two curves are likely matches, points on one curve have to be associated with points on the other curve. This is carried out by defining an extra ("cross") energy term for each pair of corresponding knots on the two curves, which aims to measure the deformation needed to transform one curve into the other. This term is required because not only do we want to measure the amount of stretching/shrinking undergone in the transformation, but also to have a term corresponding to the "stiffness" of the curves. The minimisation process now tries to position the knots so that the sum of the individual curve energies (equation 6) and the cross energies is as low as possible. In more detail, the cross energy function is composed of two parts :- a disparity gradient term (which is a way of introducing stiffness to the string), and a similarity term (which encourages corners to match against corners).

The *disparity gradient term* is similar to that used by Kass et al. [13] except that we resolve into components parallel and perpendicular to the estimated epipolar direction, and also divide by the knot separations. The latter change is so that the term corresponds more closely with the notion of the disparity gradient idea from psychophysics which forms the basis of the PMF stereo algorithm [17]. Formally, the disparity gradient component is defined as

$$E_{disparity\ gradient}(i) = \frac{\beta\,|\hat{\mathbf{e}}^{\perp}.\mathbf{d}_i| + \gamma\,|\hat{\mathbf{e}}.\mathbf{d}_i|}{\delta s_{left}(i) + \delta s_{right}(i)} \tag{7}$$

where the difference between the changes in the left and right knot positions is

$$\mathbf{d}_i = \mathbf{r_{right}}(i) - \mathbf{r_{right}}(i-1) - \mathbf{r_{left}}(i) + \mathbf{r_{left}}(i-1),$$

$\hat{\mathbf{e}}$ is an estimate of the epipolar direction, $\hat{\mathbf{e}}^{\perp}$ is orthonormal to this direction, and β and γ are constants expressing the strength of the belief in the epipolar directions. The use of these different constants is so that, for instance, there is a greater penalty paid for vertical lines being of a different length between images, than for horizontal lines.

The *similarity term* is used in the energy expression to bias corresponding knots on the two curves to be in regions of similar shape. A measure of the shape of the curve at a knot is given by the sizes of the two components of equation 6 (Figures 8 and 9 show the component energies for the left and right projections of a world contour), and so the similarity term is defined to be the sum of the differences between these components for corresponding knots (suitably scaled so that it does not dominate the other energy terms).

The similarity and disparity gradient components have different relative strengths depending on the shape of the curve at that point. Near corners matching a "corner" knot against a "non-corner" knot leads to a high similarity energy which thus becomes the dominant term. However, in low curvature regions all the knots have similar component energies and so the disparity gradient energy has the relatively larger effect.

Implementation Of The Curve Matching

Potential matches for a curve are found by finding the curves in the second image which meet the (estimated) epipolar lines. For each of these potential matches, the curves are shortened using the (loose) epipolar geometry, so that their end points not only correspond, but are not on a section of curve which is parallel to the epipolar direction. If the shortened curves are of a sufficiently similar length (say within 30%), they are passed on to the main matching stage. Here the internal energies of the curves are minimised jointly with the cross energy.

After the joint minimisation, potential matches whose disparity gradient energy (given by summing equation 7 over the number of knots) divided by the number of knots, is above a threshold are eliminated. The remaining proposed correspondences are passed on to a search stage, along with a value indicating how likely the match is. At the moment, this value is the threshold value minus the cross energy divided by the number of knots, all multiplied by the number of contour points.

Currently, no use is made of the contrast across the curves (although use of this is envisaged in the near future), but the potential matches which give rise to negative disparities or disparities greater than a threshold (usually 150 pixels) are eliminated.

Tree Search And Trinocular Consistency Check

The curve matching stage although it returns very few matches for each curve and a measure of their likelihood, cannot eliminate all of the matching problem, if only because of cases where both images are composed of isolated, vertical lines. However, the number of remaining potential matches is now of a manageable size, and a tree search can be carried out with the decision at the i^{th} level being to accept or reject match i. The best branch is the one which maximises the sum of the scores from the curve matching stage whilst subject to

1. The uniqueness constraint - a piece of curve cannot have more than one matching section of curve.

2. The ordering constraint - if curve one is met before curve two when moving along an epipolar line in the first image, then the first curve's match is met before the second curve's match in the second image.

The tree search is implemented by forming a graph whose nodes are the potential matches and where two nodes are connected unless they break one of the above conditions, and then applying the standard clique finding algorithm of Bron and Kerbosch [18]. Unfortunately, if a large number of potential matches are passed on to this stage, the search can be computationally expensive. However, if junction information were available from the edge detector, this search could be improved in terms of speed and reliability and work is currently being carried out with this aim [19] (and the use of junction information seems to be a factor in the matching of Wheatstone's images [3]). Hopefully, this will mean that after highly distinctive curves have been matched, the matches can be propagated throughout the images greatly reducing the search requirements.

So far only binocular matching has been considered. The third camera is used to match curves which almost lie along one of the epipolar lines for the other cameras, and also as a check on the reliability of the binocular matches. The trinocular check entails determining whether the binocular matches are consistent by checking, for example, that if a right curve is matched to a curve in the left image and also to a curve in the top image, then the two resulting disparities should not be significantly different. As the curve matching essentially deals with the correspondence problem at the binocular level, this trinocular consistency check is a quick and effective means of eliminating false matches which may have slipped through the net.

Results And Concluding Remarks

Figure 10 shows the results of applying the algorithm to the curves shown in Figure 7. In addition to its better performance than Ayache and Lustman's algorithm on curved objects, it is less dependent on a precisely known epipolar geometry.

The current implementation of the algorithm takes several minutes to run, with virtually all of this time being spent in the minimisation of the curve energies and in the tree search stages. As was mentioned above, work is currently in progress to greatly reduce the amount of searching needed in this latter stage by making use of junction information. Whilst the time spent in the former stage could be significantly reduced by the use of a more selective screening procedure than just using the ratios of the curves' lengths, it seems that for near real time performance, specialist hardware such as that developed for dynamic time warping in speech recognition, would be required.

The next stage of this work will involve combining the stereo algorithm with motion by tracking the curves in the left, right and top images.

Acknowledgements

This work was funded by the Science and Engineering Research Council as part of ALVEY project MMI/149. Various members of the Robotics Research Group provided help, especially Andrew Zisserman and Alan McIvor. Bryan Rogers motivated the interest in Wheatstone's stereograms.

References

[1] **Marr,D. & Poggio,T.** "A computational theory of human stereo vision," *Proceedings of the Royal Society of London B*, Vol. 204 (1979) pp.301-328.

[2] **Ayache,N. & Lustman,F.** "Fast and reliable passive trinocular stereovision," in *Proceedings of the 1st International Conference on Computer Vision*, 1987, pp.422-427.

[3] **Wheatstone,C.** "Contributions to the physiology of vision - Part the first. On some remarkable and hitherto unobserved, phenomena of binocular vision," *Philosophical Transactions of the Royal Society*, Vol. 128 (1838) pp.371-394.

[4] **Takahashi,H. & Tomita,F.** "Self-calibration of stereo cameras," in *Proceedings of the 2nd International Conference on Computer Vision*, 1988, pp.123-128.

[5] **Keriven,R. & Robert,L.** "Un algorithme de vision stéréoscopique trinoculaire par appariement de chaînes," INRIA, July 1988.

[6] **Srinivasan,R., Ramakrishnan,K.R. & Sastryy, P.S.** "A contour based stereo algorithm," in *Proceedings of the 1st International Conference on Computer Vision*, 1987, pp.677-681.

[7] **Deriche,R. & Faugeras,O.D.** "2-D curve matching using high curvature points: application to stereo vision," INRIA, May 1988.

[8] **Nasrabadi,N.M., Liu,Y. & Chiang,J.-L.** "Stereo vision correspondence using a multi-channel graph matching technique," in *Proceedings of the IEEE Conference on Robotics and Automation*, 1988, pp.1804-1808.

[9] **Gazit,S.L. & Medioni,G.** "Multi-scale contour matching in a motion sequence," in *Proceedings of the DARPA Image Understanding Workshop*, 1989, pp.934-943.

[10] **Schwartz,J.T. & Sharir,M.** "Identification of partially obscured objects in two and three dimensions by matching noisy characteristic curves," *International Journal of Robotics Research*, Vol. 6 (1987) pp.29-44.

[11] **Wolfson,H.** "On curve matching," in *Proceedings of the IEEE Workshop on Computer Vision*, 1987, pp.307-310.

[12] **Verri,A. & Yuille,A.** "Perspective projection invariants," AI Memo 832, MIT, 1985.

[13] **Kass,M., Witkin,A. & Terzopoulos,D.** "Snakes: active contour models," in *Proceedings of the 1st International Conference on Computer Vision*, 1987, pp.259-268.

[14] **Asada,H. & Brady,J.M.** "The curvature primal sketch," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 8 (1986) pp.2-14.

[15] **Weiss,I.** "Curve fitting using a varying mesh," in *Proceedings of the IEEE Workshop on Computer Vision*, 1987, pp.311-314.

[16] **Amini,A., Tehrani,S. & Weymouth,T.** "Using dynamic programming for minimizing the energy of active contours in the presence of hard constraints," in *Proceedings of the 2nd International Conference on Computer Vision*, 1988, pp.95-99.

[17] **Pollard,S.B., Mayhew,J.E.W. & Frisby,J.P.** "PMF : a stereo correspondence algorithm using a disparity gradient limit," *Perception*, Vol. 14 (1985) pp.449-470.

[18] **Bron,C. & Kerbosch,J.** "Algorithm 457 Finding all cliques of an undirected graph [H]," *Communications of the ACM*, Vol. 16 (1973) pp.575-577.

[19] **Noble,J.A.** "Morphological feature detection," in *Proceedings of the 2nd International Conference on Computer Vision*, 1988, pp.112-116.

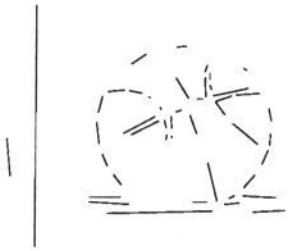

Figure 2: The matched line segments from running the Ayache and Lustman algorithm on the image shown in Figure 1.

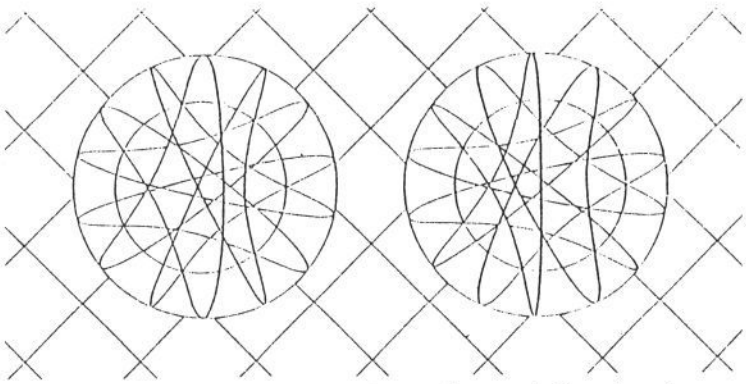

Figure 3: A Wheatstone influenced curved line drawing stereogram (published by Underwood & Underwood, London).

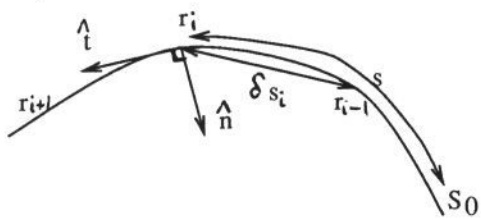

Figure 4: The notation used in Equations 5 and 6.

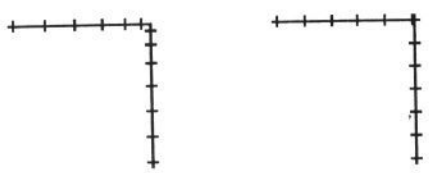

Figure 5: Different knot placements resulting from the two energy formulations discussed in the text. Left: the energy formulation of Kass et. al. Right: the energy formulation proposed in this paper concentrates energy at corners, implicitly making them act as match tokens.

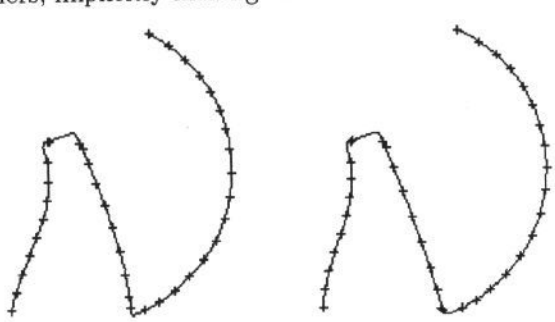

Figure 6: Knot placement along one of the curves extracted from respectively the left and right curve sets shown in Figure 7.

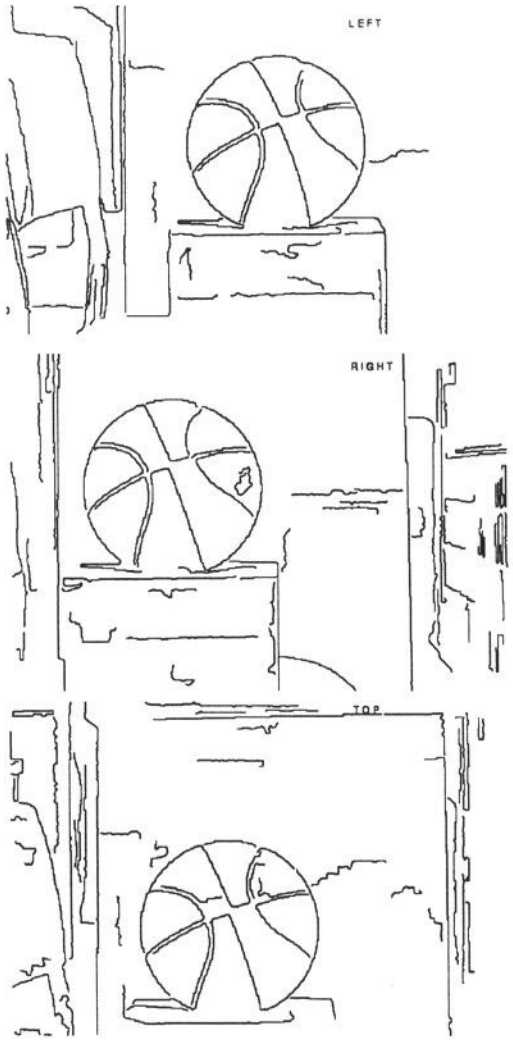

Figure 7: The curves extracted from the stereo images of the basketball.

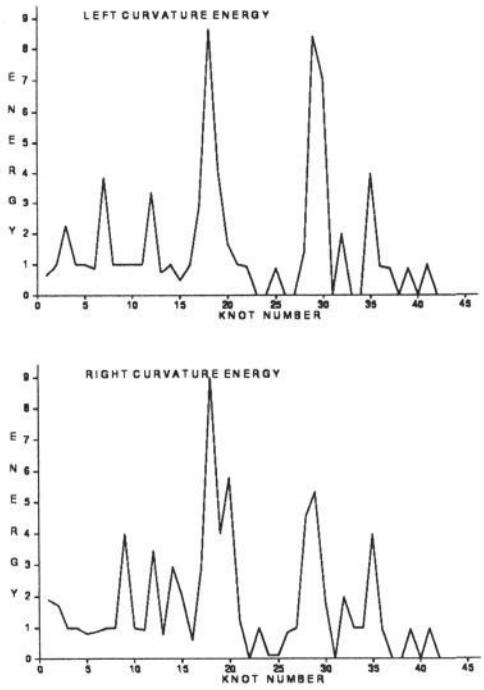

Figure 8: The final left and right curvature energies for the curve shown in Figure 6 when the knots are initially spaced 9 points apart.

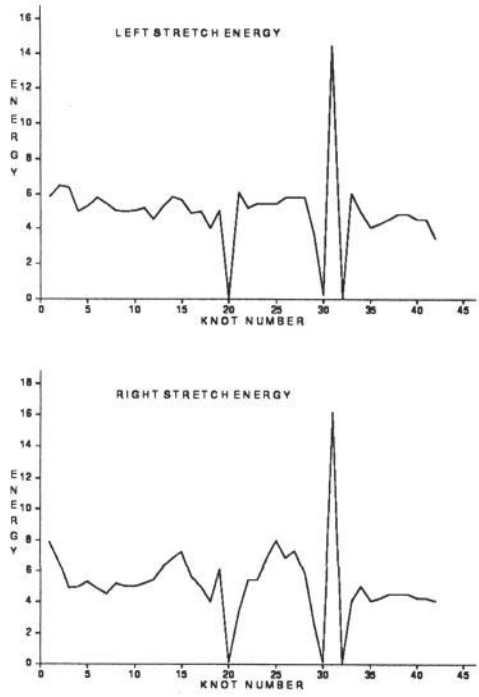

Figure 9: The final left and right stretch energies for the curve shown in Figure 6 when the knots are initially spaced 9 points apart.

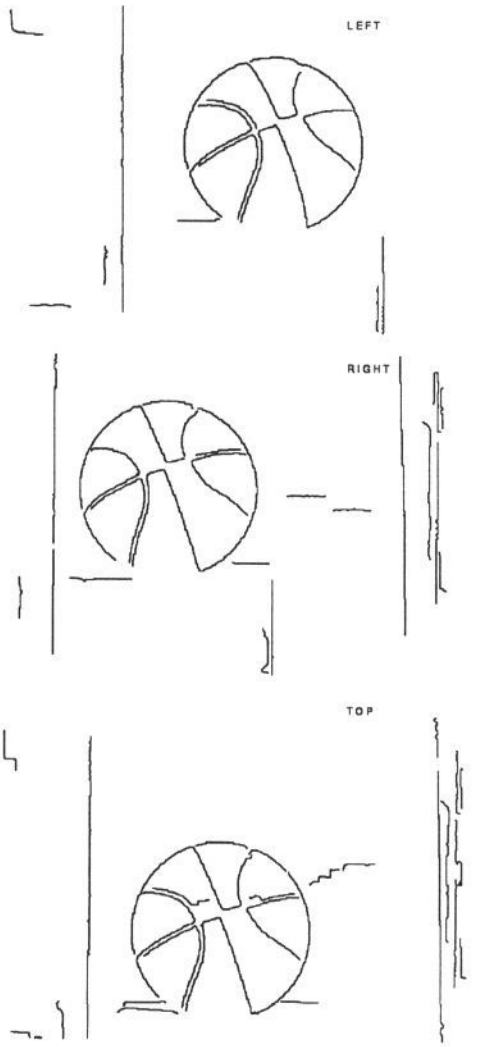

Figure 10: The curves matched from the stereo images of Figure 7.