Principal Component Analysis for the Approximation of an Image as an Ellipse

Sudanthi N.R. Wijewickrema
Computer Science & Software Engineering,
Monash University,
Clayton, VIC 3168, Australia
snw@csse.monash.edu.au

Andrew P. Papliński
Computer Science & Software Engineering,
Monash University,
Clayton, VIC 3168, Australia
app@csse.monash.edu.au

ABSTRACT

In this paper, we investigate a method of using principal component analysis(PCA) to fit an encapsulating ellipse to the image of a hypothetically ellipsoidal object. This technique is aimed at applications such as fruit sorting, where resource constraints and speed requirements necessitate the approximation of data.

Keywords

Ellipse Fitting, Principal Component Analysis (PCA)

1. INTRODUCTION

Principal component analysis (PCA) is a classical statistical technique which analyses the covariance structure of multivariate data [Hot33]. It determines the directions along which the variation of data occur and the corresponding importance of that direction. The first principal component gives the direction where the maximum variance could be observed. The second principal component is the direction of the next maximal variation and is orthogonal to the first and so on.

We take the image of an ellipsoidal object, represent it as a set of data points and use a PCA based algorithm to approximate it as an ellipse. Furthermore, we define a measure for calculating the error of fit, and outline a simple technique to determine the goodness of fit.

2. EXISTING WORK

Existing ellipse fitting algorithms could be categorized into Hough transform based methods and least squares fitting algorithms [HF98].

Hough transform based methods concentrate on mapping the data space to parameter space. Then the most likely parameters are chosen usually by clustering. They are robust against noise but have high computation time.

Least squares fitting looks at minimizing a distance measure between the set of points and the fitted curve. These methods are computationally better but are very sensitive to outliers.

Additionally, with respect to the target application, edge detection has to be done on the image to obtain the data points. In comparison, the method we propose only requires a simple segmentation of the image and needs only a few simple operations to calculate the ellipse parameters, making it ideal for resource constrained environments. A similar method is discussed in [Fan] where an ellipse is fitted to a region of interest.

3. METHOD OF CALCULATION

The proposed method of ellipse fitting is as follows:

1. Acquiring the set of data points
2. Calculating the covariance matrix
3. Solving the eigensystem
4. Fitting the ellipse

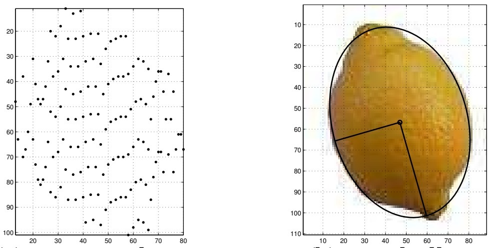
3.1 Acquiring the Set of Data Points

Before performing PCA, a set of data points should be extracted from the image. To simplify operations, the image is converted into gray scale first.

It is assumed here that the image is acquired under controlled conditions and this is easily satisfied in the application that we aim for: i.e. fruit sorting. Next, global thresholding is used to segment the object from the background to extract the set of points. A more sophisticated segmentation method could be used for uncontrolled systems.

Decimation of the points resulting from the segmentation is done next, to reduce the processing. We go for computational simplicity, selecting every d^{th} pixel from the image that falls inside the object. If d is co-prime with R or C, (the no. of rows or columns of

WSCG POSTER proceedings *ISBN 80-903100-8-7*
WSCG'2005, January 31 - February 4, 2005,
Plzen, Czech Republic.



(a) Decimated Image (b) Fitted Ellipse

Figure 1: An Example of a fitting

the image respectively, depending on the direction of selection: column-wise or row-wise), it should give good random uniform distribution. See figure 1(a).

3.2 Calculating the Covariance Matrix

After acquiring the data, the mean vector $\mathbf{c} = [c_1\ c_2]$ is calculated to get the center of the ellipse. Then the mean is removed to center the data around the origin. Let X consist of n 2-dimensional data points. The covariance matrix S, which indicates how strongly correlated the components are, is calculated as follows:

$$S = \frac{1}{n-1}\hat{X}^T \cdot \hat{X} \quad \text{where,} \quad \hat{X} = X - \mathbf{c} \tag{1}$$

3.3 Solving the Eigensystem

We calculate the two eigenvalues, λ_i and eigenvectors V_i of S as follows:

$$\lambda_{1,2} = \frac{1}{2}\left(\sigma_{11} + \sigma_{22} \pm \sqrt{(\sigma_{11}-\sigma_{22})^2 + 4\sigma_{12}^2}\right) \tag{2}$$

$$v_1 = \frac{\sigma_{12}}{\sqrt{(\lambda_1 - \sigma_{11})^2 + \sigma_{12}^2}}, \quad v_2 = \frac{\lambda_1 - \sigma_{11}}{\sqrt{(\lambda_1 - \sigma_{11})^2 + \sigma_{12}^2}} \tag{3}$$

where, $V_1 = [v_1\ v_2]^T$ and $V_2 = [v_2\ \ -v_1]^T$

3.4 Fitting the Ellipse

The mean of data, c_1, c_2, the eigenvalues $\lambda_{1,2}$, and the direction of the first eigenvalue, ϕ, uniquely define the ellipse [Hot33]. Eqn (4) shows this and figure 1(b) shows an ellipse, fitted to an image of a fruit.

$$x_1 + jx_2 = 2(\cos t + j\sin t)e^{j\phi} + (c_1 + jc_2) \tag{4}$$

4. ERROR & GOODNESS OF FIT

The error of fit determines how good a certain ellipse fitting algorithm is. For this, we define an error coefficient e, the normalized sum of squares of algebraic distances(D) from the edge points as shown in eqn (5).

$$e = \frac{1}{n}\sum_{i=1}^{n} D_i^2 \quad \text{where,} \quad n = \text{no. of edge points} \tag{5}$$

The error of fit changes with the decimation constant d. To generalize the selection of d, we define a value G, as shown in eqn (6), so that an optimal set of points that minimizes the processing and produces an acceptable error of fit can be selected. See figure 2.

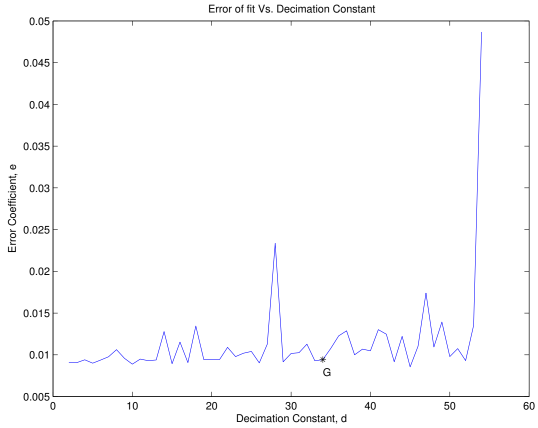


Figure 2: Error of fit Vs. Decimation constant

$$G = \text{round}(R \times \frac{3-\sqrt{5}}{2}) \tag{6}$$

The goodness of the ellipse fit can then be calculated to test the feasibility of the approximation. For this we propose a simple method to suit the target application. We define a threshold T, and take any fit that has an error measure below it as acceptable. The threshold is determined experimentally to suit the application.

5. CONCLUSION

The application of PCA to estimate the image of an object as an ellipse was investigated as a technique to be used in fruit sorting applications.

The processing here has to be done in real-time, and hence the algorithm and techniques used are aimed at simplifying the calculations and maximizing efficiency. For instance the use of a very simple segmention method was used to aquire the data points. This does not compromise the accuracy as the images are obtained under controlled conditions. Decimation of data was also done to achieve higher speed.

The above discussed method of representating an image of an object as an ellipse could therefore be considered as an efficient and suitable method to be used in fruit sorting in real-time and could be easily adapted for other similar applications.

6. REFERENCES

[Fan] Fit an Ellipse to a Region of Interest. www.dfanning. com/ip_tips/fit_ellipse.html.

[HF98] R. Halír and J. Flusser. Numerically stable direct least squares fitting of ellipse. In *Proceedings of the 6th Conference in Central Europe on Computer Graphics and Visualization*, pages 253–257, Plzen, February 1998.

[Hot33] H. Hotelling. Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24, 1933.