Efficient Region Segmentation through 'Creep-and-Merge'

Antranig Basman, Joan Lasenby and Roberto Cipolla

University of Cambridge, Department of Engineering, Cambridge CB2 1PZ, England

Abstract. We present a novel architecture for region-based segmentation of stationary and quasi-stationary statistics, which is designed to function correctly under the widest range of conditions. It is robust to the extremes of region topology and connectivity, and automatically maintains region boundaries sampled to the minimum scale at which the region configuration can be determined with statistical confidence. The algorithm is deterministic, and when operating on images from within its domain of validity, contains no adjustable parameters. In contrast to most other techniques directed at the same problem, the progress of the algorithm cannot be described by the optimisation of a global energy criterion.

We describe a specific implementation using Gaussian stationary statistics, and present test results which demonstrate superior performance to a collection of other systems.

1   Introduction

Many areas of computer vision could benefit from the replacement of its long–standing preference for local edge and corner detectors, based on linear correlation and convolution operations by more global methods, making use of region–based information. However, current region–based schemes suffer from drawbacks that make them unattractive for general use. These include:

Inflexibility : these schemes are often tied to a particular region modelling framework (MRF, piecewise polynomial, wavelet, etc.), which makes it difficult to adapt them for general use, or as new models arise. Schemes are also frequently [7, 3, 6] restricted to grayscale images.

Inefficiency : either through the performance of unnecessary extensive search through small scales of the image, or due to non-deterministic elements of the optimisation [1], region–based schemes traditionally take many orders of magnitude longer to run than localised feature detectors.

Non-regularisation : arguably the greatest challenge facing region optimisation is the regularisation of parameters, e.g. the boundaries, size and number of regions. Current schemes have difficulty accommodating regions with greatly varying size [12, 15], and often include explicit penalties for increasing the length of boundaries (a 'smoothing' regularisation) or increasing the number of regions. These introduce arbitrary parameters without supplying satisfactory algorithms to compute them for different situations.

The scheme we present here is flexible, efficient, and intrinsically regularises the necessary parameters in a dynamic manner during the course of optimisation; it thus requires no manual adjustments.

1.1 Motivation

A popular region–based segmentation paradigm (the "snakes/balloons" framework put forward in [9, 5], and generalised in [15]) idealises the image as a continuous field, and region boundaries as differentiable contours; optimisation then proceeds by steepest descent deformation of the contours with respect to some global energy criterion.

Another paradigm (split–and–merge/link, presented in [11, 2]) subdivides the image into discrete nested (square) cells, which are then recursively connected/split apart according to some homogeneity predicate [14].

The current scheme avoids the problems associated with these systems. Contour frameworks suffer from excessive locality (the image is only examined in a curve neighbourhood), troublesome discretisation – image 'forces' often involve curvature and gradient terms that are sensitive to quantisation artifacts – and topological difficulties which all restrict the scope of deformations.

On the other hand, splitting frameworks can overlook important features, since they proceed through a fixed number of recursive passes over the image.

Our framework may either be viewed as an explicit discretisation of a contour–based system, or as an iterative enhancement to region splitting methods – it shares in the good properties of both.

2 Theoretical Framework

We adopt a general modelling framework – the image is completely covered by a collection of non-intersecting connected regions $\{R_i\}$, which contain data D_i with some size measure $N_i = N_i(D_i)$. The region contents are drawn from $\alpha(D_i)$, one of a collection of models indexed by parameter vector α.

We assume the existence of an estimator $E(D_i)$ which determines, given a set of data, a 'best' model with vector $\hat{\alpha}_i$ from the collection. We place a commonplace restriction on these models:

Items of data from the modelled regions are viewed as independent, identically distributed (i.i.d.) random variables; i.e. the regions are those with some form of *stationary statistics*. This assumption is required at a point in the development mentioned in Section 2.3.

2.1 Statistical Framework

By design, the system proceeds by deterministic, minimal statistically significant discrete perturbations of the configuration, which at all times conforms with the description given above.

Which perturbations are, or are not, statistically significant, will vary during the course of the algorithm, and different sizes of perturbation will be appropriate according to the contents of the regions.

We will define this size in terms of a function $\delta(D_i, D_j)$, the *discriminability* of two sets of region contents.

The algorithm will be deemed to have converged when

1. No significant perturbations of a boundary between two regions will improve its discriminability.
2. Each boundary considered as a whole is significant.

We require that (i) δ has the dimensions of a log probability, and (ii) for fixed N_1, as $N_2 \to \infty$, $\delta \to \beta$, some $\beta \neq 0$, $-\infty$. This guides us to form δ using the hypothesis of homogeneity of two neighbouring regions; i.e. $H_\omega : \alpha(D_{12}) = \hat{\alpha}_{12}$, against the alternative hypothesis $H_a : \alpha(D_1) = \hat{\alpha}_1$ and $\alpha(D_2) = \hat{\alpha}_2$, where R_{12} with data D_{12} is the region formed by considering R_1 and R_2 one region. We choose δ to be the log level of confidence at which the null hypothesis H_ω of homogeneity would be rejected, if in fact we were required to make the decision.

This choice now constrains our estimators E to be the *maximum likelihood estimators* (m.l.e.s) of the parameters under the required hypothesis.

We now present the derivation of δ for a simple case; our parameters will be the means μ_i and standard deviations σ_i of univariate Gaussians $G(x; \mu_i, \sigma_i)$, and so $\hat{\alpha}_i = (m_i, s_i)$, where m_i and s_i are the standard (biased) maximum likelihood estimates of the parameters.

2.2 Univariate Gaussians

The log likelihood of data of size N drawn from a population $G(\mu, \sigma)$ under a hypothesis specifying $\mu = m$, $\sigma = s$ is

$$\log\left(\lambda(\mu, \sigma, m, s)\right) = -\frac{N}{2}\left(\log(2\pi\sigma^2) - \frac{s^2}{\sigma^2} - \frac{(\mu - m)^2}{\sigma^2}\right). \tag{1}$$

Consider two such regions, whose parameters have m.l.e.s m_i, s_i under the inhomogeneous hypothesis H_a, and $m_1 = m_2 = m_i^\alpha$, $s_1 = s_2 = s_i^\alpha$ under the homogeneity hypothesis H_α; the forms of these standard estimators are not presented here.

The likelihood ratio test (LRT) statistic for testing H_a against H_ω is obtained by summing three terms of the form of equation (1); it is given by

$$-\log \lambda^\alpha = N \log(s^\alpha)^2 - N_1 \log s_1^2 - N_2 \log s_2^2. \tag{2}$$

This quantity arises through the testing of a hypothesis allowing 2 degrees of freedom (the parameters m_i^α and s_i^α) as against a null hypothesis with 4. Thus, this statistic has an asymptotic (as both N_1 and N_2 become large) χ^2 distribution with 2 degrees of freedom. The reader is referred to Silvey [13], p. 114 for details of the argument, which is general for all tests of this sort.

Although formally this result is only applicable asymptotically, empirically it is highly reliable in this case for extremely small sample sizes.

Thus, the discriminability δ is defined as

$$\delta(D_i, D_j) = \log\left(1 - \int_0^{\log \lambda^\alpha(D_i, D_j)} \chi^2(2, x)\, dx\right) \tag{3}$$

i.e. the log level of confidence under which the data D_i, D_j would give rise to the rejection of H_ω.

A generalised derivation for multivariate Gaussians is found in [10], p. 140.

2.3 Optimisation Framework

A *border* B_{ij} between regions R_i and R_j is the set of points neighbouring both regions. Neighbouring a border B_{ij} are two sets L_{ij} and L_{ji} of *border elements*; these are (not necessarily non-overlapping) subregions of R_i and R_j, which are considered suitable candidates for exchange with the opposite region.

The minimum size of subregion below which the membership of R_i or R_j cannot significantly be determined is $N^*(D_i, D_j)$, the *magic number* for the border B_{ij}; members of L_{ij} must be larger than this.

To determine N^*, we consider subregions S_j^N of R_j with size N, assumed to give rise to the same model vector $\hat{\alpha}_j$ (not true in general, depending on the form of E – this is where we need the uniformity assumption mentioned in 2).

$N^*(D_j, D_i)$ is defined to be the largest N such that $\delta(D_j^N, D_i) > t_c$ where t_c is a level of confidence to be defined later.

3 Progress of Optimisation

In overview, the optimisation procedure has a three–phase structure:

Phase A – Seeding – the region configuration must be initialised in a suitable form, aiding efficiency and accuracy of convergence.

Phase B – Creeping – estimated regions are repeatedly deformed until no further significant deformation is possible.

Phase C – Merging – regions with weak (to be defined) boundaries are merged with the regions on the other side of the weak boundary.

Phase A is executed once, and alternating passes of Phases B and C are applied until both produce no further change.

3.1 Phase A

Phase A determines regions of maximum local discriminability from a selection of neighbours, where the maximum is taken through all scales of the image. This aims at allocating (a) at least one seed region to each actual image region (b) minimising seed regions straddling actual region boundaries.

The current system is rather generous in seeding regions, as figure 1(b), a seeding of image 1(a), shows. However, the remaining phases are quite robust; speed could be improved by seeding fewer, more well–positioned regions.

Each seed is converted into a region R_i, and the border configurations L_{ij} described in section 2.3 are initialised.
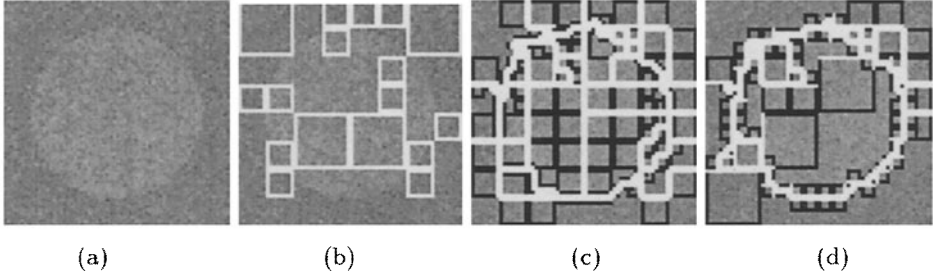
Fig. 1. (a) Original image; the regions differ in means by 15 units, and have common deviation 9. (b) Configuration after seeding; (c) After one phase of creeping; (d) After one phase of merging. The white boundaries represent regions R_i, black squares are border elements L_{ij}.

3.2 Phase B

Phase B applies operations analagous to the region deformation procedures of region growing/snake/balloon systems — in our system this procedure further generalises to the split–and–merge/link systems; we can decompose the deformation step as follows:

> Choose a border element L_{ij}^n from some border B_{ij}
> Form new regions R_i' and R_j' by exchanging it with R_j
> Evaluate the discriminability $\delta(D_i', D_j')$
> If it has increased, return L_{ij}^n to R_i; else leave with R_j.

Thus, although at a high level, the regions appear to be bounded by deformable contours, at a low level, the system is proceeding by deterministic split–and–merge. Some connectivitiy issues arise here, which there is not space to treat; for more details, see [4].

Figure 1(c) shows the configuration after Phase B is applied to figure 1(b)

3.3 Phase C

Region creeping and region merging proceed from a consistent and unified viewpoint. Creeping is a modification of the boundary configuration variables by significant increments (i.e. border elements). Merging occurs when these configuration variables as a whole are not significant, and *all possible efforts* have been made to improve their significance.

There are thus two criteria by which boundaries may be insignificant variables:

1. The discriminability $\delta(D_i, D_j)$ associated with the boundary as a whole is below the confidence threshold t_c.

<div align="center">OR</div>

2. Due to the geometry of the region, NO border elements L_{ij}^n of the requisite size as defined in section 2.3 can be formed.

Figure 1(d) shows the figure 1(c) configuration after Phase C.

3.4 Representation

The algorithm is not tied to a particular choice of representation.

Phase A depends on a hierarchical image partition, but the following operations may use e.g. the polygons or spline chains of section 1.1. However, in this implementation, a quadtree framework is used for all three phases, the best way to handle the rapid changes in connectivity and adjacency that occur in phases B and C. Cells C_i and border elements L_{ij} are therefore single quadtree cells.

3.5 Confidence

The confidence criterion t_c remains to be set; we use a base level of -5, giving a confidence level of 99.6%. A quadtree size N contains $2N/k$ pairs of adjacent cells area k; for a uniform error rate we set $t_c = -12 + \max\big(7, \log(\min(N_i, N_j))\big)$.

4 Evaluation

We demonstrate the power of the algorithm to adapt automatically to the extremes of scales/topologies present in images. Figure 2 is the final segmentation of figure 1(d) - the system correctly descends to the appropriate scale, and the circle is accurately localised after 3 iterations of phases B and C.
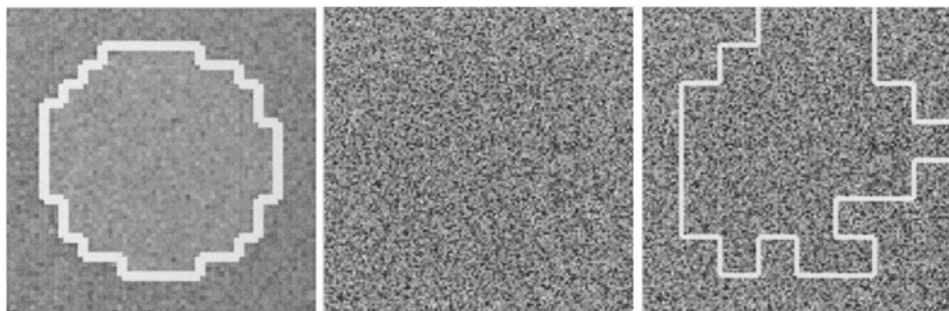

Fig. 2. Final circle image   **Fig. 3.** Original 4/32 image   **Fig. 4.** 4/32 result

4.1 Noise robustness

Figure 3 shows a severe test of the algorithm; a dark circle drawn from $G(124, 32^2)$ is successfully segmented (figure 4) from background $G(128, 32^2)$ (the magic number was artificially lowered one scale, leading to the irregular boundary). This level of noise is too severe for any contour–based technique, and indeed too severe for segmentation by the human visual system. The system continues to detect the circle until the difference in means drops below 2.

A split–and–merge system, however, only succeeds on an image where the difference in means is 5. The run time on 128x128 images of this sort is around 1 second, a factor of 2–3 slower than split–and–merge, which [8] shows to be the fastest of region segmentation schemes.

Figure 5 shows a systematic evaluation of the algorithm's breakdown under noise, using an evaluation criterion and results taken from [8]. The criterion

is a weighted sum of terms involving region size and boundary length. Our system (the dark line) is compared against five competing region– and edge–based systems, on varying noise images similar to figure 6.
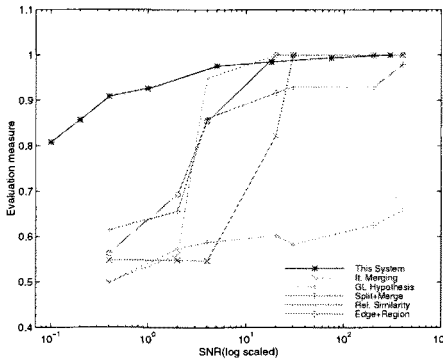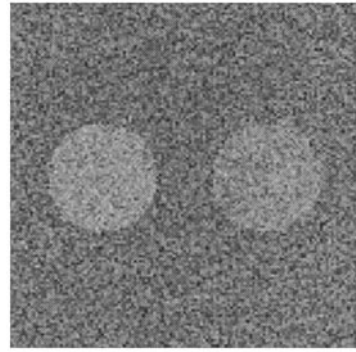

Fig. 5. Performance evaluation

Fig. 6. Typical image from [8]

Our method is considerably superior in high noise, and shows excellent performance in low noise also; the lack of smoothness regularisation leads to misclassification of a few pixels in moderate noise. Such a term could be added as post-processing; however, the score would degrade on images with corners.

We also note that the other systems' results are the result of optimising the criterion with respect to their adjustable parameters, over several dozen runs. Our system also scores heavily over the others in that it has no such parameters.
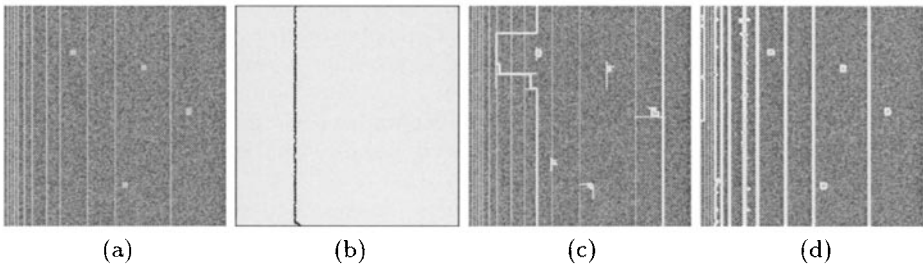

(a)   (b)   (c)   (d)

Fig. 7. (a) Original degenerate image; (b) Result from Yuille+Zhu system[15]; (c) Split and merge result; (d) Proposed system.

4.2 Scale adaptation

In the other direction, we show a successful segmentation of figure 7(a), taken from [15]; the algorithm correctly identifies the thin lines and blocks on the right–hand side and centre of the image; the left–hand strip is however still seen to be a single, high–variance region. Contrary to the claim made in [15], this success is made in a unified framework without the addition of any form of edge term. In order to segment the strip with closely-spaced lines, it will be necessary to handle pixels on an individual basis, which is inefficient in the current representation.

Both split–and–merge, and the system of [15], which claims to generalise and improve all contour–based methods, fail to segment this image.

5 Conclusions and Further Work

Our choice of criteria and optimisation procedure have been shown to resolve the problems of regularisation, boundary bias, automation, inflexibility and computational inefficiency from which region–based optimisation schemes often suffer.

Work will now turn to application of the multivariate version of the algorithm to the segmentation of coloured and textured images, and the development of appropriate surface models for the segmentation of images of real scenes.

References

1. P. Andrey and P. Tarroux. 'Unsupervised Texture Segmentation using Selectionist Relaxation'. In *Proc. 4th European Conf. on Computer Vision*, Volume I, pp. 482–491. Springer-Verlag, 1996. LNCS 1064.

2. H.J. Autonisse. Image segmentation in pyramids. *Computer Vision, Graphics and Image Processing*, 19(4):367–383, 1982.

3. J.A. Bangham, R. Harvey, P.D. Ling, and R.V. Aldridge. Nonlinear scale–space from *n*–dimensional sieves. In *Proc. 4th European Conf. on Computer Vision*, Volume I, pp. 189–198. Springer-Verlag, 1996. LNCS 1064.

4. A.M. Basman, J. Lasenby, and R. Cipolla. The Creep–and–Merge segmentation system. Technical Report CUED/F-INFENG/TR295, University of Cambridge, July 1997.

5. L.D. Cohen. On active contour models and balloons. *Computer Vision, Graphics and Image Processing*, 53(2):211–218, May 1991.

6. J.H. Elder and S.W. Zucker. Local scale control for edge detection and blur estimation. In *Proc. 4th European Conf. on Computer Vision*, Volume II, pp. 58–69. Springer-Verlag, 1996. Lecture Notes in Computer Science 1065.

7. S.A. Hojjatoleslami and J. Kittler. Region growing: A new approach. Technical report, University of Surrey, 1995.

8. H. Jiang, J. Toriwaki, and H. Suzuki. Comparative performance evaluation of segmentation methods based on region growing and division. *Systems and Computers in Japan*, 24(13):28–42, 1993.

9. M. Kass, A.P. Witkin, and D. Terzopoulos. Snakes: Active contour models. *Int. Journal of Computer Vision*, 1(4):321–331, Jan 1998.

10. K.V. Mardia, J.T. Kent, and J.M. Bibby. *Multivariate Analysis*. Academic Press, 1979.

11. T. Pavlidis. *Structural Pattern Recognition*, Chapter 5. Springer-Verlag, 1977.

12. P. Shroeter and J. Bigun. Hierarchical image segmentation by multi–dimensional clustering and orientation–adaptive boundary refinement. *Pattern Recognition*, 28(5):295–709, May 1995.

13. S.D. Silvey. *Statistical Inference*, Chapter 6. Penguin, 1970.

14. F. van der Heijden. *Image Based Measurement Systems*, Chapter 6. John Wiley & Sons, 1994.

15. S.C. Zhu and A. Yuille. 'Region Competition: Unifying Snakes, Region Growing, and Bayes/MDL for Multiband Image Segmentation'. *IEEE Trans. Pattern Analysis and Machine Intell.*, 18(9):884–900, September 1996.