ELSEVIER

Journal of Structural Geology 28 (2006) 363–375

JOURNAL OF STRUCTURAL GEOLOGY

www.elsevier.com/locate/jsg

Automated grain boundary detection by CASRG

Kingshuk Roy Choudhury [a,*], Patrick A. Meere [b], Kieran F. Mulchrone [c]

[a] Department of Statistics, National University of Ireland, Cork, Ireland
[b] Department of Geology, National University of Ireland, Cork, Ireland
[c] Department of Applied Mathematics, National University of Ireland, Cork, Ireland

Received 12 December 2005; accepted 14 December 2005
Available online 7 February 2006

Abstract

Constrained automated seeded region growing (CASRG) is an algorithm for automated grain boundary detection. It uses as input a single digitised microphotograph, such as ones obtained from a polarising microscope with an attached digital camera. In addition to this, it requires the user to click on the clasts within the microphotograph that the user wishes to obtain boundaries for. The algorithm requires no subsequent human input. The algorithm is based on the seeded region growing (SRG) algorithm of Adams and Bischof [Adams, R., Bischof, L., 1994. Seeded region growing. IEEE Transactions on Pattern Analysis Machine Intelligence 16, 641–647]. We have modified this algorithm to be guided by constraints and to adapt to the heterogeneity of colour information in the image. Imposition of these pre-determined additional conditions enables automated grain boundary detection without human intervention. The accuracy of CASRG has been validated through two benchmarking comparisons; one lithology with low tectonic strain and a second with high strain are used. The CASRG measurements are compared with those from hand drawn boundaries, which are used as a gold standard. Comparison is made using (a) a non-overlap statistic, (b) shape features, (c) strain estimates. In each case, the CASRG method compares very favourably with the gold standard.

Keywords: Grain boundaries; Seeded region growing; Adaptive threshold; Automated; Constraints; Non-overlap; Shape features; Strain analysis

1. Introduction

Strain analysis and the study of regional strain patterns are invaluable tools in interpreting the tectonic history of a region. However, in considering 12 recently published studies, a wide variability in sampling density for finite strain characterisation is demonstrated. The data in Table 1 indicates that, independent of the size of the study area, there is an upper limit of around 30 to the number of samples used. This results in very low sampling densities when the study area becomes large. There might be a number of reasons for this observation, e.g. homogeneity of deformation in an area, availability of suitable exposure, etc. However, we believe that the primary reason is the laborious and time consuming methods available for obtaining the raw data required for strain analysis.

There is one notable exception to the 30 sample limit provided by the study of Mukul and Mitra (1998). They analysed 119 samples of quartzite from an area of 200 km^2 around the Sheeprock Thrust Sheet, Sevier Fold-and-Thrust belt, Utah, USA. However, they employed a semi-automatic procedure for obtaining the data for strain analysis as described by Mukul (1998). In this paper we develop CASRG, a semi-automatic algorithm for strain analysis that enables rapid and accurate extraction of data for strain analysis. Automation of this process will allow strain analysis studies to break the 30 sample limit and introduce the possibility of statistical analysis of spatial strain variation, as in Mukul (1998). This paper concentrates on the problem of extracting data for strain analysis from sandstones and looks at deformed examples from the Variscides of southwest Ireland (Meere, 1995) and the Moine of northwest Scotland. The CASRG algorithm will yield data that is applicable to strain analysis methods based on marker shape (e.g. the mean radial length method of Mulchrone et al. (2003)) and to methods based on the relative position of markers such as those by Fry (1979) and Mulchrone (2003).

Traditional methods of measurement required sustained use of the polarising microscope with skilled manipulation of the rotating stage and knowledge of the use of various graticules (Ramsay, 1967, section 5.2). With the proliferation of digital cameras, it is now common place to obtain digital images of a field of view as seen through the microscope (Fig. 1a). Digital images are easily manipulated by computer graphic software

* Corresponding author. Tel.: +353 21 4902546.
E-mail address: K.RoyChoudhury@ucc.ie (K. Roy Choudhury).

0191-8141/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jsg.2005.12.010

Table 1
Area, number of samples and sampling density for a selection of recent studies which included strain analysis, at least in part

Author(s)	Study area (km^2)	No. Samples	Samples per km^2
Meere (1995)	70	23	0.32
Srivastava et al. (1995)	2.5×10^{-9}	4	1.6×10^{9}
Yin and Oertel (1995)	12	8	0.66
Mukul and Mitra (1998)	200	119	0.59
Roig et al. (1998)	500	8	0.02
Bresser and Walter (1999)	180	15	0.08
Althoff et al. (2000)	800	6	0.01
Hippert and Davis (2000)	12	9	0.75
Hippert and Davis (2000)	18	3	0.16
Hippert and Davis (2000)	4	6	1.50
Simancas et al. (2000)	60	28	0.46
González-Casado and García-Cuevas (2002)	2000	37	0.02
Mulchrone (2002)	55	18	0.33
Bailey and Eyster (2003)	14	8	0.57

packages that allow the outline of clast boundaries to be easily traced. Alternatively the boundaries may be manually traced from a printout and then scanned into a digital image (i.e. as suggested by Mukul (1998) and Mulchrone et al. (in review)). Given a set of such boundaries, it is possible to make the measurements required for strain analysis either manually or using automated methods (Mulchrone et al., in review).

Although methods that require manual identification of boundaries represent a striking improvement on totally manual methods for data extraction, there is room for further improvement. The primary aim of automatic clast boundary detection is to remove the manual step of marking boundaries. The task is especially onerous if one is engaged in a strain mapping study, where thousands of clasts have to be marked for reliable measurements of strain. Concomitantly, one is also seeking increased speed and accuracy in making strain measurements (Meere and Mulchrone, 2003). Speed is guaranteed not so much by the efficiency of the algorithm itself, as by the increase in processing power of computers. The issue of accuracy is of course paramount in any scientific endeavour. Given the nature of the current problem, clasts will always exist where manual marking of boundaries will be better than any automatic identification algorithm. In fact, given time, patience and practice, manual marking will be at least as good as the best automatic clast boundary detection algorithm. In practice, however, lack of dexterity with the mouse or pen can cause manually identified boundaries to deviate from the true boundary of the clast. In most cases, these errors will be negligible in terms of the accuracy of measurements made on the clast. Therefore, the aim is to develop a method that will deliver parameter estimates which are close (in an average sense) to those obtained by careful manual marking.

Previous work on automated clast boundary detection (e.g. Heilbronner, 2000; Ailleres et al., 1995; Bartozzi et al., 2000), demonstrates the difficulty of the problem. These papers address the difficulty by introducing extra information about the grain boundaries: for example Heilbronner (2000) has utilised multiple images of the same field of view and Bartozzi et al. (2000) use SEM images. Automatic clast boundary identification from a single image is an even harder problem. Clasts will be adjacent to each other and appear to be the same colour (e.g. in Fig. 1b, clasts 10 and 11). It will be very difficult to tell these apart. Fortunately, for strain measurement, we do not have to measure *all* clasts, but only enough for an accurate strain analysis (around 150 according to Meere and Mulchrone (2003)). In any given thin section image, there will be some clasts that appear well defined due to a sharp colour contrast with their immediate neighbourhood. It is the boundaries of these clasts that we will seek to identify.

2. Region based identification

Previous work on automatic clast boundary identification (e.g. Heilbronner, 2000; Bartozzi et al., 2000) utilise *edge-detection* based methods to identify the boundary of the clasts. Typically the initial boundaries produced by edge detection have many imperfections (some edges occur *within* the clasts as well as on the actual boundary and sometimes edges are absent on the real boundary). These initial boundaries are then post processed to obtain more realistic boundaries. While this approach appears to work satisfactorily for measurements such as clast count and clast size distribution, the subjectivity introduced by the post processing methods make them unsuitable for strain analysis, where crucial parameters are commonly the *physical* orientation of the clast (as opposed to

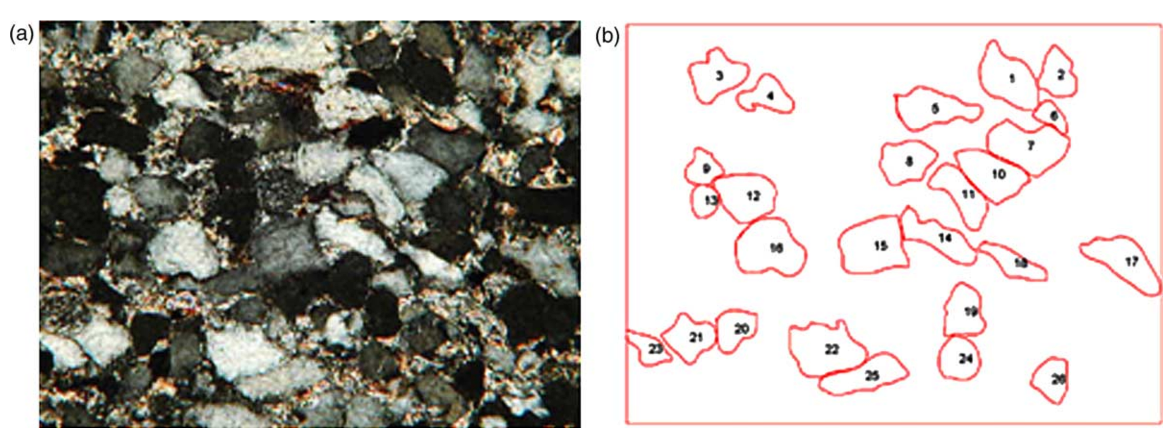

Fig. 1. (a) Original microphotograph. (b) Hand drawn boundaries.