Journal of Archaeological Science 37 (2010) 2671–2678

Contents lists available at ScienceDirect

Journal of Archaeological Science

journal homepage: http://www.elsevier.com/locate/jas

Practical quantitative lithic use-wear analysis using multiple classifiers

Nathan E. Stevens [a,*], Douglas R. Harro [b], Alan Hicklin [c]

[a] Department of Anthropology, University of California, Davis, One Shields Avenue, Davis, CA 95616, USA
[b] Applied Earthworks, Inc. 515 E. Ocean Ave, Suite G, Lompoc, CA 93436, USA
[c] Spectral Imaging Facility, NEAT ORU, Chemistry Rm 11, University of California, Davis, One Shields Avenue, Davis, CA 95616, USA

A R T I C L E I N F O

Article history:
Received 21 April 2010
Received in revised form
2 June 2010
Accepted 2 June 2010

Keywords:
Lithics
Use-wear
Microwear
Laser scanning confocal microscopy
Classification

A B S T R A C T

Although use-wear analysis of prehistoric stone tools using conventional microscopy has proven useful to archaeologists interested in tool function, critics have questioned the reliability and repeatability of the method. The research presented here shows it is possible to quantitatively discriminate between various contact materials (e.g., wood, antler) using laser scanning confocal microscopy in conjunction with conventional edge damage data. Experiments with replica and prehistoric tools suggest the quantitative method presented here provides valid functional inferences and is flexible enough to accommodate other relevant sources of data on tool function.

1. Introduction

For several decades, archaeologists have had an uneasy relationship with lithic use-wear analysis. While the data it provides are useful, the subjectivity of the method, which often relies more on experience and expertise than explicit criteria, has left many archaeologists wary. This problem has been addressed in two ways: 1) blind tests to explore accuracy and interobserver error (Bamforth, 1988; Newcomer et al., 1986; Odell and Odell-Vereecken, 1980), and 2) attempts at quantification (Evans and Donahue, 2008; Gonzalez-Urquijo and Ibanez-Estevez, 2003; reviewed in Grace, 1996; Keeley, 1980; Kimball et al., 1995; Stemp and Stemp, 2001, 2003; Van den Dries, 1998). Similar issues and solutions have developed in parallel among researchers studying dental microwear in primates (Scott et al., 2006).

Laser scanning confocal microscopy (LSCM) is a promising method for use-wear analysis because it produces 3-dimensional point data that can be presented either as a high-resolution image (Fig. 1), or as quantitative data (Evans and Donahue, 2008). Making use-wear analysis quantitative rather than qualitative will not only make the process of producing functional inferences more explicit, it will also provide a basis for further methodological improvements. Tests using both human analysts and quantitative classification (i.e., "machine learning" in the language of artificial intelligence, see Alpaydin, 2004) can be used to explore the relevant variables affecting the accuracy of the method including experience (Bamforth et al., 1990), materials (Bradley and Clayton, 1987; Lerner et al., 2007), use duration (Bamforth, 1988; Goodale et al., 2010), and post-depositional processes (Burroni et al., 2002; Evans and Donahue, 2005; Levi Sala, 1986).

If different contact materials have distinctive quantitative signatures, then it should be possible to represent use-wear data in the form of probability statements that report not only how a tool was used, but also provide information about the certainty of the attribution, something that has not been possible using traditional methods. Additionally, if the results of human and machine use-wear experts converge for both replica and archaeological datasets, then many previous use-wear studies using conventional methods can be validated.

2. Background

Blind tests typically involve analysts trading replica tools to test their ability to identify: 1) which tools or tool edges were used, 2) the tool motions employed, and 3) the contact materials on which tools were used. While such tests have generally produced satisfactory results, some high-profile published tests with poor results (e.g., Newcomer et al., 1986) led many archaeologists to doubt the efficacy of use-wear analysis.

A common misunderstanding of use-wear studies is that there is one "correct" answer and that anything less is not useful. In fact, the

* Corresponding author. Tel.: +1 805 710 3486; fax: +1 530 752 8885.
E-mail addresses: nestevens@ucdavis.edu (N.E. Stevens), dharro@appliedearthworks.com (D.R. Harro), aghicklin@ucdavis.edu (A. Hicklin).

0305-4403/$ – see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jas.2010.06.004

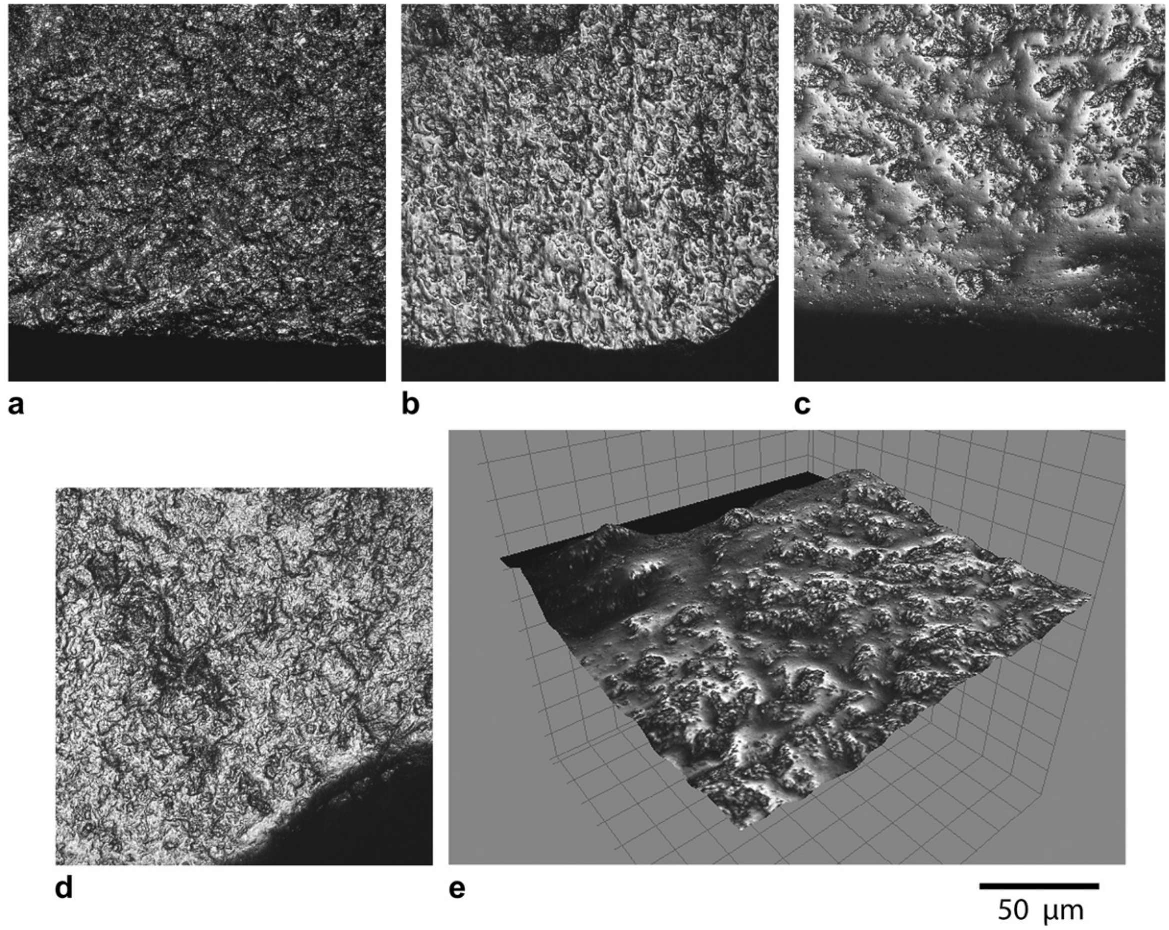

Fig. 1. LSCM images of tool edges. a: unused replica, b: replica used for scraping wood, c: replica used for cutting soft plants, d: prehistoric tool (981-3) classified as used on hard material (wood), e: 3D projection of replica tool depicted in c.

specificity of functional data needed depends on the archaeological question. In some cases, it may be enough to know simply which items are tools and which are waste flakes. In other cases, knowing whether tools were used on hard or soft substances may suffice. In still other cases, it may be desirable to know specifically the numbers of tools used to cut meat, or scrape wood.

Independent of the desired specificity of functional data are a number of factors largely beyond the control of the analyst that affect the quality of the data including raw material type, post-depositional alteration, available equipment, and others. The problem is that it is difficult to convey information about how certain an analyst is in a particular identification. It is also unclear when this lack of certainty should lead to less specific attributions (e.g., “indeterminate hard material”) or an admission by the analyst that the contact material is unknown. This problem is analogous to that of faunal analysts deciding when to identify material to the species, genus, or higher taxonomic levels owing to fragmentation, expertise, and other factors (e.g., Gobalet, 2001). These issues can create real differences in how use-wear data are produced by different analysts and interpreted by the archaeological community.

Blind tests of both the “high-power” and “low-power” use-wear approaches (see Odell and Odell-Vereecken, 1980) have shown that accuracy varies predictably with level of specificity (Bamforth, 1988). Accuracy is quite high (ca. 70–90%) if only the presence or absence of use wear is examined. Accuracy is lower (ca. 60–75%) when discriminating between general contact material classes (e.g., hard vs. soft), and lower still (ca. 20–70%) for specific contact materials (e.g., antler, meat). Low power methods, in particular, have reduced success at discriminating specific contact materials but do well at discriminating general material classes. (Odell and Odell-Vereecken, 1980).

Conventional use-wear studies have shown that identification of specific contact materials can be improved by combining more than one type of data such as polish appearance at high magnification in addition to edge damage at low magnification (Bamforth, 1988; Keeley, 1980). This approach was explicitly implemented by Grace (1989) and Van den Dries (1998), who used a combination of polish and edge damage variables obtained by conventional microscopy in the construction of expert systems designed to identify tool functions. More recently, studies using sophisticated instrumentation including atomic force microscopy (Kimball et al., 1995), laser profilometry (Stemp and Stemp, 2001), and LSCM (Evans and Donahue, 2008) have demonstrated quantitative differences between different polish classes, but have used only single descriptors (e.g., Rq, or root mean square roughness) and have not incorporated other sources of data on tool function such as edge damage. Our aim was to combine quantitative data on use-wear polishes (acquired at magnification equivalent to 1000×, using the LSCM) with qualitative data on edge damage (observed at 20×–100×, using a stereomicroscope) to arrive at a multivariate approach to classification incorporating the best attributes of each.

This was accomplished by following the lead of researchers working in the fields of artificial intelligence, pattern recognition, and machine learning, who have advocated an approach to