Shoe Print Recognition, an Introduction (UG4:AV)

John Hussey

April 14, 2009

1 Introduction

The use of the shoeprint in crime scene investigation can be a very useful tool. Identifying the same shoeprint at different crime scenes allows crimes scenes to be linked. Similarly, suspects can be linked to crime scenes. Since the Serious Organised Crime Police Act (2005)[1] the shoe print is now recognised as having the same legal status as DNA and fingerprint evidence [7].

Shoeprint information is gathered from crime scenes by photography, a cast or electrostatic lifting. This impression is then used to arrive at three main types of classification. Firstly there is the basic match in shoe size, secondly matching a shoe print to a particular brand or model. This can be achieved by keeping a database of product information from shoe manufactures (this data is abundant) and then matching the crime scene print against the catalogue. Thirdly there is the task of exactly matching a shoe print found at the crime scene with that of the shoeprint of a suspect. This can only be achieved with confidence if the crime scene shoeprint is captured in sufficiently high quality.

Until recently the approach used in shoeprint recognition involved manually sifting through a large database of captured shoe prints, trying to find a match with the print found at the crime scene. This is time consuming and is presumably very tedious. Given the repetitive nature of the task and the subtle differences between shoeprints, for example a slightly worn left heel, the task is vulnerable to human error. The noise inherent in many shoe print samples compounds the problem.

In light of these considerations, the automation of the classification process becomes very attractive. The problem of shoeprint recognition can be divided into two areas, the actual retrieval of information from a crime scene and the identification of scene evidence with a stored reference. Successful approaches to the latter of the two areas are explored below.

2 Approach

Assuming the shoe print has been captured digitally by a forensic scientist at the crime scene, several tasks remain. These are; noise removal, feature detection, appropriate encoding of feature information and finally, classification.

2.1 Noise Removal

Captured shoe prints are very delicate and are not very robust to traditional noise removal algorithms. For example it is unacceptable to use a basic background thresholding algorithm. There are many subtle features that will not be preserved by such an approach as a threshold used in one part of the image may be unsuccessful at thresholding another part of the image. It is important to remove the noise on the shoe print without destroying the information encoded in the shoe print. An approach developed by the Image and Vision Group at Queen's University Belfast attempts to do just this [3].

The technique implemented used a robust denoising technique based on non-local mean filtering. The image is divided into pixel regions. The effect of a particular operation on a pixel region is calculated by first applying that operation to pixel regions that are sufficiently similar. This calculates the expected result of an operation. Each pixel is classified as foreground or background by using the expected result of an operation on the local pixel.

2.2 Feature Detection

Given a clean shoe print, the next task is to extract the relevant features that describe that shoe. For example many trainers have distinctive shapes on the sole. Some shoes will have been worn in such a way that will have eroded the toe or heel. It is desirable that such features can be automatically identified. Recent research highlights Maximally Stable Extreme Region (MSER) [6] feature detection and Harris-Affine region detection as being particularly sound algorithms for the detection of shoe print features [5]. Both of these methods are invariant to rotation, scaling and translation. This is essential as without this constraint, accurate comparisons between shoe features are unattainable.

2.2.1 Harris-Affine

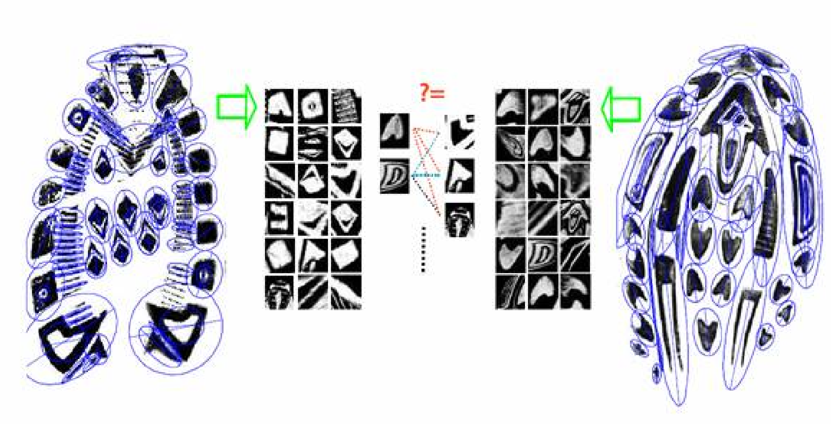
The HA region detector is a feature detection algorithm. It uses the Harris corner measure [2] which seeks areas of the image where the image intensity changes largely in multiple directions. This lends it to the detection of features such as grooves and worn areas of shoe prints. However it is important to note that the MSER method is "better suited to discriminating general patterns"[5]. The HA algorithm can be broken down into 5 main steps:

1. Initial region points are detected with the scale-invariant Harris-Laplace Detector.
2. Each point is normalised to be affine invariant.
3. Iteratively estimate the affine region where a point is localised.
4. Update the affine region using these scales and spatial localizations.
5. Return to 3. until stopping criterion is met.

The components of the algorithm are detailed however the idea at the core of it is that at a corner feature, the image intensity will dramatically change when the focus is shifted in an arbitrary fashion to another close part of the image.

2.2.2 MSER

MSER is a segmentation algorithm that works best when identifying similar features separated by well defined boundaries. This makes it ideal for classifying many types of shoe print. The image depicts the features extracted from two shoeprints when using MSER [5].



2.3 Encoding

Once the shoe print's local features have been extracted they must be encoded in such a way that is convenient for classification. Scale invariant feature transform (David Lowe)[4] is able to describe local features in a manner that is amenable to classification. The SIFT algorithm takes an image, in the case of shoe print recognition a feature extracted from the shoe, and compresses it into a vector of SIFT features. The sift features are based on local features within the image. The sift features are affine invariant, that is to say that the features are preserved under linear transformations. Once the image has been converted into SIFT features it is traditionally the case that classification would be implemented by using a nearest neighbor approach, comparing the test example with the SIFT translated training examples. However, in the case of the shoe print recognition the classifier component of the SIFT approach is not desirable. This is because one individual shoe is made up of many SIFT feature vectors, making it inappropriate for use with the one nearest neighbor method. Also by only encoding the local information associated with the shoe (particular features on the shoe) the SIFT algorithm does not take into account important factors, such as how those features are related to each other. To reiterate, the SIFT features are only used as a method of encoding the shoes features, it is not used for classification.

2.4 Classification

Having extracted the shoes features and appropriately encoded them, the final task involves finding a correspondence between shoe prints recorded from the crime scene and the shoeprints already stored. However many distinct shoes share similar features. Therefore it is necessary to use the geometric position of features to disambiguate potential matches. There are three main techniques for finding the correct correspondences. These include point pattern matching, graphical models and spectral methods. Point pattern matching attempts to identify whether a particular arrangement of points occurs in two images. However the point pattern matching and spectral methods fail to take into account the additional information that is necessary when trying to identify similarities between shoe prints, for example the geometry of the shoe (where particular features are positioned).

In [5] the positional information is encoded with proximity matrix of selected feature locations. A feature similarity matrix is also used. This is created by taking the feature descriptors (SIFT vectors) F_a and F_b of two images. By applying a Gaussian function against all pairs of features, a Gaussian similarity matrix can be formed (illustrated below).

$$K(i,j) = \exp((-\|F_i - F_j\|^2) \div 2\sigma^2)$$

This is then multiplied with the proximity matrix to encode feature information combined with positional information that gives an overall similarity between two shoe prints. This rating is used to rank possible shoe matches.

2.5 Summary

Assuming that the system has a photo of a shoe print as an input and a database of reference examples where we try to find a match a shoeprint recognition system would have four main components. These components would be noise removal, feature extraction, feature encoding and finally classification.

References

[1] *Serious Organised Crime Police Act 2005 (Part 3 section 118 of this document).* 2005.

[2] Konstantinos G. Derpanis. *The Harris Corner Detector.* York University http://www.cse.yorku.ca/~kosta/CompVis_Notes/harris_detector.pdf, 2004.

[3] H.Su, D.Crookes, and A.Bouridane. *Thresholding of Noisy Shoeprint Images based on Pixel Context.* Pattern Recognition Letters 28(2): 301-307, 2007.

[4] David G Lowe. *Object recognition from local scale-invariant features.* International Conference on Computer Vision, 1999.

[5] M.Pavlou and Nigel. M. Allinson. *Automatic extraction and classification of footwear patterns.* Springer Berlin/Heidelberg, 2006.

[6] Andrea Vedaldi. *http://vision.ucla.edu/~vedaldi/code/mser/mser.html.*

[7] David Wark. *Shoe Print Recognition.* CVonline: On-Line Compendium of Computer Vision [http://homepages.inf.ed.ac.uk/rbf/CVonline/]. R. Fisher (ed), 2008.