ESANN'2004 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 28-30 April 2004, d-side publi., ISBN 2-930307-04-8, pp. 549-554

Using classification to determine the number of finger strokes on a multi-touch tactile device

Caspar von Wrede and Pavel Laskov

Fraunhofer FIRST.IDA
Kekuléstr. 7, 12489 Berlin, Germany
{wrede,laskov}@first.fhg.de

Abstract. On certain types of multi-touch touchpads, determining the number of finger stroke is a non-trivial problem. We investigate the application of several classification algorithms to this problem. Our experiments are based on a flat prototype of the spherical Touchglobe touchpad. We demonstrate that with a very short delay after the stroke, the number of touches can be determined by a Support Vector Machine with an RBF kernel with an accuracy of about 90% (on a 5-class problem).

1 Introduction

Tactile interfaces constitute a dynamically developing field in human-computer interaction research and applications. They are no longer primarily seen as the means of communication for visually impaired people but rather as a possible technology for a fast and ergonomic computer interface. Further impetus for the development of tactile interfaces was given by the emergence of the notion of "wearable computing" in which touchpads play the key role as an input device.

The hardware technology underlying the design of touchpads can roughly be categorized in two classes: the projective sensor matrices and the sensor arrays. In the former, used in the majority of commercially available single-touch touchpads, e.g. Synaptics TouchPad™[1], the input is generated by the sensors corresponding to the directions on the active area. The advantage of the projective sensor touchpads is their simplicity and the small number of sensors required (proportional to roughly the square root of the active area). While adequate for detection of one touch, the projective sensor matrices suffer from ambiguity arising when multiple touches are made [8]; at certain locations two touches cannot be distinguished from three touches etc. The ambiguity problem

The authors gratefully acknowledge partial support from the PASCAL Network of Excellence (EU #506778). We also thank Klaus-Robert Müller for valuable discussions and advice.

ESANN'2004 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 28-30 April 2004, d-side publi., ISBN 2-930307-04-8, pp. 549-554

can be solved for three fingers by using special disambiguation techniques [2, 7], or, similarly to computer tomography, by using multiple scanning directions [9, 13], but no solution is currently known for more than three fingers. The sensor arrays utilize touch-sensitive sensors allocated throughout the active area of a touchpad [6, 3, 17]. While the sensor array technology is constantly improving, with some new commercial products, e.g. by Tactex Controls Inc. [15] and Fingerworks Inc. [5], available on the market, it increases hardware complexity and has potentially lower spatial resolution.

In this paper we propose using machine learning technology to alleviate the ambiguity problem of projective sensor touchpads. In particular, we present the techniques for learning to classify the number of fingers applied to a touchpad (one through five; for one hand only). We show that with relatively simple preprocessing, the classification accuracy of about 98% can be attained on a single-class classification of the number of touches (for example one touch against all other number of touches), or about 90% accuracy can be attained on multi-class classification, i.e. determining how many fingers are pressed at a single time instance. The proposed methods are online in the sense that they yield an answer shortly after pressing the touchpad (with about 200ms delay).

2 The prototype touchpad

We use a prototype of a Touchglobe spherical touchpad [12] for the application of machine learning techniques. The hardware of the touchpad is built as a criss-crossed pattern of 100 sensor channels layed out along both the rectangular and the diagonal directions.

The flat version of the touchpad we used consists of an array of $30 \times 20 = 600$ taxels attached to 100 sensor channels. Combination of rectangular and diagonal directions allows for preprocessing to be applied in order to somewhat alleviate the ambiguity. If one of the four channels attached to a taxel has a zero signal, the value at a taxel is set to zero; otherwise the value at a taxel is computed as $\sqrt[4]{\prod_{i=1}^{4} c_i}$ of the channel signals c_i. As a result of such preprocessing a graphical image can be obtained showing the intensity of pressure at certain areas on a touchpad. As another data format we also consider the direct channel readings. These two data types are referred to as graphical and raw data. Figure 1 shows 25 exam-

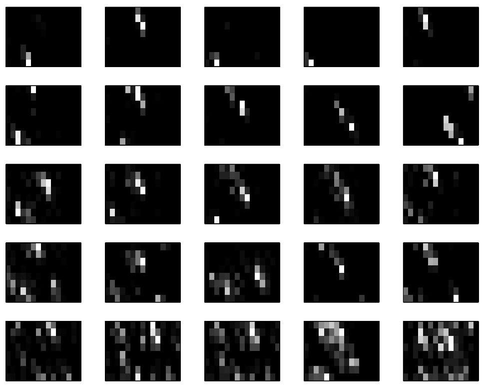

Figure 1: Graphical data samples. The first row shows samples from class one, the second from class two, and so on.

ESANN'2004 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 28-30 April 2004, d-side publi., ISBN 2-930307-04-8, pp. 549-554

ples of graphical data samples, 5 of each class (one through five fingers pressed) in each row.

3 Data collection and preprocessing

At this point we are concerned only with static gestures, i.e. touching the device with a certain number of fingers and releasing the fingers without moving them along the touchpad. Data was collected by following this procedure, at irregular time intervals, 250 times for each class. The resulting multidimensional time series, single vectors being either channel values or taxel values from the graphical representation, had to be further preprocessed to identify the time interval during which the touchpad is pressed and to obtain the features suitable for training a classifier.

3.1 Segmentation of the continuous time series

Segmentation is essential for applying classification during the operation of a touchpad since for it to be used for disambiguation, the output should be obtained as soon as possible after the touch has occurred. To perform segmentation, we use the energy of the signals, which we define as the sum of the absolute values across all channels at one time instance. Figure 2 shows the signals recorded in the data channels of both type—raw and graphical—and their respective energies.

One can see that the onset of the touch is easily detected by thresholding the rising front of the energy signal (with some heuristic rules to prevent jitter). Detection of the release of the fingers is somewhat more difficult, but since we are interested in the classification of the gesture as soon as possible, we cannot wait for the release event anyway.

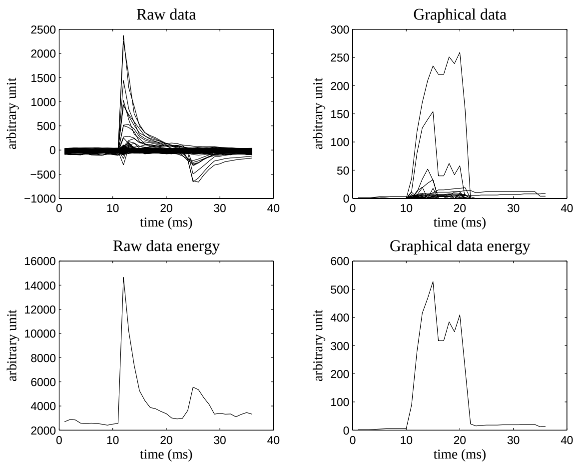

Figure 2: Two different data types (raw and graphical) of the same sample along with their respective data energies.

3.2 Feature selection: temporal aggregation

The features to be used in classification are computed by adding up the values in each channel over the window of size 10, after the onset of the touch is

ESANN'2004 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 28-30 April 2004, d-side publi., ISBN 2-930307-04-8, pp. 549-554

detected. Besides ensuring the swift response of the classifier (the delay before the classification is applied amounts to 50ms at the sampling rate of 200Hz), temporal aggregation keeps the number of dimensions relatively low: 100 for the raw data and 120 for the graphical data.

3.3 Centering of the graphical data

An additional kind of preprocessing can be applied to the graphical data to take into account the invariance to the geometric location of points of contact. Since we are only interested in the *number* of contact points and not (at this point) in their location, we identify a "center of gravity" of the contact points and re-normalize the images so as to put the center of gravity in a certain location, namely the upper-left corner of the image.

To compute the center of gravity along dimensions x and y of the graphical image, the image histograms along each dimension, v_x and v_y are computed. Then a center of gravity for each dimension is computed as:

$$g_x = \left\lceil \frac{\sum_{i=1}^{d_x} v_x^i i}{\sum v_x^i} \right\rceil,$$

and similarly for the dimension y. Let $a = I_{(g_x, g_y)}$. Then the image matrix I is transformed as follows:

$$\begin{bmatrix} I_{11} & \vdots & I_{12} \\ \cdots & a & \cdots \\ I_{21} & \vdots & I_{22} \end{bmatrix} \longrightarrow \begin{bmatrix} a & \cdots & \cdots \\ \vdots & I_{22} & I_{21} \\ \vdots & I_{12} & I_{11} \end{bmatrix}.$$

4 Experiments and results

We evaluate the use of three popular modern classification algorithms: the classical k-nearest-neighbor (KNN) algorithm [4], the Support Vector Machines (SVM) with polynomial and RBF kernels [16, 11, 14] and the Kernel Fisher Discriminant [10]. Multi-class classification is implemented by training 5 one-against-the-rest classifiers and the class with the largest classifier output as the class of an example.

The experiment is carried out as follows. One third of the data is reserved for validation; on the remaining two thirds 5-fold cross-validation is used to determine the optimal model parameters (k in the KNN, C, and kernel parameters γ and polynomial degree for SVM and KFD). The process is repeated 10 times to obtain statistically significant results.

Table 1 shows the classification error of SVM with the RBF kernel and optimal parameters of one-against-the-rest classifiers for the graphical and the raw data. One can see that the classification is more accurate on the graphical data. In the remaining part we present the results only for the graphical data.

Classification error of multi-class classifiers is shown in Table 2. Here one can see a significant advantage of both kernel methods with the RBF kernel.

ESANN'2004 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 28-30 April 2004, d-side publi., ISBN 2-930307-04-8, pp. 549-554

	Graphical data				Raw data			
class	mean	std	C	gamma	mean	std	C	gamma
1	0.037	0.012	39.54	0.25	0.019	0.005	46.5	1.33
2	0.077	0.015	17.34	0.16	0.052	0.016	25.2	0.34
3	0.079	0.016	45.94	0.46	0.092	0.014	9.6	0.46
4	0.059	0.012	21.44	0.46	0.096	0.014	13.0	0.54
5	0.014	0.005	9.60	1.20	0.043	0.010	11.6	0.71

Table 1: Classification error of one-against-the-rest classifiers for SVM with RBF kernel.

classifier	mean	std
SVM with RBF kernel	0.121	0.016
SVM with RBF kernel and centered data	0.106	0.007
SVM with polynomial kernel	0.263	0.018
K nearest Neighbour	0.171	0.020
Fischer discriminant with RBF kernel	0.134	0.017
Fischer discriminant with polynomial kernel	0.284	0.024

Table 2: Classification error of multi-class classifiers.

Finally, the impact of the centering on the classification accuracy on the graphical representation is presented in Table 3. One can observe significant accuracy improvement in classes 1–3 and deterioration of accuracy in classes 4 and 5. This suggests that the notion of the “center of gravity”, is advantageous for simpler gestures, with some local contextual focus, but leads to problems in the representation when the gesture is more distributed across the active area.

5 Conclusions and future work

We have shown that classification algorithms such as Support Vector Machines and the Kernel Fischer Discriminant can be successfully applied to determine the number of finger strokes on a tactile device with the projective matrix sensoring. This offers a possibility to extend applicability of such touchpads, simpler and cheaper than the ones with the array sensoring, from one-touch to

class	mean	std	C	gamma
1	0.025829	0.0057451	61.71	0.79393
2	0.037441	0.0070473	5.7017	0.16681
3	0.059242	0.013907	8.4812	0.5469
4	0.081161	0.011009	6.7101	0.24115
5	0.037441	0.0041497	6.2964	0.46416

Table 3: Single class results for centered graphical data

ESANN'2004 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 28-30 April 2004, d-side publi., ISBN 2-930307-04-8, pp. 549-554

multi-touch applications. Our future work will address the issues of determining the exact location of contact points as well as their tracking during chordic manipulation on a multi-touch device.

References

[1] Synaptics. http://www.synaptics.com/products/touchpad.cfm.

[2] S. Bisset and B. Kasser. Multiple fingers contact sensing method for emulating mouse buttons and mouse operations on a touch sensor pad, 1998. US Patent 5825352.

[3] R. Boie, L. Ruedisueli, and E. Wagner. Computer mouse or keyboard input device utilizing capacitive sensors, 1995. US Patent 5463388.

[4] R.O. Duda and P.E. Hart. *Pattern classification and scene analysis*. John Wiley & Sons, 1973.

[5] Fingerworks, Inc. http://www.fingerworks.com/igesture_tech.html.

[6] G. Gerpheide and D. Lay. Capacitance-based proximity and interference rejection apparatus and methods, 1994. US Patent 5565658.

[7] B. Kasser, B. Joss, and S. Bisset. Touch sensing method and apparatus, 1998. US Patent 5790107.

[8] S. Lee. A fast multiple-touch-sensitive input device. Master's thesis, University of Toronto, 1984.

[9] P. McAvinney. Telltale gestures. *BYTE*, 15(7):237–240, 1990.

[10] S. Mika, G. Rätsch, J. Weston, B. Schölkopf, A.J. Smola, and K.-R. Müller. Constructing descriptive and discriminative nonlinear features: Rayleigh coefficients in kernel feature spaces. *IEEE Transactions on Patterns Analysis and Machine Intelligence*, 25(5):623–627, May 2003.

[11] K.-R. Müller, S. Mika, G. Rätsch, K. Tsuda, and B. Schölkopf. An introduction to kernel-based learning algorithms. *IEEE Transactions on Neural Networks*, 12(2):181–201, 2001.

[12] Touchglobe project. http://www.first.fraunhofer.de/german/profile/institute/first/areas/vista/first_f_vista_pro_01.html.

[13] D. Rubine and P. McAvinney. Programmable finger-tracking instrument controllers. *Computer Music Journal*, 14(1):26–41, 1990.

[14] B. Schölkopf and A.J. Smola. *Learning with Kernels*. MIT Press, Cambridge, MA, 2002.

[15] Tactex Controls, Inc. http://www.tactex.com/prodOEM.htm.

[16] V.N. Vapnik. *The nature of statistical learning theory*. Springer Verlag, New York, 1995.

[17] W. Westerman. *Hand tracking, finger identification and chordic manipulation on a multi-touch surface*. PhD thesis, University of Delaware, 1999.