ENHANCED SURFACE CLASSIFICATION FROM TACTILE DATA BY IMAGE FUSION

Anwesha Khasnobish[1], Monalisa Pal[2], Amit Konar[3], Dewaki Nandan Tibarewala[4], Atulya K. Nagar[5]
[1,4]School of Bioscience & Engg., [2,3]Dept. of Electronics & Telecommunication Engg., [5]Dept. of Math & Computer Science
[1,2,3,4]Jadavpur University, Raja S. C. Mullick Road, Kolkata – 700032, West Bengal, India
[5]Liverpool Hope University, United Kingdom
[1]anweshakhasno@gmail.com, [2]monalisap90@gmail.com, [3]konaramit@yahoo.co.in, [4]biomed.ju@gmail.com,
[5]nagara@hope.ac.uk

ABSTRACT
This work deals with recognizing surface by processing local information from their corresponding tactile images and fusing them to obtain the global pattern of surface irregularities. Tactile images are acquired while exploring surfaces with four kinds of texture patterns. Texture information is obtained from each of the images by edge detection of the region where higher amount of pressure is felt. These edge detected images are fused to obtain the pattern of surface irregularities. The fused images are classified using hierarchical multi-class Support Vector Machine which yields an accuracy of 83.334% in 0.083 seconds. It is observed that the classification accuracy is enhanced by image fusion than that obtained by concatenating features of each component image which formed the fused images in the former case. When noise is gradually added to the features, the classifier shows an accuracy of 75% even when SNR is 8dBW, indicating the robustness of the classifier. Also, the performance of the algorithm is tested by adding white Gaussian noise to the raw images. Finally, McNemar Test validates the results. Thus, the algorithm can be integrated into a tactile-sensing system in real-time scenario for identifying surfaces based on texture.

KEY WORDS
Tactile Images, Textures, Support Vector Machine.

1. Introduction

Design and development of artificial hands is a well-researched area in fields of rehabilitation, robotic surgery, tele-navigation and several other Human Computer Interaction (HCI) areas. However, the functionality of the artificial hands is incomplete without haptic perception. It is the sense of touch that allows us to distinguish objects around based on their properties like shape, size, deformability, etc. This study deals with the surface irregularities or textures which are important parameters for controlling friction and interfacing implants. While exploring a surface, human brain sequentially processes the spatial distribution of the irregularities in the form of images (i.e. pictorial processing) to perceive the surface, and thereby recognize the object [1-3]. With this insight, a pre-processing stage is presented for classification of surfaces based on textures where edges of the region sensed at different instants during exploration are fused to reconstruct the 2-D texture pattern. Here, lies the novelty of our research.

Several research has shown object recognition based on its properties like shape [4], size [5], etc. Also classification of surfaces from properties like deformability [6], texture [7], etc. can be found in literature. Based on Kalman filtering, 3-D object shapes have been reconstructed from tactile information and thereby recognized [8]. Although fine texture based classification has previously been done to distinguish different types of clothes, papers [9] and tailor-made surfaces [10], yet classification of house-hold items based on the pattern of surface irregularity has not been previously done as to the best of author's knowledge.

In this work, the information obtained from tactile images at different instants is combined to recognize four kinds of surface patterns. Some are tailor-made surfaces with 1mm embossed uniformly distributed patterns while the others are house-hold items. Images have been acquired over the duration of exploration. Each image is converted to grayscale image and from each of them the region of interest is obtained by histogram based thresholding. The edge of the region of interest is obtained using Sobel mask. All the images belonging to a single exploration are fused to obtain a single image representing that exploration which reveals the 2-D texture of the surface. The gradient information of this image forms the feature-space. Hierarchical multi-class Support Vector Machine is used to classify the features of the four surface classes. Results are compared with other standard classifiers. Also, a comparison of the result without the fusion step is shown. These results validate our hypothesis that forming an image by combining information from all the frames during exploration of the surface reveals the spatial distribution of its texture, thereby providing higher classification accuracy. The algorithm is also tested in presence of noise to claim its robustness.

Section 2 describes the entire course of the work. Results are discussed in Section 3. Finally, section 4 concludes the paper while mentioning future scope of research in this direction.

DOI: 10.2316/J.2014.216.818-0033

2. Methods

The process of experimental approach undertaken to acquire tactile images, extract features and classify surfaces based on the extracted features, are explained here.

2.1 Image Acquisition

Data has been acquired from 15 healthy subjects (8 female and 7 male) in the age group of 25±3 years who are given one of the surfaces shown in Fig. 1 and are asked to dynamically explore them with a fingertip, for 10 seconds after which the given surface is replaced by another surface. The surfaces are provided to the subjects in the orientation shown in Fig. 1. This is repeated 10 times for all the surfaces. The surface and the tactile sensor are fixed together using a micro-pore so that their relative orientation does not vary during exploration. By fixing the initial orientation of the surface and the relative frames of the surface and the sensor, the need for image registration is eliminated in this work. Thus, a data-set consisting of 15 (subjects) × 10 (number of times each surface is explored) × 12 (number of surfaces) = 1800 observations is obtained. A segment of the Tekscan grip sensor [11] is used as the tactile sensor and its sampling rate is fixed to 10 images per second.

Figure 1. Different surfaces having different textures: 1-dots (embossed pattern, lid of plastic box, hair brush), 2-dash (embossed pattern, switches of a remote control, strip of capsules), 3-horizontal lines (embossed pattern, comb, guitar strings) and 4-grid (embossed pattern, biscuit, badminton racket).

2.2 Pre-Processing and Feature Extraction

For each observation, there are 10 (sampling rate) × 10 (duration of exploration) = 100 raw component images. These images are processed to obtain relevant features for classification. The pre-processing and feature extraction consist of the following seven steps.

- At first, the RGB images are converted to grayscale images or intensity (I) images [12] using (1).

$$I = 0.2989 \times R + 0.5870 \times G + 0.1140 \times B \quad (1)$$

- Next, from each of these, an intensity histogram is constructed as shown in Fig. 2(i). Last bin of these histograms does not include the intensities of foreground as most of the background is white (intensity close to 255). From the intensity variation of all the images, it is noted that a threshold of the 240 can be chosen for converting the grayscale image to binary image. Thus, the output of the second stage consists of binary images revealing the region of interest.
- As the applied pressure is more where irregularity is more, we erode the region of interest slightly to obtain a more specific region of interest. For this morphological operation, a square structural element with dimension of 10 pixels is selected [12].
- These eroded images are sent from the third stage to fourth stage for edge detection using the Sobel mask [12]. Sobel mask is chosen for its superior performance than other edge detection masks as it places emphasis on the pixels closer to the centre of the mask. The Sobel mask is shown in Fig. 2(ii). The edge detection provides an outline of the surface irregularities as is perceived by the subject during exploration.
- At the fifth step, all the 100 component images corresponding to a single observation are fused by performing logical OR on the binary edge detected images.
- Following this, the gradient information is obtained which consists of the gradient magnitude (M) and the gradient direction (α) matrices where each matrix has the dimension of the image. These matrices are then converted to arrays in row-major format for extracting the statistical features.
- Finally, at the seventh step, six statistical features are extracted from each of the gradient magnitude and direction arrays. These six features are mean, standard deviation, skewness, kurtosis, 95 percentiles and mean absolute deviation. This yields a 12-dimensional feature vector (6 statistical features from the magnitude array and 6 statistical features from the direction array).

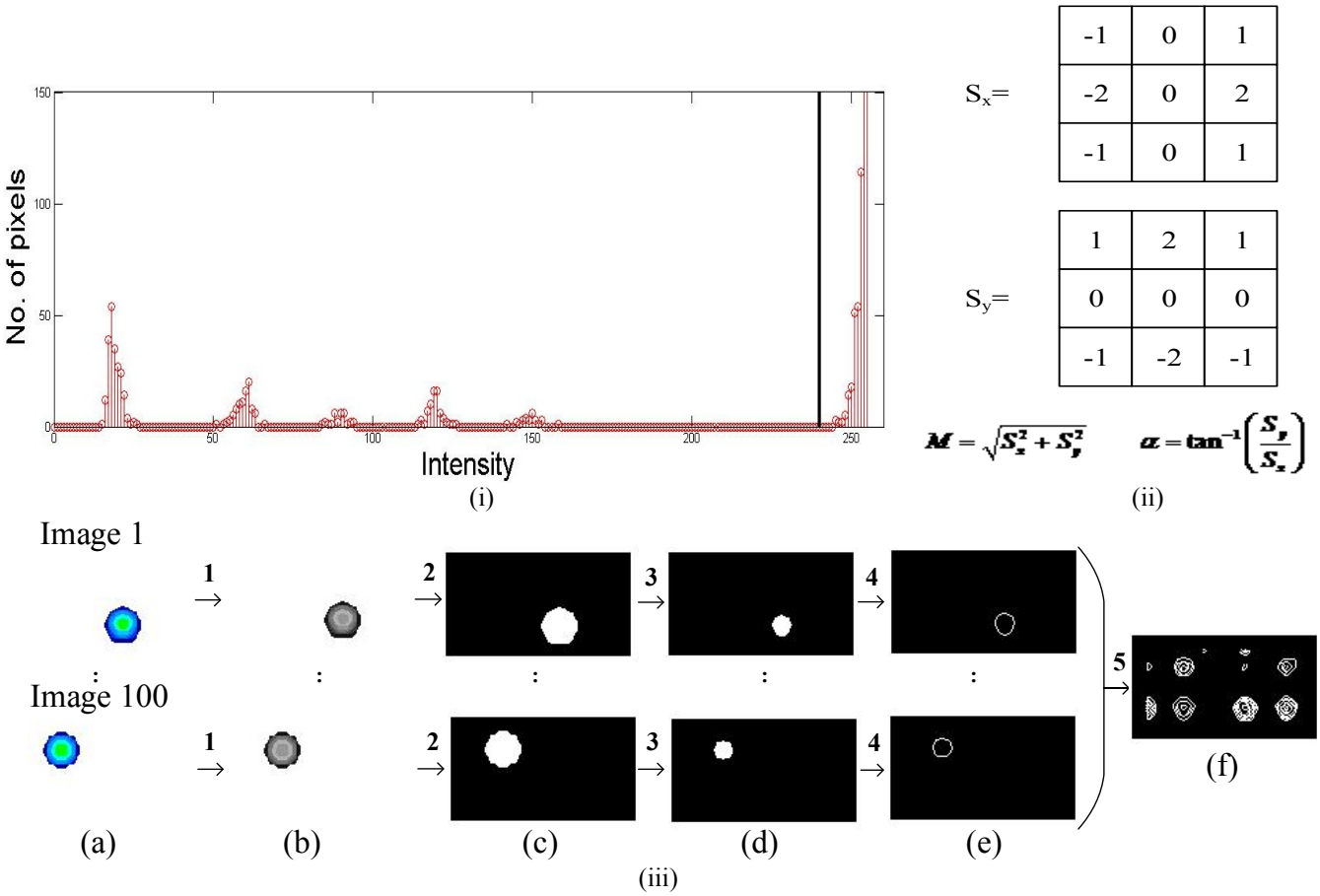

Figure 2. Pre-processing stages: (i) Histogram of an grayscale image with the threshold marked by bold black line; (ii) Sobel mask; (iii) Steps 1 to 5: (a) RGB image; (b) Grayscale Image; (c) Binary Image; (d) Eroded Image; (e) Edge of the region of interest; (f) Combined Image from information of all 100 frames of an observation.

Now, if the fifth step, where edge detected images are fused, is excluded and instead, the statistical features of the 100 component images are concatenated, we would have obtained a comparatively large 1200-dimensional feature space. The first five steps of image processing is briefly outlined in Fig. 2(iii).

2.3 Classifier: Hierarchical Multi-Class Support Vector Machine

The standard Support Vector Machine (SVM) is a binary classifier. In this work, the following strategy has been adopted to convert it into a multi-class classifier. This is called as the hierarchical multi-class SVM (HM-SVM) [13]. The dataset is recursively partitioned into non-overlapping subsets having maximum margin. For this, the mean feature vectors of each class are taken, and are clustered into two-sets using k-means algorithm [14]. The corresponding sets become the two-classes for the current SVM classifier. The clustering followed by classification of the resulting sets are continued until there is data from a single class. Each class forms a leaf node of the SVM hierarchy (binary tree).

As an example Fig. 3 shows the training procedure for a dataset having 5 classes. The mean feature vectors of five classes {1, 2, 3, 4, 5} are clustered into two sets i.e. {1, 2, 5} and {3, 4} at level-1. Thus, the dataset corresponding to classes 1, 2 and 5 forms one of the two classes for the SVM classifier at the top level and other class is formed from the data of class 3 and 4. Proceeding in a similar manner, the SVM classifiers at level-2 separates {1, 2, 5} into {1, 5} and {2}. On the other hand, {3, 4} is divided into {3} and {4}. Finally, {1, 5} gets divided into {1} and {5} at level-3.

During testing, the classification occurs in a top-to-bottom manner. At each SVM node, the predicted class decides whether to pass the unknown sample to the left or the right branch and finally, the classification stops at the leaf node deciding the class of the unknown sample.

3. Results and Discussions

For testing the proposed method, the dataset is 5-fold cross-validated. 60% of the dataset is used for training the classifiers, 20% for validating the chosen parameters and the remaining dataset for testing the trained classifier.

The result of multi-class classification for the four surfaces is mentioned in Table 1. For each of the surfaces, accuracy, Type-I error and Type-II error [15] is noted.

Accuracy is how many samples of a particular class are correctly identified by a trained classifier. With respect to a given class, Type-I error indicates how many samples of the other classes are identified as the given class and Type-II error indicates how many samples of the given class are classified as a different class. Along with this, the time taken to classify an unknown sample is also noted.

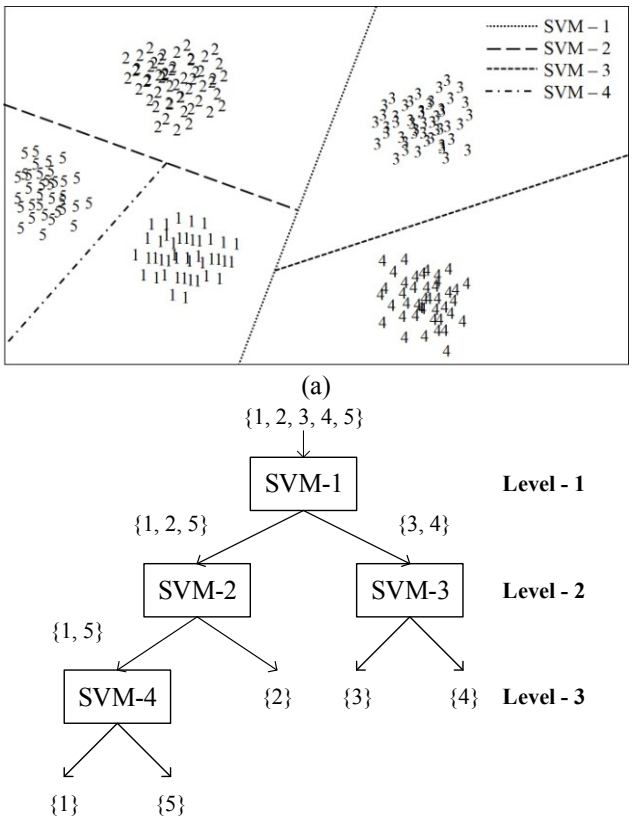

Figure 3. Training of HM-SVM: (a) Successive clustering of the feature space, (b) SVM tree.

Table 1
Multi-class Classification Results using HM-SVM

Class	Accuracy (%)	Type-I error (%)	Type-II error (%)
1	70.001	22.332	29.999
2	93.334	5.000	6.666
3	96.667	10.500	3.333
4	73.334	19.833	26.666
Mean	83.334	14.416	16.666
Time (s)	0.083		

The performance of the dataset constructed using proposed method (Dataset-1) is compared with other standard classifiers viz. Linear Discriminant Analysis (LDA) [16] and k-Nearest Neighbor (*k*NN with 3 neighbors, city-block distance and inverse distance weighting) [16] as shown in Table 2. Also, the classification results are compared by constructing the dataset concatenating the statistical features of the 100 component images i.e. without the fusion of images. This dataset is referred as Dataset-2 in Table 2.

Table 2
Comparison of Classification Results

Dataset	Performance Metrics (Mean)	HM-SVM	LDA	*k*NN
Dataset-1	Accuracy (%)	83.334	76.667	73.336
	Type-I error (%)	14.416	25.835	18.887
	Type-II error (%)	16.666	23.333	26.664
	Time (s)	0.083	0.0031	0.0016
Dataset-2	Accuracy (%)	45.835	45.005	36.667
	Type-I error (%)	23.020	46.744	31.552
	Type-II error (%)	54.165	54.995	63.333
	Time (s)	0.0909	0.0050	0.0017

From Table 2, it is observed that HM-SVM provides higher recognition rate than other standard multi-class classifiers. However, due to recursive partitioning of the data-set the classification time significantly increases. Moreover, recognition is very poor with Dataset-2. This may be attributed to the fact that the high-dimensional features has redundant and irrelevant information that increases the computational load (higher classification time) and reduces the accuracy such that the surfaces become unrecognizable (accuracy less than random guess i.e. 50%).

To assess the robustness of the designed HM-SVM classifier, additive white Gaussian noise (AWGN) is gradually added to the features at specified level of Signal-to-Noise Ratio (SNR). The mean accuracy for classifying the data-set at different SNR is mentioned in Table 3. It is noted from Table 3 that as the noise power increases (or equivalently the SNR decreases) the classification accuracy deteriorates. However, even when the SNR is 8dBW the classifier shows 75% classification accuracy (averaged over 10 trials). This shows that the HM-SVM is sturdy even in presence of high noise.

After this, white Gaussian noise is gradually introduced in the raw images (which is a representative of noise at the source or data acquisition in noisy environment). The Gaussian noise has constant mean (m) and variance (v) over the entire image. The mean and variance of the noise are normalized intensity values. The classification accuracy of these noisy images is tabulated in Table 4 at different mean and variance. To visualize the effect of noise, the first three pre-processing steps for four different cases (low mean, low variance; low mean, high variance; high mean, low variance; and high mean, high variance) are shown in Fig. 4. From both Table 4 and Fig. 4 it is observed that the algorithm performs well in presence of both low and medium noise. Its reason can be attributed to the intelligent use of erosion step in the pre-processing algorithm. However, it fails in presence of high noise.

Table 3
Classification Accuracy for the same dataset using HM-SVM in presence of AWGN

SNR (dBW)	Classification Accuracy (%)										Mean	Standard Deviation
	Trials											
	1	2	3	4	5	6	7	8	9	10		
30	75.00	83.33	75.00	91.67	83.33	91.67	83.33	83.33	83.33	83.33	83.332	5.557
28	91.67	91.67	83.33	75.00	91.67	75.00	83.33	83.33	75.00	83.33	83.333	6.805
26	75.00	91.67	83.33	83.33	91.67	75.00	75.00	91.67	83.33	83.33	83.333	6.805
24	100.00	83.33	75.00	83.33	83.33	91.67	75.00	83.33	75.00	83.33	83.332	7.857
22	91.67	83.33	75.00	83.33	100.00	75.00	75.00	91.67	83.33	75.00	83.333	8.785
20	91.67	83.33	75.00	91.67	75.00	75.00	100.00	75.00	83.33	83.33	83.333	8.785
18	75.00	91.67	83.33	83.33	83.33	91.67	75.00	91.67	75.00	83.33	83.333	6.805
16	75.00	75.00	91.67	83.33	83.33	75.00	91.67	91.67	83.33	75.00	82.500	7.298
14	91.67	83.33	83.33	66.67	83.33	75.00	83.33	66.67	83.33	91.67	80.833	8.827
12	66.67	83.33	91.67	91.67	83.33	83.33	66.67	83.33	75.00	75.00	80.000	8.957
10	66.67	83.33	83.33	75.00	83.33	66.67	75.00	75.00	91.67	83.33	78.333	8.049
8	75.00	91.67	66.67	83.33	83.33	75.00	66.67	66.67	66.67	75.00	75.001	8.783
6	66.67	66.67	75.00	75.00	66.67	50.00	66.67	58.33	58.33	75.00	65.834	8.288
4	66.67	66.67	50.00	75.00	58.33	58.33	50.00	58.33	58.33	41.67	58.333	9.622
2	50.00	41.67	50.00	66.67	41.67	41.67	66.67	41.67	58.33	58.33	52.501	9.662
1	50.00	50.00	58.33	41.67	58.33	58.33	50.00	50.00	50.00	41.67	51.667	7.657

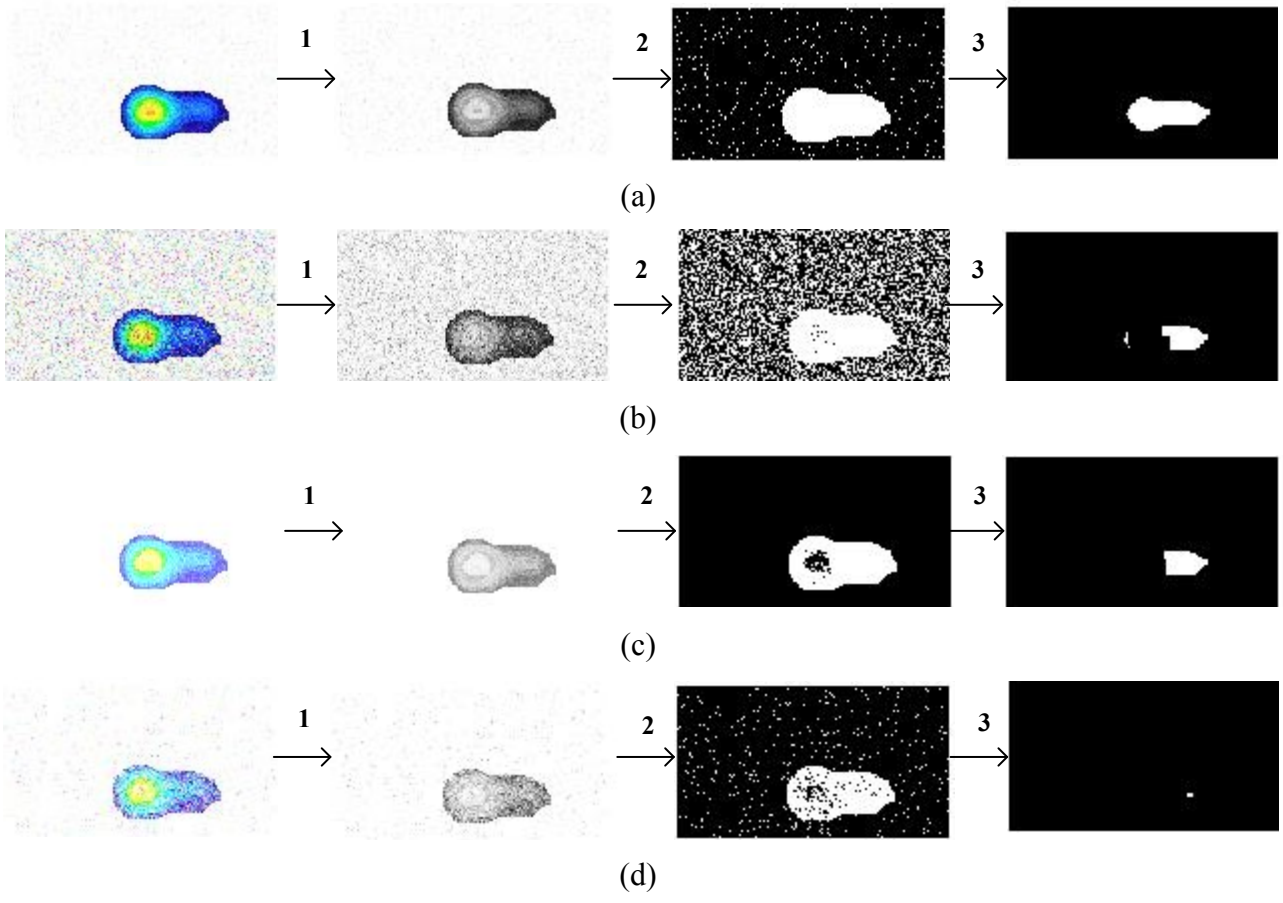

Figure 4. Initial three pre-processing steps of raw images affected by Gaussian noise with mean *m* and variance *v*. (a) *m*=0.1, *v*=0.01; (b) *m*=0.1, *v*=0.09; (c) *m*=0.5, *v*=0.01; (d) *m*=0.5, *v*=0.09.

Table 4
Mean Classification Accuracy (%) of images affected by white Gaussian noise

		Variance of Gaussian noise (v)				
		0.01	0.03	0.05	0.07	0.09
Mean of Gaussian noise (m)	0.1	83.33	81.67	78.33	72.67	65.00
	0.2	75.00	70.00	68.34	65.00	61.67
	0.3	73.33	66.00	65.00	55.83	46.67
	0.4	63.33	58.33	55.00	46.67	36.67
	0.5	56.67	51.67	46.67	40.00	33.33

Following this, the classification accuracy of the proposed algorithm is compared with LDA and *k*NN by means of McNemar Test which validates the results. McNemar Test [17] compares two algorithms and assesses which one is better among them. Here, we consider HM-SVM as the reference algorithm and compare it with either LDA or *k*NN at a time. The symbols given in Fig. 5 are used in McNemar Test where this table is called a contingency table used for comparing algorithms A and B.

n_{00}=number of samples misclassified by A and B	n_{01}=number of samples misclassified by A but not by B
n_{10}=number of samples misclassified by B but not by A	n_{11}=number of samples misclassified neither by A nor by B

Figure 5. Contingency Table McNemar Test.

According to the null hypothesis, all the classifiers are equivalent and thus, n_{01} and n_{10} are equal. In this study, the data-set is 5-fold cross-validated and the classifiers are tested on each fold of the data-set. Thus, 20% of the dataset = 20% of 1800 samples = 360 samples forms the test-set. The values of the contingency table of the 5-folds of classification are added to get full coverage of the data-set. The McNemar's statistic with one degree of freedom considering the correction factor is given by (2).

$$\chi^2 = \frac{\left(|n_{01} - n_{10}| - 1\right)^2}{n_{01} + n_{10}} \qquad (2)$$

The critical value of chi-square for 95% confidence interval is 3.84. If the chi-square value obtained from the contingency table is greater than the critical value $\chi^2_{1,0.05}$ then the null hypothesis is correct only with a probability less than 0.05 (in other words null hypothesis is rejected for 95% confidence interval). The parameters from contingency table used in the test, the chi-square value and the acceptance or rejection of null-hypothesis are indicated in Table 5.

Table 5 shows that the null hypothesis is rejected. Thus the classifiers are not equivalent. As for both the cases n_{10} is greater than n_{01} i.e. the number of samples misclassified by LDA or *k*NN but not by HM-SVM is greater than the opposite, thus, our claim that the HM-SVM algorithm is better in recognizing the surfaces than LDA and *k*NN is justified.

Table 5
Statistical Test: McNemar Test

Reference Algorithm (A) = HM-SVM				
Classifier Algorithm used for comparison (B)	Parameters for McNemar's Test		χ^2	Acceptance/ Rejection of null hypothesis
	n_{01}	n_{10}		
LDA	36	60	5.51	Rejected
*k*NN	18	54	17.01	Rejected

4. Conclusion

The aim of this work is to process tactile images, combine them to obtain enriched information of surface irregularities and thus using them to enhance classification of the surfaces. Tactile images are acquired by exploring four surfaces of different textures (realized by different embossed patterns and found in different household items). For each observation of 10 seconds, we have 100 RGB images. These are converted to grayscale images. From the intensity histogram of these images, a threshold is obtained using which the grayscale images are converted to binary image. The images are eroded to obtain more precise region of interest. Edge of the required regions are obtained. The edges from all the 100 images are fused to obtain a single image from which 12 statistical features based on the gradient, are extracted. The feature space is then classified by hierarchical multi-class Support Vector Machine (HM-SVM). From the results, we note superior performance of the HM-SVM with respect to LDA and *k*NN providing an accuracy of 83.334%. Also, the fusion of edge information from 100 images of an observation provides more accurate identification of surfaces rather than concatenating the feature space of each of the 100 component images. This is because the images are fused before feature extraction and hence the information content of all the images over a duration of exploration gets combined thereby increasing the information content of the single fused image from which the features are extracted. The feature vectors constructed in this manner contain the parameters of the fused images. Thus, fusion at the image level rather than at the feature level increases the information content without increasing the feature dimension and thereby enhancing the classification accuracy without compromising space complexity. The performance of the proposed pre-processing steps and the classification strategy are tested in presence of system-generated-noise which influences the feature space and sensor-acquired-noise which influences the raw images. HM-SVM shows

satisfactory performance providing an accuracy of 75% even when SNR is as low as 8dBW. The pre-processing stages shows good performance in low and moderate noise conditions. It is concluded that the reason for the good performance is the intelligent use of erosion step in the pre-processing stage of the raw images.

Although the surface irregularity sensed depends to some extent on the resolution of the sensor, but in future we would like to extend this work to include more intricate real-life surface textures. Our next step is the hardware realization of the proposed method using prosthetic device or robotic arm fitted with tactile sensors having vibro-tactile actuators for feedback to create a real-time HCI system.

Acknowledgement

This study has been supported by University Grants Commission (UGC), India; University with Potential for Excellence Program (UGC-UPE) (Phase II) in Cognitive Science, Jadavpur University and Council of Scientific and Industrial Research (CSIR), India.

References

[1] A. Amedi, N. Raz, H. Azulay, R. Malach, & E. Zohary, Cortical activity during tactile exploration of objects in blind and sighted humans, *Restorative neurology and neuroscience, 28*(2), 2010, 143-156.

[2] T. W. James, S. Kim, & J. S. Fisher, The neural basis of haptic object processing. *Canadian Journal of Experimental Psychology, 61*(3), 2007, 219-229.

[3] M. Grunwald, T. Weiss, W. Krause, L. Beyer, R. Rost, I. Gutberlet, & H. J. Gertz, Power of theta waves in the EEG of human subjects increases during recall of haptic information. *Neuroscience Letters, 260*(3), 1999, 189-192.

[4] H. Nakamoto, W. Fukui, F. Kobayashi, F. Kojima, N. Imamura, & H. Shirasawa, Shape classification based on tactile information by Universal Robot Hand. In *Industrial Electronics, IECON'09, 35th Annual Conference of IEEE*, 2009, 2360-2365.

[5] M. Craddock, & R. Lawson, The effects of size changes on haptic object recognition, *Attention, Perception, & Psychophysics, 71*(4), 2009, 910-923.

[6] M. Pal, A. Khasnobish, A. Konar, D. N. Tibarewala, & R. Janarthanan, Classification of Deformable and Non-Deformable Surfaces by Tactile Image Analysis. In *Control, Instrumentation, Energy & Communication (CIEC 14), IEEE International Conference on*, 2014, 701-705.

[7] N. Jamali, P. Byrnes-Preston, R. Salleh, & C. Sammut, Texture recognition by tactile sensing. In *Australasian Conference on Robotics and Automation (ACRA)*, 2009.

[8] M. Meier, M. Schopfer, R. Haschke, & H. Ritter, A probabilistic approach to tactile shape reconstruction, *Robotics, IEEE Transactions on, 27*(3), 2011, 630-635.

[9] F. De Boissieu, C. Godin, B. Guilhamat, D. David, C. Serviere, & D. Baudois, Tactile texture recognition with a 3-axial force MEMS integrated artificial finger. In *Robotics: Science and Systems,* 2009.

[10] S. H. Kim, J. Engel, C. Liu, & D. L. Jones, Texture classification using a polymer-based MEMS tactile sensor, *Journal of micromechanics and microengineering, 15*(5), 2005, 912-920.

[11] http://www.tekscan.com/grip-pressure-measurement

[12] R. C. Gonzalez, R. E. Woods, & S. L. Eddins, *Digital image processing using MATLAB* (Vol. 2, Knoxville: Gatesmark Publishing, 2009).

[13] S. Liu, H. Yi, L. T. Chia, & D. Rajan, Adaptive hierarchical multi-class SVM classifier for texture-based image classification. In *Multimedia and Expo, ICME 2005, IEEE International Conference on,* 2005, 4.

[14] A. K. Nazeer, & M. P. Sebastian, Improving the Accuracy and Efficiency of the k-means Clustering Algorithm. In *Proceedings of the World Congress on Engineering*, 2009, 1-3.

[15] M. D. Lieberman, & W. A. Cunningham, Type I and Type II error concerns in fMRI research: re-balancing the scale. *Social Cognitive and Affective Neuroscience, 4*(4), 2009, 423-428.

[16] S. Theodoridis, & K. Koutroumbas, *Pattern recognition* (Fourth Edition, Academic Press, MA, USA, 2008).

[17] P. A. Lachenbruch, & C. J. Lynch, Assessing screening tests: extensions of Mc Nemar's Test, *Statistics in medicine, 17*, 1998, 2207-2217.