A New Approach of a Touchless Virtual Hand for SixthSense Technology

S. M. Kamrul Hasan[1]
[1]Department of Electrical and Electronic Engineering
Khulna University of Engineering & technology (KUET)
Khulna-9203, Bangladesh

Mohiuddin Ahmad[1, 2]
[1]Department of Electrical and Electronic Engineering
[2]Department of Biomedical Engineering
Khulna University of Engineering & technology (KUET)
Khulna-9203, Bangladesh

ABSTRACT

With the advancement of technology particularly the Human Computer Interaction (HCI), life has become easier and more comfortable. Taking the advantage of HCI we present a concept of Touchless Virtual Hand (TVH) to aid users a new and cheap way of experiencing multi-touch capabilities, without the need of expensive hardware or software. Our proposed Touchless Software Development Kit (SDK) incorporates a convenient Bengali Word Recognition System with the application of thrilling potential of SixthSense technology that can superimpose the digital information on the physical world. We use RGB color model for this application that can effectively enable the users to interact the physical world with the digital world.

General Terms

C#, Feature Extraction, Gesture Recognition.

Keywords

Bicubic Interpolation, RGB Color Model, SixthSense Technology, Touchless Writer, Touchless Virtual Hand, Virtual Hand Gaming, Virtual Image Zooming.

1. INTRODUCTION

Now we live in a technologically sound world. It has blessed us with many advanced technologies that have made our life more comfortable. HCI has made this possible. Present days, keypad technology is replaced by touchscreen technology [1]. But due to expensive hardware requirements this this technology is becoming more complex, especially for developing country like Bangladesh. So, we need a cheap and user friendly system that can make our world like a magic world. Touchless technology based on gestures instead of clicks and typing may have been an element of a sci-fi movie in 2002 but it's no longer science fiction today. So a motivational saying about our proposed technology is: "DO, WHAT MY HAND SAYS".

Touchless system is based on some processing: a real time webcam captures images, a marker is added to track the color processed by a color model that distinguish it from the surrounding color, thus webcam track the color of that particular marker and same color exists during writing. But Touchless system is poor for smooth positioning than touch system. It is also affected by lighting condition. We develop such a system that is not bound to any specific application like only character recognition. We develop a Touchless Virtual Hand (TVH) that can write any word or sentence without the use of any external device in associated with zooming in &

out of any image especially any Map. Our system can point any location on the map by moving & zooming in & out on the map. So, it acts like an online location finder. Our TVH can do a lot of things also. It can play games. We develop a snake game and a defense game like football or Hockey.

The outline of our paper is as follows: Section II describes some previous work on HCI. In Section III overall setup for our work is described. Results have been discussed in Section IV. Finally the discussion & conclusion part is included in Section V & VI.

2. LITERATURE REVIEW

In the recent years, a lot of researchers have tried to make a Touchless device. Almost all these methods are based on character recognition. In Statistical method, though the character is recognized successfully, but it is affected by noise [2], [3]. For word recognition implicit segmentation approach is implemented. But it is not suitable for gradual changing words [4]. For recognition task Principal Component Analysis (PCA) is used in [5] that can effectively reduce the dimension and extract the efficient features. Other approaches like Hidden Markov Model (HMM), Artificial Neural Network (ANN) [6], Fuzzy logic [7] etc. are proposed for Touchless character recognition.

None of them are effective for word or sentence recognition. Our motto is to discover a new generation computer system that can emphasis not only character recognition but also SixthSense technology. We develop an almost complete Touchless Virtual Hand (TVH) that can do a lot of things except handwritten word recognition.

3. PROPSED TOUCHLESS VIRTUAL HAND (TVH)

We divide our work into three sections:

- Hardware Arrangements
- Software Requirements.
- System Architecture.

3.1 Hardware Arrangements

For Hardware setup, we use three tools:

- Laptop
- Webcam
- Any marker for color detection.

After starting the application, a webcam attached to Laptop captures an object in view and tracks the markers. For making the system user-friendly, we keep an option to choose any type of object for detection. The marker is used to distinguish the pointer from the surrounding objects. Multiple markers can be added for zooming in and out of any roadmap or world map or any type of image. The overall hardware setup of our work is shown in Fig. 1.
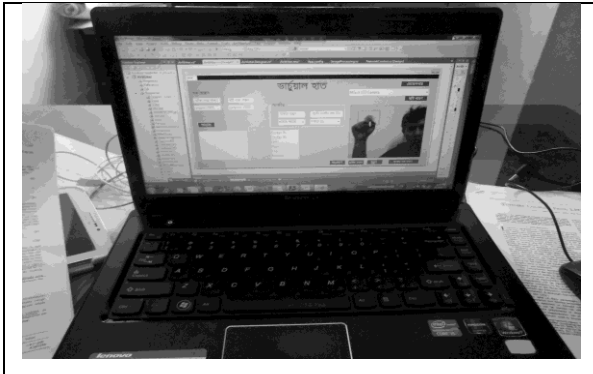
Fig. 1: Hardware setup of our system.

3.2 Software Requirements

In our work, we use different development tools. They are:

- Visual Studio 2010
- Microsoft SDK
- XML
- Avro Keyboard
- Windows Platform.
- Avro Keyboard installed.
- DotNet Framework 3.5.
- Aforge.net Framework.

We use Visual C# rather than C++ due to its access to the .NET library that contains a huge amount of pre-written code. For accessing into the API (Application Programming Interface), we use SDK (software Development Kit) that includes headers, tools, libraries etc. We use the AForge.net library function for video capturing. AForge.Video is used for video processing.

3.3 System Architecture

We propose a TVH that can be used for handwritten word, especially Bengali word even sentence recognition. A webcam tracks the marker color processed by a color model. For color detection we use a hybrid RGB-YCbCr-HSV color model that distinguishes the object color from the surrounding objects. Our proposed system uses the combination of Principal Component Analysis (PCA) & Linear Discriminant Analysis (LDA) for dimension reduction. We use Support Vector Machine (SVM) for training and classification. The reason for choosing SVM is due to its better classification capability. The reasons are written below:

- In SVM there is no overfitting problem.
- It automatically selects the model size by selecting the support vectors.
- It is a very easy training system.
- Artificial Neural Network (ANN) suffers from multiple local minima, but SVM has unique global minima.

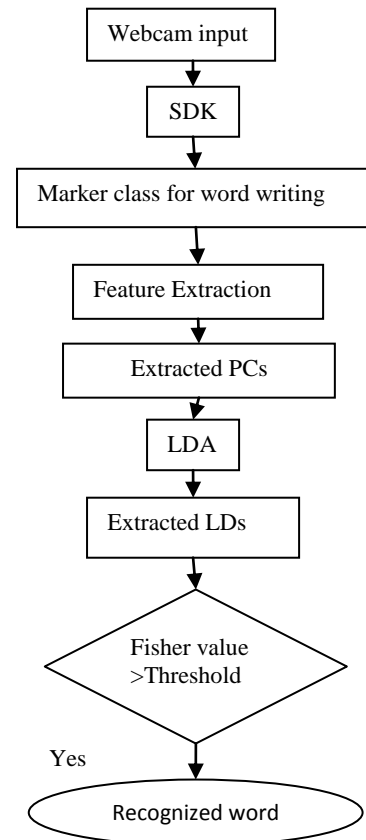We rearrange our overall TVH architecture in an up-bottom approach as Fig. 2.

Fig. 2: Overall system architecture.

3.3.1 Pre-processing

Webcam is the first stage of HCI. We use the AForge.Video library for accessing the webcam in real time mode. The System starts with choosing marker color through webcam. Many researchers have chosen the HSV or other color model for marker color detection, but it is time consuming. Therefore, we choose an RGB color model so that the computational time becomes fast. In this process, we need the color information not the intensity and so we choose RGB rather than HSV. The system starts with the manual selection of marker region & for this we use a circular red colored blob so that within this region color properties become almost the same. Then threshold of image is done by a simple decision rule where the region inside the blob is 1 i.e. white and outside the circle its value is 0 i.e. black. Thresholding is shown in Fig. 3.
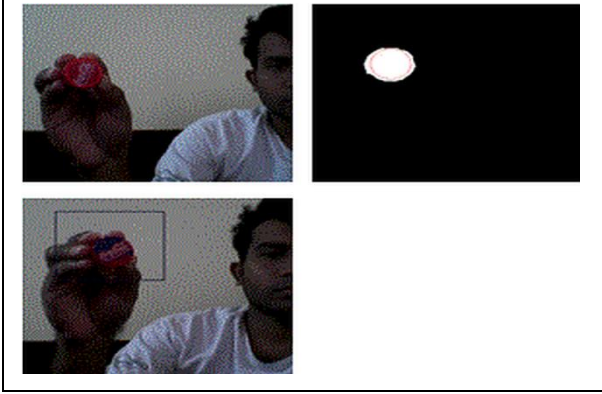
Fig. 3: Marker selection steps (From left to right and top to bottom) manual blob selection, thresholding, object detection.

For making the system user-friendly, we use two different colors: red & green. For handwriting we use a red circular object. To track the object a circular blob is selected manually to choose the area that should be tracked by the webcam depending on the object color. Once the area is selected, then a rectangular bounding box will show the tracked object area dynamically with the movement of the object using *EuclideanColorFiltering ().* The writing color can be chosen from the List box. Most attractive feature of our system is to select multiple markers. We use dual markers for zooming in & out of an image. We use Bangladesh Map for this purpose to locate any area, pointing by a circle.

3.3.2 Zooming Algorithm

Zooming operation corresponds to the gradient interpolation between points of several pixels on the input image. To collect any information from the image we develop a Virtual Image Zooming (VIZ) technique. Different interpolation algorithms are used for this operation. Amongst these, we use Bicubic Interpolation (BI) as it works best. Interpolation is mainly a process to get the continuous value from the sample that is discrete in form [8]. For bicubic interpolation we calculate the grey value of the output points. We sequentially convert RGB image into a matrix of double and convert it into a grey scale image as shown in Fig. 4. At the starting of BI algorithm, we convolute four 2-D cubes horizontally & then one more convolution vertically & thus calculates 16 closest pixels. The Bicubic Interpolation function that we use is shown in (1).

$$i(x, y) = \sum_{x=-1}^{2} \sum_{y=-1}^{2} g_{l+x,m+y} \, p(\frac{3}{2}|d|^3 - \frac{5}{2}|d|^2 + 1)_x$$

$$p(\frac{3}{2}|d|^3 - \frac{5}{2}|d|^2 + 1)_y \qquad \qquad \textbf{(1)}$$

Where p ( ) is the 1-D interpolation function along x or y direction, $g_{l+x,m+y}$ is a non-boundary interpolation coefficients, d is the distance between the grid and the interpolated point.

Our zooming algorithm is better than any other algorithm. The salient features of BI technique are as follows:

- Adequate for high frequency signal interpolation.

- Smoother appearance level.

- Small loss of image resolution.

```
public static Bitmap ToImage(double[]
Matrix, int MatrixRowNumber, int
MatrixColumnNumber,

int ImageHeight, int ImageWidth)
     {
        double HRate =
((double)ImageHeight / MatrixRowNumber);
        double WRate =
((double)ImageWidth / MatrixColumnNumber);
        Bitmap Result = new
Bitmap(ImageWidth, ImageHeight);

        for (int i = 0; i <
ImageHeight; i++)
        {
            for (int j = 0; j <
ImageWidth; j++)
            {
                int x =
(int)((double)j / WRate);
                int y =
(int)((double)i / HRate);

                double temp = Matrix[y
* MatrixColumnNumber + x];
                Result.SetPixel(j, i,
Color.FromArgb((int)((1 - temp) * 255),
(int)((1 - temp) * 255), (int)((1 - temp)
* 255)));    } }
      return Result;
    } } }
```

Fig. 4: Gray scale conversion code.

BI function for zooming is shown for Bengali character 'ব' is shown in Fig. 5.
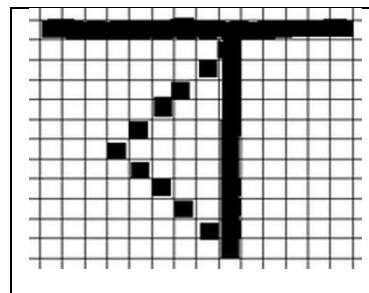

Fig. 5: Zooming of Bengali character 'ব'.

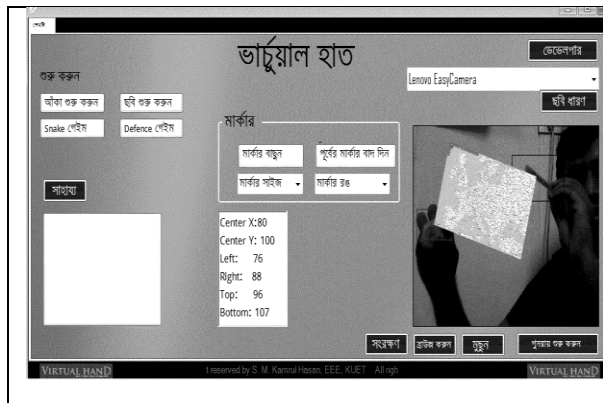The functioning framework view for zooming in and out of a map is shown in Fig. 6.

Fig. 6: Framework view with zooming.

3.3.3 *Dimension Reduction*

Each character has its distinctive feature that distinguishes it from another. To recognize word these features should be extracted. So, it is a vital part in handwritten word recognition performance. In feature extraction, least amount of elements maximizes the recognition rate. Principal Component Analysis (PCA) can be used to extract the most representative features to reduce the dimensionality, but keeping the features same to represent the data [9]. But the main problem of PCA is that often being selecting irrelevant Principal Components (PCs) form a poor space [10] and thus results in redundancy of the PCs.

To overcome the above problems we approach to add the most representative PCs to the feature extracted with Linear Discriminants (LDs). We divide our model into two stages. In the first stage we project the original image space to a lower dimensional space using PCA & in the second stage the between–class separability is maximized by projecting the PCA data to even lower dimensional space. This results in improved learning process by maximizing the distance between the classes, whereas minimizing the variance within-classes. To our research study over recognition, **this is the first attempt to successfully** recognize handwritten Bangla word using our proposed Touchless Virtual Hand (TVH). In this method, it scans the binary image until it finds the boundary. The searching follows according to the clockwise direction. For any foreground pixel say p, the set of all foreground pixels connected to it is called connected component containing p. The pixel p and its 8-neighbors are shown in Fig. 7. Once a white pixel is detected, it checks another new white pixel and so on. After feature extraction our input approximately reduces to 46%.
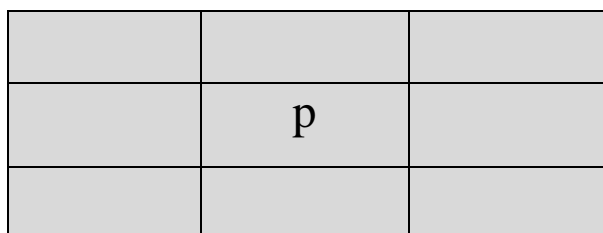
	p	

Fig. 7: Pixel p with its 8 adjacent.

3.3.4 *Classification*

In the field of machine learning, the objective of classification is to use the characteristics of an object which is known as feature value to identify which group it belongs to and to do this work classifier is used. It classifies given sign & also identifies the meaning of it. As a classifier we propose an almost newly emerging classifier named Support Vector Machine (SVM) [11]. The reason behind choosing it is its higher accuracy & better performance in handling small sample size. SVM mainly performs classification by constructing a hyper-plane that maximizes the distance between two data classes. For the simplicity we select linear SVM. The SVM is trained with the features obtained from the Linear Discriminants (LDs). We use SVM.NET library function based on .NET platform. For Touchless Virtual Hand (TVH), it is 1st time we use this method. The reasons behind the choosing of this method are:

- There is no overfitting problem.

- It automatically selects the model size by selecting the support vectors.

- It's a very easy training system.

- Faster training time.

- Higher accuracy than Artificial Neural Network (ANN).

3.3.5 *Game Development*

Using the same TVH Framework we develop a very easy game named ''Snake Game''. After choosing marker we have to select the button named "Snake গেইম" to start the game. The concept of snake game is to collect the food from the random grow by a small snake. A bounding box will prohibit the snake to go outside. We set the logic as below in Fig. 8.

Update	Draw
-Continuously check input	-Draw player
	-Draw food
-Collision check up	-Draw score

Fig. 8: Logic of Snake game.

3.3.5.1 *Snake Movement*

The movement of the snake is the main part of our game and we need to control the direction i.e. left, right, up & bottom using our chosen marker. A display view is shown in Fig. 9 where the black colored '1' like the shape line is a snake, the blue colored object is the food and red colored small circle is the pointer that scrolls the snake in different directions.
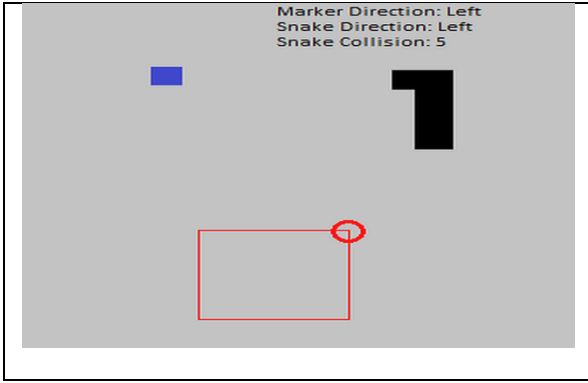
Fig. 9: Display view of our Snake game.

Starting from the back *for* loop is used to check whether '*i*' is equal to 0 or not. The logic is shown in Fig. 10.

```
Private void UpdateSnake(){
for (int i=snake.Count-1;i>=0;i--)
{
    if(i==0)
  {  //Head Logic
}
else{// Body Logic
}}}
```

Fig. 10: Snake movement logic.

3.3.5.2 Defense Game

In Similar way we develop another easiest game named "Defense Game" and the display looks like in Fig. 11. It can be started after pressing the button "Defense গেইম".
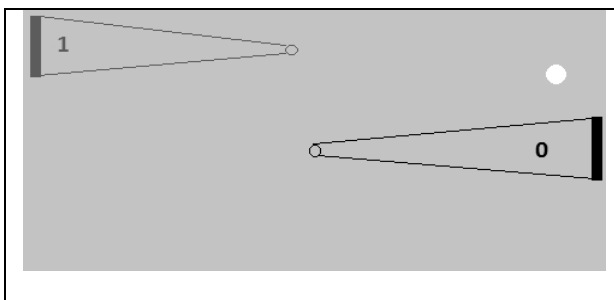

Fig. 11: Defense game.

4. RESULTS

In this experiment we use 59 Bengali characters for training purpose using SVM. Our TVH has less error due to better dimensionality reduction method. Table 1 shows the comparative comparison with our TVH with the conventional Touched module. The result shows that although our TVH requires more time than Touched Device, the error rate reduces significant levels due to better feature extraction technique.

Table 1. TVH vs. Conventional Touched Module

| Module	Training Sets	Feature Extracted Inputs for Recognition	Error Rate %	Average Time (min)
TVH	59	27	12.3	6.6
Touched Device	59	39	19.2	3.4

4.1 Handwritten Word in TVH

Our TVH can write any word as handwritten text. We focus on Bengali handwriting for Bangladeshi people. After the word is written in the Display as a rough writing, the duration of the writing is identified & then the data are collected. These data are then sent to the LDA (Linear Discriminant Analysis) block for both feature extraction and dimensionality reduction. After proper classification, the writing is shown in proper Bengali format. Fig. 12 shows the process in TVH.
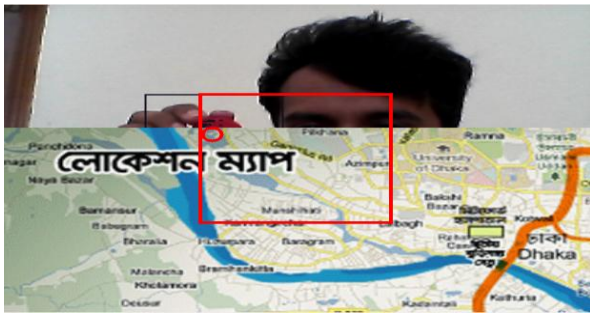

(a)   (b)
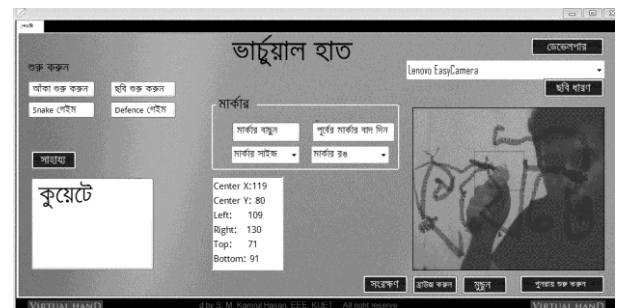
(c)

Fig. 12: Output Bengali handwriting "কুয়েটে" (a) free hand writing in the display, (b) after recognition, (c) overall framework view.

4.2 Map View in TVH

Our proposed model can perform some operation on any image like zoom in or zoom out, move or scroll. To locate any area under the map our system is a good one. Fig. 13 shows in details to locate 'নিউ ঢাকা সিটি ম্যাপ'.

(a)

(b)

Fig. 13: Location finding from the map.

5. DISCUSSIONS

During the selection of marker light intensity affect the webcam to detect the color information from the marker. As it is a freehand writing the classifier always cannot recognize the word. To overcome these problems good pixel camera and also a small object should be needed during marker selection.

6. CONCLUSION

In our research work, we introduced a new Touchless Virtual Hand (TVH) that will meet the demand of HCI. We develop a newly approach algorithm for Bengali word writing that works better than an external mouse cursor. Our VIZ technique can be used in defense application to locate any place from the map. Due to using RGB color model, there is no conversion problem and so it works well to choose color from complex background. In the future, we will make an android version of this technology so that it can be used in portable devices like TabletPCs, video games etc. We have a plan to use this technology to navigate a GPS via the internet so that it can be used in online. We will let our drawing facility in architectural design to make any presentation. We are trying to make our written text as a searching text in search engines like Google via the internet so that our technology can be utilized to full-fledged use. To make the written text for better use, we have a plan to ensure the printing facility of the text and also to compose emails to send via the internet. Thus, our designed module will be helpful for real life applications to make the world virtual.

8. REFERENCES

[1] P. Krejov and R. Bowden, "Multi-touchless: real-time fingertip detection and tracking using geodesic maxima," *IEEE Int'l Conf. Automatic Face and Gesture Recognition (FG),* pp.1-7, 2013.

[2] R. Plamondon and S. N. Srihari, "On-line and off-line handwritten recognition: a comprehensive survey," *IEEE Trans. Pattern analysis and machine intelligence, vol. 22, no. 1, pp. 63-84, 2000.*

[3] J. Cai and Z. Q. Liu, "Integration of structural and statistical information for unconstrained handwritten numeral recognition," *IEEE Trans. Pattern analysis and machine intelligence, vol. 21, no. 3, pp. 263-270, 1999.*

[4] A. Rehman, D. Mohamad, and G. Sulong, "Implicit vs. explicit based script segmentation and recognition: a performance comparison on benchmark database," *Int'l J. Open Problems Compt. Math., vol. 2, no. 3,* Sep. 2009.

[5] C. Liu, "Gabor-Based Kernel PCA with Fractional Power Polynomial Models for Face Recognition, " *IEEE Trans. Pattern analysis and machine intelligence, vol. 26, no. 5, pp. 572-581, May 2004.*

[6] L. L. Cheng and M. Nakagawa, "handwritten numeral recognition using neural networks: improving the accuracy by discriminative training," *Proc. IEEE Int'l Conf. Document Analysis and Recognition,* pp. 257-260, 1999.

[7] M. M. Hoque, Md. Mainul Islam, and M. M. Ali, "An efficient fuzzy method for Bangla handwritten numerals recognition," *IEEE Int'l Conf. Electrical and Computer Engineering,* pp. 197-200, 2006.

[8] R. R. Schultz and R. L. Stevenson, "Extraction of high resolution frames from video sequences," *IEEE Trans. Image Processing, vol. 5, no. 6, pp. 996-1011, 1996.*

[9] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: recognition using class specific linear projection," *IEEE Trans. Pattern analysis and machine intelligence, vol. 19, no. 7, pp. 572-581, Jul. 1997.*

[10] Z. Nan, W. Mio, and L. Xiuwen, "A hybrid PCA-LDA model for dimension reduction," *IEEE Int'l Joint Conf. Neural Networks, pp. 2184-2190, 2011.*

[11] H. Miyao, M. Maruyama, Y. Nakano, and T. Hananoi, "Off-line handwritten character recognition by SVM based on the virtual examples synthesized from on-line characters," *Proc. IEEE Int'l Conf. Document Analysis and Recognition, vol. 1, pp. 494-498, 2005.*