Journal of Information Hiding and Multimedia Signal Processing ©2014 ISSN 2073-4212
Ubiquitous International Volume **5**, Number **4**, October **2014**

A Multi-Modal Oculography-Based Mouse Controlling System: Via Facial Expressions & Eye Movement

Samaa M. Shohieb
Information Systems dept., Faculty of Computers and Information Sciences,
Mansoura Universit, Egypt,
sm.shohieb@yahoo.com

Hamdy K. Elminir
Department of Electrical Engineering,Faculty of Engineering,
Kafr El-Sheikh University,
hamdy_elminir@eng.kfs.edu.eg

A.M.Riad
Faculty Dean, Faculty of Computers and Information Sciences,
Mansoura University,
amriad2000@yahoo.com

Received November, 2013; revised April, 2014

Abstract. *Hand free mouse controlling systems is an effective and essential research area in human computer interaction (HCI), as they can be an ideal communication and interaction means for the physical disabled persons, who don't have means for communication except their moving eyes. In addition, these types of application provide usability and welfare for the normal users in the computer interaction. This paper presents a fully automatic multi-modal oculography-based mouse controlling system. This is done via integration between the recognized facial expressions with the eye movement and utilizing them in the mouse controlling. This paper presents a powerful multi-detector technique to localize the key facial feature points so that contours of the facial components are sampled such as the eyes, nostrils, chin, and the mouth. Also, the beholder's gaze is estimated. Based on the 66 extracted facial features points, 20 geometric formulas (GF) and 15 ratios (Rs) are calculated, and the classifier based on rule-based reasoning approach are then formed for both of the gaze direction and the facial expression (Normal, Smiling or Surprising). Tests were performed for 7 users with different ages and different facial characteristics then their overall satisfaction was measured. Our proposed multi-modal system obtained about 91.1% to 94.6% user satisfaction.*

Keywords: Facial expression recognition, Mouse controlling systems, Eye detection, Gaze tracking, Human-Computer interaction, Oculography based system

1. **Introduction.** Human face, including eyes and their movements, play very important role in expressing a persons needs, desires, pain, neurological disorders, cognitive processes, interpersonal rapports or even emotional states [1].

Despite active research and significant progress, gaze estimation and tracking still make a challenging due to the individuality of eyes, variability in scale, light conditions, occlusion and location. Building a video-based oculography tracking system is a very important research field in HCI studies. As it can play an ideal role for the physical disabled persons, who don't have means for the computer interaction except their moving eyes. Ditto, these types of application provide more usability and welfare for the normal users computer interaction.

To the best of our knowledge, all the available researches in the oculography tracking systems [2 - 9] depend mainly on the velocity of eye blinking to determine the correct action that will be taken by the computer. This, might has a bad affect that eyes get tired quickly. We found that combining between the facial expressions, controlled by the complex mesh of nerves and muscles beneath the face skin, and the gaze tracking will strengthen the computer interaction model and may reduce the user fatigue. Consequently, this paper contribution is integrating the recognized facial expressions with the eye gaze tracking to build a cursor-controlling interactive user interfaces.

This paper is organized as follows. Section 2 reviews the related work. Section 3 contains a general overview of the proposed model. Then facial feature points localization are in the 4th section. This is followed by a description of our proposed multi-modal oculography-based mouse controlling system via facial expressions & eye movement is presented in the 5th section. After that, the experimental results are in section 6. Finally, the conclusion and the future work are presented.

2. **Related Work.** As our proposed application utilizes both of oculography tracking and facial expression recognition. This section is divided to the three related parts. Facial expression recognition techniques, gaze estimation techniques and finally Gaze tracking applications.

2.1. **Facial Expression Recognition.** The existing methods of the facial expression recognition can be divided to three clusters according to how the features are extracted. The categories are the texture-based methodology, the geometric-based methodology, and the hybrid-based methodology.

2.1.1. *The Texture-Based Methodology.* Texture-based methodology process the entire image by applying a set of filters to extract facial features. Gupta et al. [10] used Discrete Cosine Transform, Gabor Filter, Wavelet Transform and Gaussian distribution for the feature extraction then classified these features using the adaboosting classifier. Researchers in [11] proposed adaptive filter selection (AFS) algorithm that is applied to choose the best subset of Gabor filters with different scales and orientations. The extracted features are then classified by adopting a multiple linear discriminant analysis (LDA) classifier.

Lirong et al. [12] used LBP features from training data, then they used these descriptors to train SVM classifier. Also Zhong et al. in [13] used the LBP for the feature extraction and then proposed a multi-task sparse learning (MTSL) framework locate those discriminative patches to train and classify these features.

2.1.2. *The Geometric-Based Methodology.* A face image is represented geometrically via fiducial points or the shape of facial regions [14]. LU and ZHANG [14] used optical flow to track facial feature points in sequential facial frames, normalized different positions of key feature points were computed and certain standardized geometric distances to form a facial-expression-arising-dataset (FEAD) matrix.

Also Pantic et al. [15] proposed a technique for detecting facial actions by analyzing facial components contours, including the mouth and the eyes. They used a multi-detector technique to sample the contours spatially and detect all facial features. Then they used a rule-based classifier to recognize the individual facial muscle action units (AUs).

2.1.3. *The Hybrid-Based Methodology.* The hybrid-based methodology raised from combining texture and geometric features Zhang and Ji [16] proposed a multi-feature technique that was based on the detecting edges in the forehead area, the facial points, and naso-labial folds. In their method, facial features were extracted by linking each AU with a set of movements, and then were classified using a Bayesian network model. Also in [17] the classification between smiling and neutral faces is done via a novel neural architecture that combines fixed, directional filters and adaptive filters in cascade. The fixed, directional filters extract primitive edge features, whereas the adaptive filters are trained to extract more complex features, which is classified by the linear classifier.

2.2. **Gaze Estimation Techniques.** The main task of gaze trackers is to define the gaze. Gaze should in this context be understood as either the point of regard (PoR) or the gaze direction. Thus, Gaze modeling focuses on the relations between the point of regard/gaze direction and the image features.

All gaze estimation techniques need to determine a set of parameters [18] through calibration. We clarify the calibration procedures into the following:

1. Geometric-calibration: determining relative locations and orientations of different units in the setup such as camera, light sources, and monitor.
2. Camera-calibration: determining intrinsic camera parameters.
3. Personal calibration: estimating cornea curvature, angular offset between visual and optical axes.
4. Gazing mapping calibration: determining parameters of the eye-gaze mapping functions.

According to [18], the gaze estimation techniques can be divided into two main categories; Feature-Based Face Estimation and Appearance-Based Face Estimation. These categories are detailed as follows.

2.2.1. *Feature-Based Face Estimation.* Gaze estimation methodologies using extracted local features such as eye corners, contours, and reflections from the eye image are called Feature-based methodologies. The main reasons for using feature-based methodologies are that the pupil and glints (under active light models) are relatively easy to find and that these features can be mainly related to gaze. This embrace aspects related to the geometry of the system as well as to eye physiology.

The glints fundamentally means that, when the curved cornea of the eye receives the falling light some of that light is reflected back in a narrow ray pointing directly toward the light source. Several reflections occur on the boundary between the lens and the cornea, producing the so-called Purkinje images [19]. The first Purkinje image or corneal reflection is often referred to as the glint. [2 - 8] that assume the mapping from image features to gaze coordinates (2D or 3D) have a particular parametric form such as a polynomial [9] or a nonparametric form such as in neural networks [20].

2.2.2. *Appearance-Based Methods.* The implied function for estimating the point of regard, relevant features, and personal variation can be extracted implicitly, without requirements of scene geometry and camera calibration. That is done to avoid the difficulties and difficiencies of the features-based methodologies. One such approach employs cropped images of eyes to train regression functions, as seen in multilayer network [21 - 23] or Gaussian processes [24] or manifold learning [25]. Images are high-dimensional representations of data, which are defined on a lower dimensional manifold.

2.2.3. *Gaze Tracking Applications.* Gaze tracking presents a powerful experimental tool for the study of real-time cognitive processing and information transfer. Gaze tracking applications include two main fields of application, namely, diagnostic and interactive [26].

Diagnostic eye trackers provide an quantitative and objective technique for recording the beholder's point of regard. This information is useful when examining people watching commercials, interacting with user interfaces, using instruments in plane cockpits, and in the understanding and analysis of human attention [27 - 30].

On the other hand, gaze-based interactive user interfaces react to the user's gaze either as a control input [31 - 34] or as the basis of gaze-contingent change in display [35]. Gaze-contingent means that the system is aware of the users gaze and may adapt its behavior based on the visual attention of the user, e.g., for monitoring human vigilance [26], [36 - 39]. Thus, the system tends to adapt its behavior according to the gaze input, which, in turn, reflects the person's desires.

3. **The Proposed Model.** A general overview of the components of our proposed system is shown in Figure 1. The system is divided into two parallel components; the oculography module and the facial expression recognition module. The oculography module obtain information from one camera (Input Image). The eye location in the image is detected and is subsequently tracked over frames. Based on the information obtained from the eye region, the gaze is estimated. Afterthat, the system calculates the GFs and Rs for the direction of gaze determination. On the other hand, the other facial key points are localized, tracked and used to calculate the GFs and Rs for the facial expression recognition process (that uses the rule-based classifier). Finally, the recognized gaze direction and the facial expression are used to control the movement and the right and left click for the mouse respectively.

4. **Facial Feature Points Localization.** Facial feature points are generally referred to as facial salient points such as the corners and centers of the eyes, corners of the nostrils, corners of the eyebrows, corners and outer mid points of the lips, tip of the nose, and the edge of the face. Permanent facial features are facial components such as eyebrows, eyes, and mouth. Their shape and location can alter immensely with expressions (e.g., pursed lips versus delighted smile). Transient facial features are any facial lines and bulges that did not become permanent with age but appear with expressions. We propose in this paper a simple method to detect automatically facial feature points.

4.1. **Face Detection.** Face detector is the first module in our facial expression recognition system, to localize the face in the image. This step allows an automatically labeling for facial feature points in a face image. We use a real-time face detector proposed in [40]. OpenCV library [41] which represents a m version of the original Viola-Jones face detector.

4.2. **Facial Feature Extraction.** To spatially sample the outline of the eyebrows, eyes, nostrils, and lips from an input frontal-view face image. We apply a simple analysis of image histograms in a combination with various filter transformations to locate six regions of interest (ROIs) in the face region segmented from an input frontal-view face image: two eyebrows, two eyes, nose, and mouth. The followed procedure is the updated version of the procedure in [42] and [43]. We supported head rotation of - 20 to 20 degrees of both in-plane and out-of-plane rotation. Detections speed ranges from 0.15 to 1.5 seconds, and the returned information of detected face based on (x, y) coordinates of face center, width and rotational angle. The facial features extraction components are detailed as following.

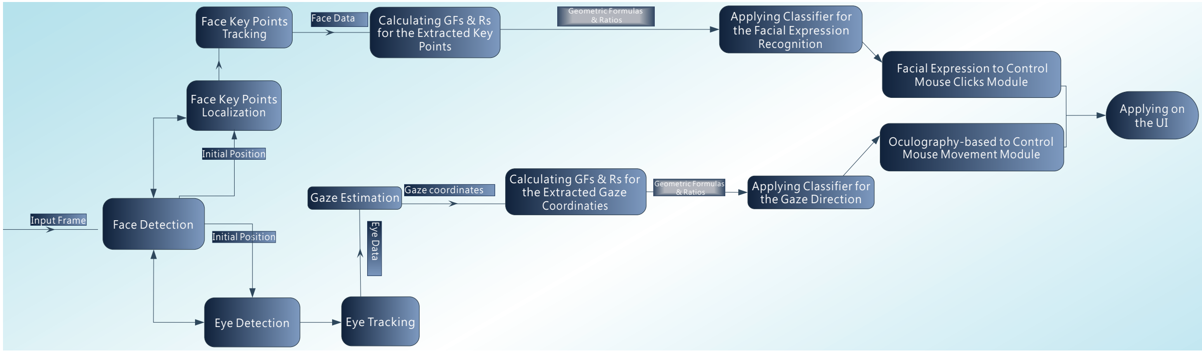

FIGURE 1. Components of The proposed Multi-Modal Oculography_Based Mouse Controlling System : Via Facial Expressions & Eye Movement.

1) Eye Detection, Initial Gaze Estimation, and gaze Tracking
Viola-Jones method integrated with Simple Elliptical Shape Models are used together to detect eye and iris respectively. The Viola- Jones classifier employs AdaBoost in a cascade basis to learn a high detection rate. The Viola-Jones classifier function is employed from OpenCV [41]. The Circle Hough Transformation (CHT) [44] is used which supports the strong and reliable detection of circular objects in an image. Let Cr be a circle with radius r:

$$Cr = \left\{ (i,j) \left| i^2 + j^2 = r^2 \right. \right\} \tag{1}$$

The Hough space H(x, y, r) is created using an extended Hough transform. It is applied on search mask which is generated based on a threshold segmentation of the red channel image corresponds to a three dimensional accumulator that indicates the probability that there is a circle with radius r at position (x, y) in the input image (I). Then the Hough Space (H) is described by:

$$H\left(x,y,r\right) = \sum \left(i,j\right) \varepsilon Cr I \left(x+i, y+j\right) \tag{2}$$

According to anthropometric datum [45], the regular eyeball radius ranges from 12 mm to 13 mm. Eye gaze in our model is estimated based on pupil location in addition to iris location to be more accurate and robust to a variety of users and conditions changes.

Since the iris contains little red hue, only the red channel is extracted from the eye region which contains a high contrast between skin and iris. Using some anthropometric statistics [45] we found the line contains eyes [46] which corresponds to many transitions: sclera to iris, skin to sclera, iris to pupil and the same thing for the high gradient side.

For the pupil detection, we used three types of knowledge [47]. We use pupil shape, color, and size as the first knowledge. First, we applied adaptive threshold method to eye image. Then, threshold T is obtained from average pixel value (mean) g of eye image. Threshold is set value is 27% below from the mean experimentally.

$$\mu = 1/N \sum_{i=0}^{N-1} Ii \tag{3}$$

$$T = 0.27\mu \tag{4}$$

According to eye movement, pupil shape and size are changed when user looks at right or left directions. This condition makes the pupil detection is hard to find.

To solve this problem the previously detected pupil location was utilized. The pupil is decided using following equation.

$$P\,(t-1)\;-C < P\,(t)\;< P\,(t-1)\;+C \tag{5}$$

The reasonable pupil location P(t) always in the surrounding previous location P(t-1) with the search area C.

Thereafter, pupil locations are tracked using the previously detected location using Kalman filter [48] for prediction of pupil location. At each time instance (frame), Pupil motion is figured with velocity and position. Let $(i_t,\ j_t)$ be the centroid of pupil at time t and $(u_t,\ v_t)$ be velocity at time t in i and j directions. The state vector at time t can therefore be represented as $x_t = (i_t\ j_t\ u_t\ v_t)t$. The system can therefore be modeled as

$$X_{t+1} = X_t + W_t \tag{6}$$

where Wt represents system disorders. Our assumption is that a fast feature extractor estimates $z_t = \left(\hat{i}_t\,,\hat{j}_t\right)$ the pupil position at time t. Thence, the measurement model in the form needed by the Kalman filter is

$$z_t = HX_t + U_t \tag{7}$$

2) Eyebrows Detection
Two different detectors localize the contours of the eyebrows in the eyebrow region of interests (ROIs). One applies the contour-following algorithm based on four-connected chain codes and the other fits a 2-D model of the eyebrow consisting of two second degree parabolas. The details of these algorithms are reported in [49].

Also for more accurate results for the eyebrows we excluded the eyes from the eye region detected in the face detection step [50].

3) Nostrils and Chin Detection
We used the method in [51]. In which the contours of the nostrils are detected in the nose ROI by applying a method that fits two 2-D small circular models onto the two small regions delimited as the nostril regions by a seed-fill algorithm. The seed-fill algorithm is also used to color eyes and mouth regions in the face region. An adapted version of the Vornoi-diagrams-based algorithm delimits the symmetry line between them [52]. The tip of the chin is localized as the first peak after the third deepest valley (the mouth) of the brightness distribution along the symmetry line [42].

4) Lips Border Detection
The natural color of the lips is red, so the ROI of the mouth is transformed into the HSV color space (Hue, Saturation, Value), in order to separate color from intensity, and then into red domain in order to segment mouth area from the rest of the strip image [53]. Image in red domain is obtained from the hue and saturation components of the input frame by applying the following thresholds [52]:

$$120 < H < 200 \tag{8}$$

$$S > 70 \tag{9}$$

By applying a method that fits 2-D small circular models onto the two small regions delimited as the Nasolabial_Fold regions by a seed-fill algorithm. The seed-fill algorithm is also used to color eyes and mouth regions in the face region. An adapted version of the Vornoi-diagrams-based algorithm [52] delimits the symmetry line between them. The tip of the chin is localized as the first peak after the third deepest valley (the mouth) of the brightness distribution along the symmetry line [42].

Figure 2 shows the facial feature extraction steps. In Figure 2 each key point of the feature extracted from the original frame is highlighted with a point in the figure. These points figure the x and y coordinate of each extracted feature.

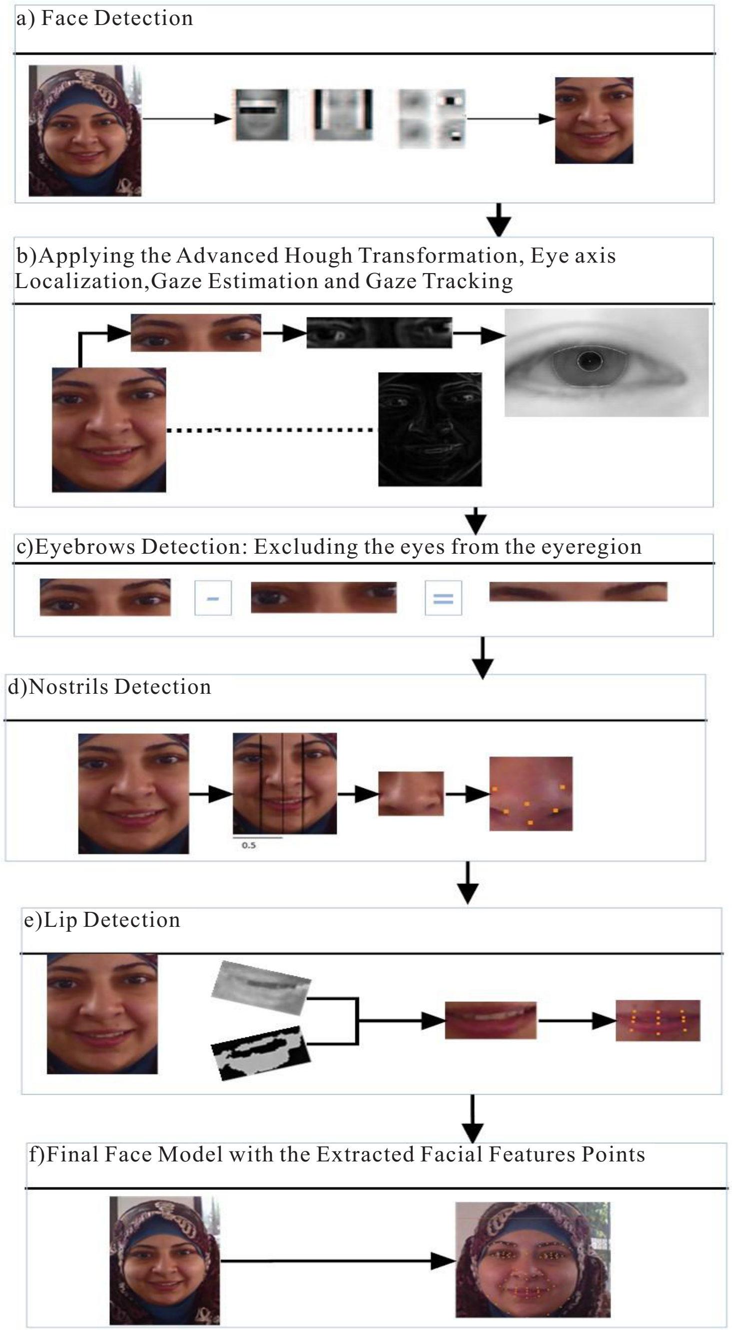

FIGURE 2. Facial Feature Points Localization Steps (Applied on Samaa Shohieb).

The extracted 66 features x and y coordinates are listed in Table 1 in the appendix. The features names followed the naming rules used in the Luxand FaceSDK v.4 documentation [54] but the values differ from those in Luxand library.

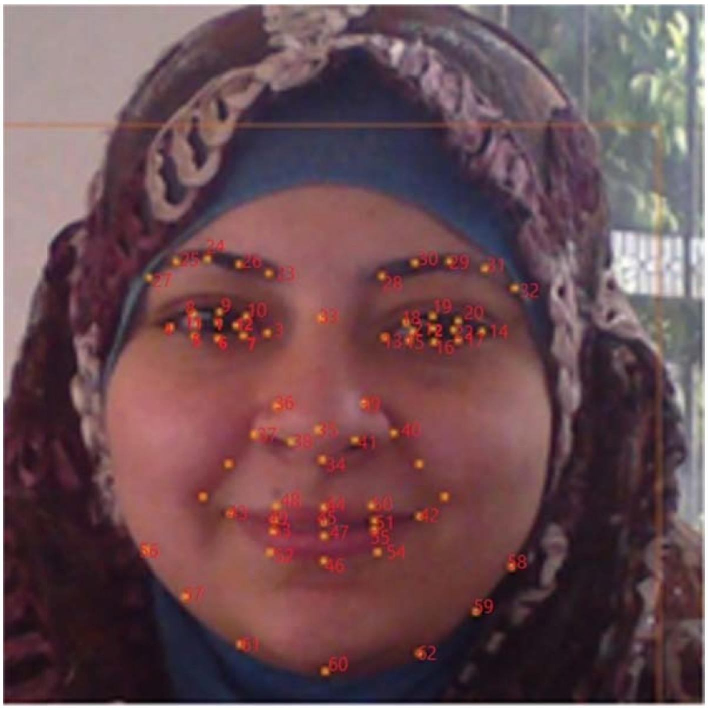

FIGURE 3. Facial Feature Points Localization Steps (Applied on Samaa Shohieb).

4.3. **Extracted Facial Points Tracking.** The facial feature points detected in the first image of the sequence will be tracked using pyramidal Lucas-Kanade algorithm [55]. This algorithm suppose that the brightness of every point of a static or moving object remains stable in time

5. **The proposed Multi-Modal Oculography-Based Mouse Controlling System : Via Facial Expressions & Eye Movement.** We extract standard displacements of main feature points to represent expression characteristics such the model in [56]. Twenty five more points are added to describe the movement of eyebrows, nose, and cheek more subtly. Figure 3 illustrates all 66 adopted facial feature points.

There are many classifiers that are used for classifying the facial features. One of these classifiers is the rule-based classifier. It has many advantages that its very simple, isnt a time consuming, can be used independently from the trained set [57]. However, it cant produce efficient results for large set of patterns.

To build our mouse input controlling system, we deduced to recognize three facial expressions; smiling and surprising in addition to the direction of the beholder's gaze. Thus, we built a simple rule-based classifier to detect the intended facial expressions. The followed rules reached from testing the values on ten persons with different ages, facial characteristics, genders , and one of them is a different nationality.

We calculated some geometric distances from each image that we addressed in the calculated GF. The formed GFs are more than those used in our system for future extensions in the facial expression recognition.

Width of mouth:

$$GF1 = F43.x - F42.x \tag{10}$$

Width of the lower nasolabial fold:

$$GF2 = F59.x - F57.x \tag{11}$$

Eyes axis:

$$GF3 = F14.x - F4.x \tag{12}$$

Left side of lip Openness:

$$GF4 = F53.y - F49.y \tag{13}$$

Right side of lip Openness:

$$GF5 = F55.y - F51.y \tag{14}$$

Height of the left eyebrow from the eye:

$$GF6 = F3.y - F26.y \tag{15}$$

Height of the right eyebrow from the eye:

$$GF7 = F13.y - F30.y \tag{16}$$

Width bet. the two eyebrows

$$GF8 = F28.x - F23.x \tag{17}$$

For the Right direction of sight

- X Distances bet. the left iris and the eye's outer corner

$$GF9 = F11.x - F4.x \tag{18}$$

- X Distances for the right iris from the eye's inner corner

$$GF10 = F21.x - F13.x \tag{19}$$

For the Left direction of sight

- X Distances bet. the left iris and the eye's inner corner

$$GF11 = F3.x - F12.x \tag{20}$$

- X Distances for the right iris from the eye's outer corner

$$GF12 = F14.x - F22.x \tag{21}$$

For the UP direction of sight

- Y Distances for the right iris from the eye's upper line

$$GF13 = F2.y - F19.y \tag{22}$$

- Y Distances for the left iris from the eye's upper line

$$GF14 = F1.y - F9.y \tag{23}$$

For the Down direction of sight

- Y Distances for the right iris from the eye's lower line

$$GF15 = F16.y - F2.y \tag{24}$$

- Y Distances for the left iris from the eye's lower line

$$GF16 = F6.y - F1.y \tag{25}$$

The following are more GFs for determining the lip openness.

Lip Openness In the Inner Center

$$GF17 = F47.y - F45.y \tag{26}$$

Lip Openness In the Outer Center

$$GF18 = F46.y - F44.y \tag{27}$$

Lip Openness In the Outer Left

$$GF19 = F52.y - F48.y \tag{28}$$

Lip Openness In the Outer Right

$$GF20 = F59.y - F57.y \tag{29}$$

Then, we calculated fifteen Rs as follows.

$$R1 = GF1/GF2 \tag{30}$$

$$R2 = GF2/GF3 \tag{31}$$

$$R3 = GF1/GF3 \tag{32}$$

$$R4 = GF4/GF5 \tag{33}$$

$$R5 = GF6/GF7 \tag{34}$$

$$R6 = GF8/GF1 \tag{35}$$

$$R7 = GF9/GF11 \tag{36}$$

$$R8 = GF10/GF12 \tag{37}$$

$$R9 = GF13/GF15 \tag{38}$$

$$R10 = GF14/GF16 \tag{39}$$

$$R11 = GF17/GF19 \tag{40}$$

$$R12 = GF17/GF18 \tag{41}$$

$$R13 = GF5/GF20 \tag{42}$$

$$R14 = GF5/GF4 \tag{43}$$

$$R15 = GF20/GF19 \tag{44}$$

Experiments have been done on about *5670 frames* from captured videos. We used about 3400 example (60% of the training sets out of 5670) for the training process and deducing the ratios Rs values between each GFs. And used about 1130 example (20% of the training sets out of 5670) as a cross-validation sets. Finally, we tested our classifier on about 1130 example (20% of the training sets out of 5670). We proved that the Rs between some specific two GFs (not only for the same ROIs but also it may be the ratios between the GFs of different ROIs) remain in the same range even if the observed face differs in its characteristics. The following are two examples to declare the relationship between the facial expression and the Rs.

For example, for the "smile" facial expression, the effective Rs are those affected by the Width of mouth (GF1), Width of the lower nasolabial fold (GF2), and Eyes axis (GF3). The former example declares that features from different ROIs affects the smile facial expression. Another example, for the right direction of sight, the effective Rs are those affected by the X distances between the left iris and the eye's outer corner (GF9), X distances between the right iris and the eye's inner corner (GF10), X distances between the left iris and the eye's inner corner (GF11), and X distances between the right iris and the eye's outer corner (GF12). The former example declares that features from the same ROIs affects the smile facial expression.

The following are the set of Rs that were conducted from our experiments.

```
IF (R7 < 0.8) AND (R8 < 0.8)
    TEHN the Line of sight is RIGHT;
    Move the cursor to the RIGHT side by a determined amount (25)
    Cursor.Position = new Point(Cursor.Position.X + 25, Cursor.Position.Y);
ELSE IF (R7 > 1.4) AND (R8 > 1.4)
    THEN the Line of sight is LEFT;
    Move the cursor to the LEFT side by a determined amount (25)
    Cursor.Position = new Point(Cursor.Position.X - 25, Cursor.Position.Y
    );
ELSE IF (R 9 < 0.5) AND (R 10 < 1)
    THEN the Line of sight is UP;
    Move the cursor to UP by a determined amount (25)
    Cursor.Position = new Point(Cursor.Position.X, Cursor.Position.Y-
    25);
ELSE IF (R 9 > 0.9) AND (R 10 > 0.9)
    THEN the Line of sight is DOWN;
    Move the cursor to DOWN by a determined amount (25)
    Cursor.Position = new Point(Cursor.Position.X, Cursor.Position.Y
    + 25);
ELSE
    THEN the Line of sight is STRAIGH";
    Do nothing for the cursor;
END IF
// Now determining the facial expression
IF (R 1 > 0.7) AND (R 2 > 0.7) AND (R 3 > 0.55)
    THEN Smiling Facial Expression
    Calling the ClickSomePoint() function to perform left click
ELSE IF (R 11 > 0.2) AND (R 12 > 0.2) AND (R 13 > 0.2) AND (R 4 <
2) AND (R 4 > 1) AND (R 1 > 0.6) AND (R 1 < 0.7)
    THEN Surprising Facial Expression
    Calling the RightClickSomePoint() function to perform Right click
END IF
```

According to that, all the used rules are simple linear equations, the running time of our program is varying from **0.15 to 1.6 seconds** with an addition of only 0.1 second from the localization step.

6. **Experimental Results.** Experiments show that the system works reliably under rotation angle between -20 to 20 degree. However it fails when the rotation angle is beyond this limit, or there is heavy occlusion. In the case of a tracking failure, the system will be reset and resume tracking shortly.

Figure 4 shows snapshots of our running system. Tests have been performed on 7 different test data that differ in ages (ranging from 5 to 27 years old), facial characteristics, facial expressions, length from the camera, direction of sight, scaling and rotation angles.

The overall satisfaction [58] was measured for each user. The details are shown in table 2 that clarify five satisfaction measurements for each user. And we applied on a pbskids.org game called "Drive Trolley" [59].This game take from 5 to 10 seconds for the trolley to complete one round using the hand mouse. We asked them all for the satisfaction with the idea. They, all, were satisfied with the idea. Then we made them complete the game without using the hand-mouse, and they completed the whole of it hand free. Afterthat, we calculated the elapsed time and it ranged from 6 to 20 seconds. We divided the elapsed time (if it was more than the maximum mouse-hand time for the trolley round; 10 seconds) over ten. If the time was lower than 10 seconds, it is estimated as100%. The accuracy was measured by finding how long time the user made a wrong direction over the overall time. Accuracy ranged from 50% to 100% depending on the user age. The accuracy and ease of use increase with increasing the user age as well. The overall user satisfaction was 91.1%. We can discarded the 5 years old child results as his results are considered as noisy results because of the difficulty of controlling his eye directions and his concentration. As a result the overall satisfaction for the other six users will increase to 94.6% as shown in table 3.

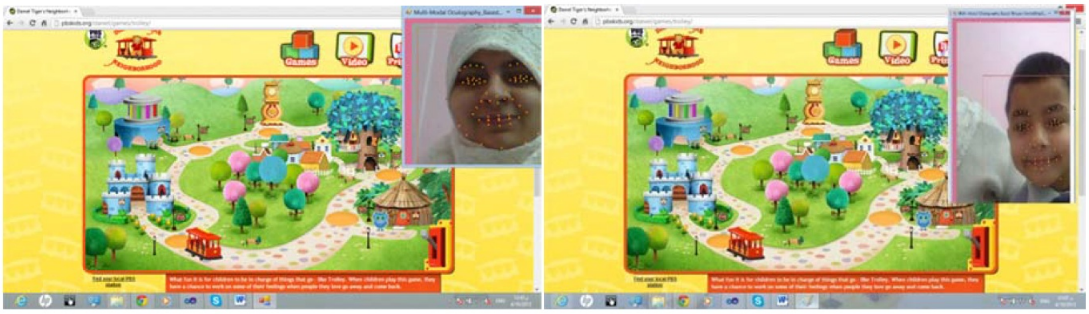

Figure 4. Snapshots for the running system including two users (27 and 5 years old respectively) during playing "Drive Trolley" Game.

7. **Conclusions.** In this paper we have proposed a multi-modal system that integrates both of estimating eye movement direction and recognizing the facial expressions for controlling the mouse movements and mouse left and right clicks. The system is divided into two parallel modules; the oculography module and the facial expression recognition module. The oculography module obtain information from one camera (Input Image). The eye location in the image is detected and is subsequently tracked over frames. Based on the information obtained from the eye region, the gaze is estimated. Afterthat, the system calculates 8 (GFs) and 4 (Rs) for the direction of gaze determination. On the other hand, the other facial key points are localized, tracked and used to calculate 12 GFs and 11 Rs for the facial expression recognition process (that uses the rule-based classifier). Finally, the recognized gaze direction and the facial expression are used to control the movement and the right and left click for the mouse using the recognized surprising and smiling facial expressions respectively.

Table 1. The User Satisfaction for the System (Including the 5 years old child)

Aged User	**Idea of the Application**	**Completed the Game without Using Mouse**	**Time in seconds**	**Accuracy**	**Ease of Use**	**Total**
User 1 (5years)	*100%*	*100%*	*20 (50%)*	*50%*	*50%*	*70%*
User 2 (8 years)	100%	100%	15 (75%)	70%	90%	87%
User 3 (10 years)	100%	100%	6(100%)	80%	92%	94.4%
User 4 (10 years)	100%	100%	8(100%)	80%	92%	94.4%
User 5 (13 years)	100%	100%	7(100%)	80%	94%	94.8%
User 6 (17 years)	100%	100%	8(100%)	90%	99%	97.8%
User 7 (27 years)	100%	100%	9(100%)	100%	96%	99.2%
Total	100%	100%	89.3%	78.57%	87.57%	91.1%

Table 2. The User Satisfaction for the System (Excluding the 5 years old child)

Aged User	**Idea of the Application**	**Completed the Game without Using Mouse**	**Time in seconds**	**Accuracy**	**Ease of Use**	**Total**
User 2 (8 years)	100%	100%	15 (75%)	70%	90%	87%
User 3 (10 years)	100%	100%	6(100%)	80%	92%	94.4%
User 4 (10 years)	100%	100%	8(100%)	80%	92%	94.4%
User 5 (13 years)	100%	100%	7(100%)	80%	94%	94.8%
User 6 (17 years)	100%	100%	8(100%)	90%	99%	97.8%
User 7 (27 years)	100%	100%	9(100%)	100%	96%	99.2%
Total	100%	100%	89.3%	78.57%	87.57%	94.6%

Our proposed system is a person independent hand free mouse controlling system. We made our tests on 7 persons and measured their overall satisfaction. We found that it obtained 94.6%.

In future, we intend to improve our system, increasing its accuracy. Also, implementing a reliable and fast full computer controlling system with more facial expressions to handle a virtual keyboard is a challenge we plan to face very soon.

Appendix. This appendix contains the detailed names and values for the extracted Facial features key points.

TABLE 3. Extracted Facial Features Key Points

Feature Value	Facial Feature Name	Feature Value	Facial Feature Name
F1	Left_Eye_Iris_Center	F34	Nose_Bottom
F2	Right_Eye_Iris_Center	F35	Nose_Bridge
F3	Left_Eye_Inner_Corner	F36	Nose_Left_Wing
F4	Left_Eye_Outer_Corner	F37	Nose_Left_Wing_Outer
F5	Left_Eye_Lower_Line1	F38	Nose_Left_Wing_Lower
F6	Left_Eye_Lower_Line2	F39	Nose_Right_Wing
F7	Left_Eye_Lower_Line3	F40	Nose_Right_Wing_Outer
F8	Left_Eye_Upper_Line1	F41	Nose_Right_Wing_Lower
F9	Left_Eye_Upper_Line2	F42	Mouth_Right_Corner
F10	Left_Eye_Upper_Line3	F43	Mouth_Left_Corner
F11	Left_Eye_Left_Iris_Corner	F44	Mouth_Top
F12	Left_Eye_Right_Iris_Corner	F45	Mouth_Top_Inner
F13	Right_Eye_Inner_Corner	F46	Mouth_Bottom
F14	Right_Eye_Outer_Corner	F47	Mouth_Bottom_Inner
F15	Right_Eye_Lower_Line1	F48	Mouth_Left_Top
F16	Right_Eye_Lower_Line2	F49	Mouth_Left_Top_Inner
F17	Right_Eye_Lower_Line3	F50	Mouth_Right_Top
F18	Right_Eye_Upper_Line1	F51	Mouth_Right_Top_Inner
F19	Right_Eye_Upper_Line2	F52	Mouth_Left_Bottom
F20	Right_Eye_Upper_Line3	F53	Mouth_Left_Bottom_Inner
F21	Right_Eye_Left_Iris_Corner	F54	Mouth_Right_Bottom
F22	Right_Eye_Right_Iris_Corner	F55	Mouth_Right_Bottom_Inner
F23	Left_Eyebrow_Inner_Corner	F56	Nasolabial_Fold_Left_Upper
F24	Left_Eyebrow_Middle	F57	Nasolabial_Fold_Left_Lower
F25	Left_Eyebrow_Middle_Left	F58	Nasolabial_Fold_Right_Upper
F26	Left_Eyebrow_Middle_Right	F59	Nasolabial_Fold_Right_Lower
F27	Left_Eyebrow_Outer_Corner	F60	Chin_Bottom
F28	Right_Eyebrow_Inner_Corner	F61	Chin_Left
F29	Right_Eyebrow_Middle	F62	Chin_Right
F30	Right_Eyebrow_Middle_Left	F63	Face_Contour1
F31	Right_Eyebrow_Middle_Right	F64	Face_Contour2
F32	Right_Eyebrow_Outer_Corner	F65	Face_Contour12
F33	Nose_Tip	F66	Face_Contour13

REFERENCES

[1] G. Underwood, *Cognitive Processes in Eye Guidance*, Oxford Univ. Press, pp. 325-354, 2005.

[2] X.L.C. Brolly, J.B. Mulligan, Implicit Calibration of a Remote Gaze Tracker, *Proc. 2004 Conf. Computer Vision and Pattern Recognition Workshop*, pp. 134, 2004.

[3] Y. Ebisawa, S. Satoh, Effectiveness of Pupil Area Detection Technique Using Two Light Sources and Image Difference Method, *Proc. 15th Ann. Intl Conf. IEEE Eng. in Medicine and Biology Soc.*, pp. 1268-1269, 1993.

[4] D. Witzner Hansen, J.P. Hansen, M. Nielsen, A.S. Johansen, M.B. Stegmann, Eye Typing Using Markov and Active Appearance Models, *Proc. IEEE Workshop Applications on Computer Vision*, pp. 132-136, 2003.

[5] D.W. Hansen and A.E.C. Pece, Eye Tracking in the Wild, Computer Vision and Image Understanding, *Special Issue on Eye Detection and Tracking*, vol. 98, no. 1, pp. 155-181, 2005.

[6] Q. Ji and X. Yang, Real-Time Eye, Gaze, and Face Pose Tracking for onitoring Driver Vigilance, *Real-Time Imaging*, vol. 8, no. 5, pp. 357-377, 2002.

[7] C.H. Morimoto, M.R.M. Mimica, Eye Gaze Tracking Techniques for Interactive Applications, *Computer Vision and Image Understanding*, vol. 98, no.1 , pp. 4-24, 2005.

[8] O.M.C. Williams, A. Blake, and R. Cipolla, Sparse and Semi- Supervised Visual Mapping with the s3p *Proc. of Computer Vision and Pattern Recognition*, pp. 230-237, 2006.

[9] D.M. Stampe, Heuristic Filtering and Reliable Calibration Methods for Video-Based Pupil-Tracking Systems, *Behavior Research Methods, Instruments & Computers*, vol. 25, no.2, pp. 137-142, 1993.

[10] Sandeep K. Gupta, ShubhLakshmi Agrwal, Yogesh K. Meena, Neeta Nain, Hybrid Method of Feature Extraction for Facial Expression Recognition, *Proc. of the IEEE 7th International Conference on Signal Image Technology & Internet-Based Systems*, 2011.

[11] S.M. Lajevardi, Z.M. Hussain, A Novel Gabor Filter Selection Based on Spectral Difference and Minimum Error Rate for Facial Expression Recognition, *proc. of the IEEE International Conference on Digital Image Computing: Techniques and Applications*, pp. 137-140, 2010.

[12] L.R. Wang, X,G. Yan, J.l. Wang, J. Xu, J. Zhao, Facial Expression Recognition Based on Local Texture Features, *proc. of the 14th IEEE International Conference on Computational Science and Engineering* , pp. 543-546, 2011.

[13] Z. Lin, Q.S. Liu, P.Yang, B. Liu, J.Z. Huang, D.N. Metaxas Learning Active Facial Patches for Expression Analysis, *proc. of the IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 2562-2569, 2012.

[14] K. Lu, Z. Xin, Facial Expression Recognition from Image Sequences Based on Feature Points and Canonical Correlations, *proc. of the IEEE International Conference on Artificial Intelligence and Computational Intelligence* , vol. 1, pp. 219-223, 2010.

[15] M. Pantic, L. J. M. Rothkrantz, Facial action recognitionfor facial expression a nalysis from static face images, *IEEE Trans.Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 34, no.3, pp. 1449-1461, 2004.

[16] Y. Zhang, Q. Ji, Active and dynamic information fusion for facial expression understanding from image sequences, *IEEE Trans. Pattern Analysis and Machine Intelligence* , pp. 699-714, 2005.

[17] P. Li, A. Bouzerdom, F. H. C. Tivive, Automatic Recognition of Smiling and Neutral Facial Expressions, *proc. of the IEEE International Conference on Digital Image Computing: Techniques and Applications* , pp. 581-586, 2010.

[18] D.W.Hansen, Ji Qiang, In the Eye of the Beholder:A Survey of Models for Eyes and Gaze *IEEE Trans. Pattern Analysis and Machine Intelligence* , pp. 478-500, 2010.

[19] A. Duchowski, *Eye Tracking Methodology: Theory and Practice*, Springer, PP. 30-45, 2003.

[20] Q. Ji, X. Yang, Real-Time Eye, Gaze, and Face Pose Tracking for Monitoring Driver Vigilance, *Real-Time Imaging*, vol. 8 , no. 5, pp. 357-377, 2002.

[21] S. Baluja, D. Pomerleau, Non-Intrusive Gaze Tracking Using Artificial Neural Networks *Technical Report CMU-CS-94-102*, Carnegie Mellon University Pittsburgh, PA, USA, 1994.

[22] R. Stiefelhagen, J. Yang, A. Waibel, Tracking Eyes and Monitoring Eye Gaze, *Proc. of the International Conference on Perceptual User Interfaces Workshop*, pp. 98-100, 1997.

[23] L.Q. Xu, D. Machin, P. Sheppard, A Novel Approach to Real-Time Non-Intrusive Gaze Finding, *Proc. of the International Conference on British Machine Vision* , 1998.

[24] Mapping with the s3p, *Proc. of the IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 230-237, 2006

[25] K.H. Tan, D.J. Kriegman, N. Ahuja, Appearance-Based Eye Gaze Estimation, *Proc. of the IEEE 6th International Conference on Applications of Computer Vision Workshop*, pp. 191-195, 2002.

[26] A. Duchowski, *Eye Tracking Methodology: Theory and Practice*, Springer, PP. 55-165, 2003.

[27] G. Anders, Pilots Attention Allocation during Approach and LandingXEye- and Head-Tracking Research, *Proc. of the 11th International Conference on Symp. Aviation Psychology*, 2001.

[28] J.H. Goldberg, A.M. Wichansky, Eye Tracking in Usability Evaluation: A Practitioners Guide. *Elsevier Science*, pp.493-516, 2003.

[29] K. Rayner, C. Rotello, C. Steward, A. Keir, A. Duffy, *J. Experimental Psychology: Applied*, vol. 7, no. 3, pp. 219-226, 2001.

[30] M. Wedel and R. Peiters, Eye Fixations on Advertisements and Memory for Brands: A Model and Findings, *Marketing Science*, vol. 19, no. 4, pp. 297-312, 2000.

[31] R. Bolt, Gaze-Orchestrated Dynamic Windows,*Proc. of the 8th annual conference on Computer graphics and interactive techniques* pp. 109-119, 1981.

[32] J.P. Hansen, K. Torning, A.S. Johansen, K. Itoh, A. Hirotaka, *Proc. of the symposium on Eye tracking research & applications*, pp. 131-138, 2004.

[33] C. Lankford, Effective Eye-Gaze Input into Windows, *Proc. of the symposium on Eye tracking research & applications*, pp. 23-27, 2000.

[34] D.J. Ward, D.J.C. MacKay, Fast Hands-Free Writing by Gaze Direction, *Nature: International weekly journal of science*, vol.418, no. 6900, pp. 838, 2002.

[35] P. Baudisch, D.D. Andrew, T. Duchowski, W.S. Geisler, Focusing on the Essential: Considering Attention in Display Design, *Magazine Communications of the ACM*, voi.46, no. 3, pp. 60-66, 2003.

[36] A. Hyrskykari, P. Majaranta, K.J. Raiha, From Gaze Control to Attentive Interfaces, *Proc. of the 11th International Conference on Human-Computer Interaction*, 2005.

[37] A. Hyrskykari, P. Majaranta, K.J. Raiha, Proactive Response to Eye Movements, *Proc. of the IFIP International Conference on Human-Computer Interaction*, 2003.

[38] Q. Ji, X. Yang, Real-Time Eye, Gaze, and Face Pose Tracking for Monitoring Driver Vigilance, *Real-Time Imaging*, vol. 8, no. 5, pp. 357-377, 2002.

[39] R. Vertegaal, I. Weevers, C. Sohn, GAZE-2: An Attentive Video Conferencing System, *Proc. of the CHI 02 Extended Abstracts on Human Factors in Computing Systems*, pp. 736-737, 2002.

[40] P. Viola, M.J. Jones, Robust real-time object detection, *proc. of 2nd international workshop on statistical and omputational theories of vision modeling, learning, computing, and sampling Vancouve*, 2001.

[41] G. Bradski, T. Darrell, I. Essa, J. Malik, P. Perona, S. Sclaroff, C. Tomasi, 2013, http ://sourceforge.net/projects/opencvlibrary/.

[42] M. Pantic, L. J. M. Rothkrantz, Expert system for automatic analysis of facial expression, *Image and Vision Computing*, vol. 18, no. 11, pp. 881V905, 2000.

[43] M. Pantic, M. Tomc, L. J. M. Rothkrantz, A hybrid approach to mouth features detection, *Proc. of the IEEE International Conference on Systems, Man, Cybernetics*, pp. 1188V1193, 2001.

[44] Hough_ transform,2013, http://en.wikipedia.org/wiki/Hough_ transform.

[45] Ketki K. Patil, S.D. Giripunje, Preeti R. Bajaj, Facial Expression Recognition and Head Tracking Using Gabor Filter, *proc. of the 3th International Conference on Emerging Trends in Engineering and Technology, IEEE Computer Society*,2010.

[46] F. Abdat, C. Maaoui, A. Pruski, Real facial feature points tracking with pyramidal lucas-kanade algorithm, *Proc. of the 17th IEEE International Symposium on Robot and Human Interactive Communication*, pp. 71-76, 2008.

[47] K. Arai, R. Mardiyanto, Eye-based HCI with full specification of mouse and keyboard using pupil knowledge in the gaze estimation, *Proc. of the 8th International Conference on Information Technology: New Generations*, pp. 423-428, 2011.

[48] Z.W. Zhu, Q. Ji, K. Fujimura, K. Lee,Combining Kalman filteringand mean shift for real time eye tracking under active IR illumination, *Proc. of the 16th International Conference on Pattern Recognition*, vol. 4, pp. 318-321, 2002.

[49] B. Raducanu, M. Pantic, L. J. M. Rothkrantz, M. Grana, Automatic eyebrow tracking using boundary Chain code, *Proc. of the International Conference on advanced School Computing Imaging*, pp. 137-143, 1999.

[50] J.L. Moreira, A. Braun, S.R. Musse, Eyes and Eyebrows Detection for Performance Driven Animation, *Proc. of the 23rd International Conference on Graphics, Patterns and Images*, pp. 17-24, 2010.

[51] M. Pantic, L.J.M. Rothkrantz, Facial Action Recognition for Facial Expression Analysis From Static Face Images, *IEEE Trans. on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 34, no. 3, pp. 1449-1461, 2004.

[52] F. ABDAT, C. Maaoui, A. Pruski, Human-computer interaction using emotion recognition from facial expression, *Proc. of the 5th UKSim European Symposium on Computer Modeling and Simulation*, pp. 196-201, 2011.

[53] S. Sural, Q. Gang, S. Pramanik, Segmentation and histogram generation using the hsv color space for image retrieval, *Proc. of the International Conference on Image Processing*, vol. 2, pp. 589-592, 2002.

[54] Luxand Face 4.0, Inc. http://www.luxand.com , 2013.

[55] J.Y. Bouguet, Pyramidal implementation of the lucas kanade feature tracker, *Intel Corporation, Microprocessor Research Labs*,2000.

[56] K. Lu, X. Zhang, Facial Expression Recognition from Image Sequences Based on Feature Points and Canonical Correlations, *Proc. of the International Conference on Artificial Intelligence and Computational Intelligence*, vol. 1, pp. 219-223, 2010.
[57] K.F. Kraiss, *Advanced Man-Machine Interaction*, Springer, Berlin Heidelberg , New York, 2006.
[58] X. Li, S. Dasgupta, Measurement Of User Satisfaction With Web-Based Information Systems:An Empirical Study, *Proc. of the 8th Americas Conference on Information System*, pp. 1149-1155, 2002.
[59] http://pbskids.org/daniel/games/trolley, 2013.