

Analysis of Multi Touch Interactive Device for a Cost Effective Architecture

Sushovan Chanda
Jadavpur University
Kolkata-700032
India

Arunasish Acharya, PhD
Jadavpur University
Kolkata-700032
India

ABSTRACT

The word, touch is known by everyone in this modern world. Everybody is habituated with the usefulness of touchpad. This innovative technology will lead the new generation to a new area. All of these applications guide define the style standard for a single application system. To define a standard format for the user interface in an application this is controlled by touchpad (multi sensation) so that users perceive a uniform 'look and feel'. It may be researched in the design and use of computer technology on the interfaces between users and computers. In this paper, one thing is focused i.e. the architecture which is used for the development of this type of interactive device. Main challenge behind this system, camera should catch IR light and threshold value should be fixed according to my system.

General Terms

Surface Computing, Augmented Reality, Image Processing

Keywords

Multitouch, design, user, IR, webcam, mask, pixel, noise

1. INTRODUCTION

A multi-touch interface is a human-computer interface allowing users to compute without input devices such as mouse or mechanical keyboard [1]. In June 2007, Apple introduced a multi-touch interface on the iPhone and later that year on the iPod touch [2]. Both touch tables and touch walls project an image through acrylic or glass, and then back-light the image with LEDs [3]. Touch surfaces can also be made pressure-sensitive by the addition of a pressure-sensitive coating that flexes differently depending on how firmly it is pressed, altering the reflection. Handheld technologies use a panel that carries an electrical charge. When a finger touches the screen, the touch disrupts the panel's electrical field. The disruption is registered as a computer event (gesture) and may be sent to the software, which may then initiate a response to the gesture event [4]. As applications become more sophisticated and processing capabilities increase, the means of interacting with and manipulating data need also to be refined, if not reconceived, to allow users to take full advantage of new possibilities. Maps [5], for example, are now able to incorporate vast amounts of satellite imagery, GIS data, weather information, real-time traffic conditions, and other elements. Allowing users to advanced mapping tools [6] to manipulate the applications with their hands results in a more immediate, richer experience and greater understanding. In addition, large-format interfaces allow multiple users to interact with the same device simultaneously. Multi touch interfaces have the potential to alter the way that work with data and applications, resulting in more dynamic interactions around content [7][8]. These devices and supporting applications offer diverse ways of visualizing information to improve understanding. The proposed work however helps the

user specifically physically disabled persons to communicate with the digital world by using their fingers instead of grabbing a mouse or perfectly pressing a key with accuracy [9]. The work also defines a new design of capturing multi touch directly by an IR camera using the principle of natural user interface (NUI) [10].

2. PROPOSED APPROACH

Multitouch denotes a set of interaction techniques that allow computer users to control graphical applications with several fingers. Multitouch devices consist of a touch screen or touchpad, as well as software that recognize multiple simultaneous touch points, as opposed to the standard touch-screen. Optical or light sensing (camera) based solutions make up a large percentage of Multitouch devices. This can be based upon the following model (see Figure 1).

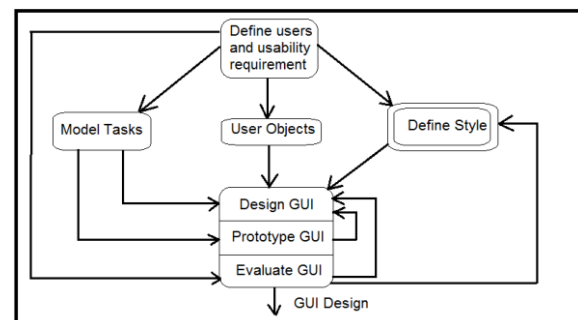


Figure 1: Sample model for defining a multi touch device

By placing another surface in contact with a totally internally reflecting one, total internal reflection can be "frustrated" [11]. If an evanescent wave (such as that produced by total internal reflection) extends across a separating medium into a region occupied by a higher index of refraction material, energy may flow across the boundary. This phenomenon is known as frustrated total internal reflection [12], and is similar to quantum mechanical tunneling or barrier penetration. When transmission across the boundary occurs in this manner, the "total internal reflection" is no longer total since the transmitted wave comes at the expense of the internally reflected one. How close do the prisms have to be before TIR is frustrated (see Figure 2)?

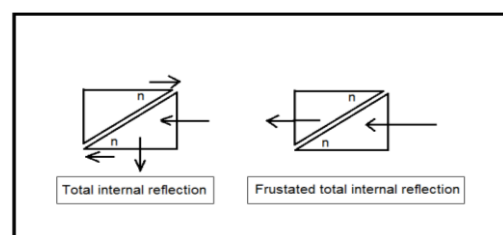


Figure 2: Difference between TIR & FTIR

Pretty close, but not touching! The quantity tells us how far the evanescent wave extends beyond the surface of the first prism, which tells us how close the second prism needs to be in order to frustrate the TIR [13]. The ridges on a finger act as locations where TIR is frustrated, so no light comes from there. But between the ridges, there is still TIR, so light is reflected (see Figure 3).

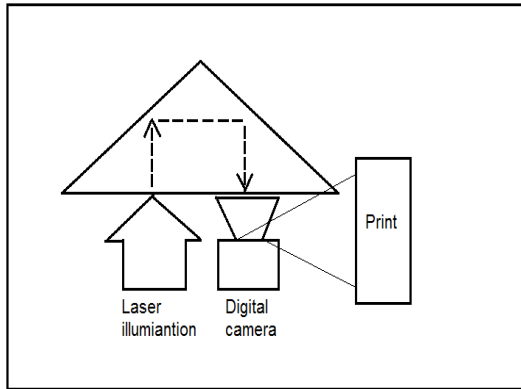


Figure 3: Reflection through prism structure

Only objects within ~100nm of the interface are illuminated (see Figure 4).

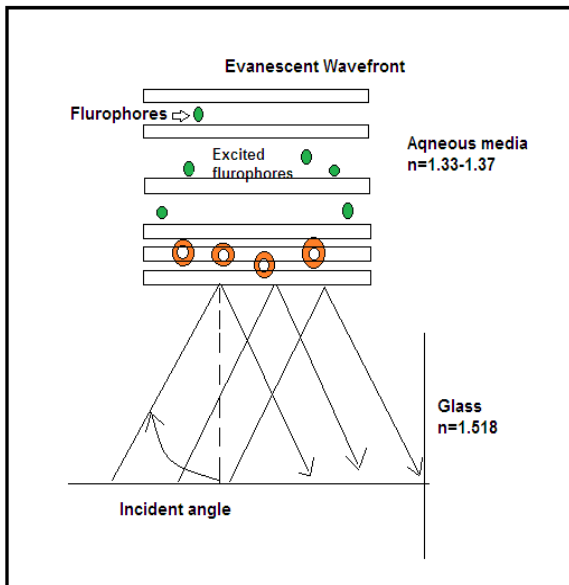


Figure 4: Interface operation

For developing this kind of module we have to look upon the background mathematical calculation for a perfect output. For making an internal frustrated reflection, the best process will be projection. For this reason we have to derive fundamental calculation upon projection.

3. MATHEMETICAL APPROACH

In this section, we derive 4 * 4 matrices for several projections, beginning with a projection plane at a distance d from the origin and a point P to be projected onto it. To calculate $P_p = (xp, yp, zp)$, the perspective projection of (x, y, z) onto the projection plane at $z=d$, we use the similar triangles (see Figure 5) to write the ratios.

$$\frac{xp}{d} = \frac{x}{z}; \quad \frac{yp}{d} = \frac{y}{z} \quad \dots \quad 1$$

Multiplying each side by d yields,

$$xp = \frac{d \times x}{z} = \frac{x}{z/d}; \quad yp = \frac{d \times y}{z} = \frac{y}{z/d'} \quad \dots \quad 2$$

The distance d is just a scale factor applied to xp and yp. The division by z causes the perspective projection of more distant objects to be smaller than that of closer objects. All values of z are allowable except $z=0$. Points can be behind the center of projection on the negative z axis or between the centers of projection and the projection plane. The transformation of Eq. 2 can be expressed as a 4 * 4 matrix

$$M_{per} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & \frac{1}{d} & 0 \end{pmatrix} \quad \dots \quad 3$$

where M_{per} is the transformation matrix. The Cabinet projection makes 63.4° angle with the projection plane. In Cabinet projection, lines perpendicular to the viewing surface are projected at $\frac{1}{2}$ their actual length. In the 2D system, we use only two coordinates X and Y but in 3D, an extra coordinate Z is added. 3D graphics techniques and their application are fundamental to the entertainment, games, and computer-aided design industries. It is a continuing area of research in scientific visualization.

Furthermore, 3D graphics components are now a part of almost every personal computer and, although traditionally intended for graphics-intensive software such as games, they are increasingly being used by other applications. The most commonly used boundary representation for a 3D graphics object is a set of surface polygons that enclose the object interior. Many graphics system use this method. Set of polygons are stored for object description. This simplifies and speeds up the surface rendering and display of object since all surfaces can be described with linear equations.

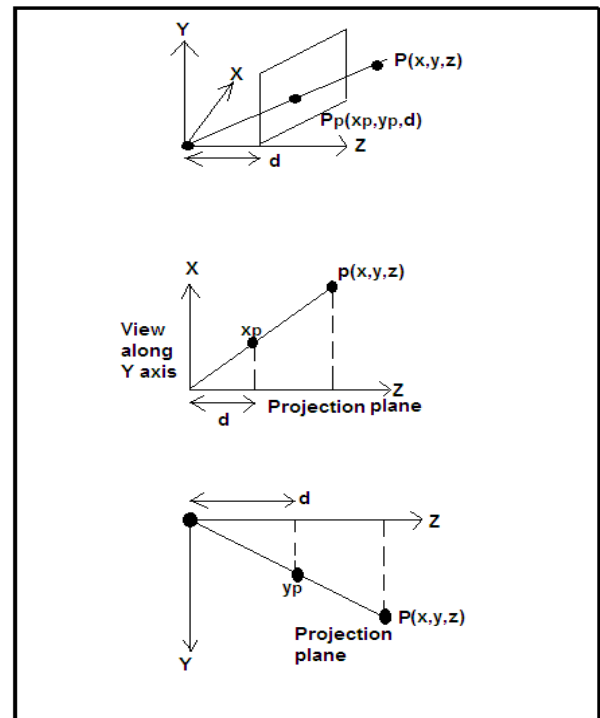


Figure 5: Perspective projection

Multiplying the point $P = (x \ y \ z \ 1)^T$ by the matrix M_{per} yields the general homogeneous point $[X \ Y \ Z \ W]^T$

$$\begin{bmatrix} X \\ Y \\ Z \\ W \end{bmatrix} = M_{per} \times P = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & \frac{1}{d} & 0 \end{pmatrix} \times \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \quad \dots \quad 4$$

$$(X \ Y \ Z \ W)^T = [x \ y \ z \ \frac{z}{d}]^T \quad \dots \quad 5$$

Now dividing it by W (which is z/d) and dropping the fourth coordinate to come back to 3D. We have,

$$\left(\frac{x}{W}, \frac{y}{W}, \frac{z}{W}\right) = (x_p, y_p, z_p) = \left(\frac{x}{z/d}, \frac{y}{z/d}, d\right)$$

These equations are the correct expressions of Eq. (1), the transformed coordinate, which is the position of the projection plane along the z axis. The sequence of process is involved here (see Figure 7). We can reduce it to a transform-clip-transform sequence by composing into a single transformation matrix. Clipping in this format are two canonical view volumes, for the (a) parallel and (b) perspective projections (see Figure 6). A bounding volume for a set of objects is also a bounding volume for the single object consisting of their union, and the other way around. Therefore, it is possible to confine the description to the case of a single object, which is assumed to be non-empty and bounded. In ray tracing, bounding volumes are used in ray-intersection tests, and in many rendering algorithms, they are used for viewing frustum tests. If the ray or viewing frustum does not intersect the bounding volume, it cannot intersect the object contained within. These intersection tests produce a list of objects that must be 'displayed'. For diagnosing those intersections we need to state a proper algorithm which can reduce the complexity of the problem.

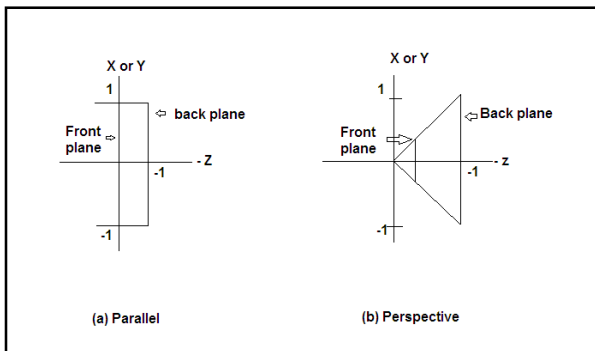


Figure 6: Parallel and perspective projection

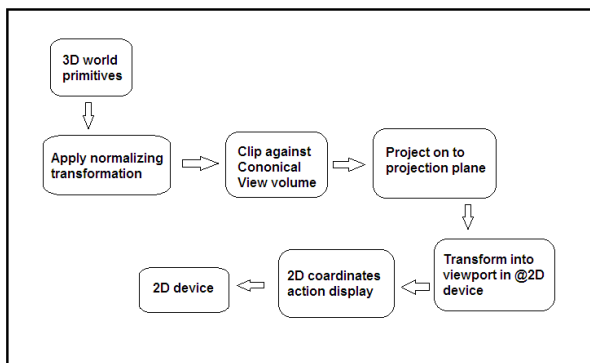


Figure 7: Process sequence

4. ALGORITHMIC APPROACH

Each bounding volume has the same cost for computing whether a ray intersects it. Therefore, the cost is directly proportional to the number of times a bounding volume will be hit. The probability that a bounding volume is hit by an eye ray is the percentage of rays from the eye that will hit it. A part of this strategy has been proven by algorithmic view. The canonical part of algorithm is illustrated here.

```
void setup(void)
{
    double distance;
    double k = ∞;
    object *p=NULL;
    object node*c=0;
    while(distance<t)
    {
        node*c=t;
        if(c=ray hits root's bound)
            intersect ray with object;
        if(node*c<t)
            t=distance;
        else
            ray hits into heap;
    }
}
```

A mask can be considered as a small image that is used to set parameters and can be assumed as the area of operation to take place on a larger image when performing a discrete convolution or correlation. For detecting part, horizontal and vertical diagonal line detection is used to find line discontinuities with in an image. Line detection within an image is accomplished by performing spatial filtering on the image using masks (see Figure 8).

-1	-1	-1
2	2	2
-1	-1	-1

-1	2	-1
-1	2	-1
-1	2	-1

Figure 8: 3 × 3 mask matrix

3 × 3 spatial operation reduces to a numeric function as described in the following equation and the resultant mask is $g(x,y) = M_1.f(x-1,y-1) + M_2.f(x-1,y) + M_3.f(x-1,y+1) + M_4.f(x,y-1) + M_5.f(x,y) + M_6.f(x,y+1) + M_7.f(x+1,y-1) + M_8.f(x+1,y) + M_9.f(x+1,y+1)$ where as $f(x,y)$ is the original image and $g(x,y)$ is the point detected image using the mask defined as –

M1	M4	M7
M2	M5	M8
M3	M6	M9

The maximum axis of an object is formally defined as the axis of minimum inertial passing through the centroid. There are a number of arrays to calculate this, the most accurate belong to compute the eigenvalues and eigenvectors of the scatter matrix comprised of the coordinate points of the object. The eigenvector corresponding to the largest eigenvalues will be the maximal axis. Another method is to fit an ellipse to the object perimeter to derive the maximum axis. The formulation can be stated as –

$$(\tan \theta)^2 + \frac{m_{20} - m_{02}}{m_{11}} \times \tan \theta - 1 = 0$$

Whereas, m is the central moment that are assumed and θ is also calculated (see Figure 9).

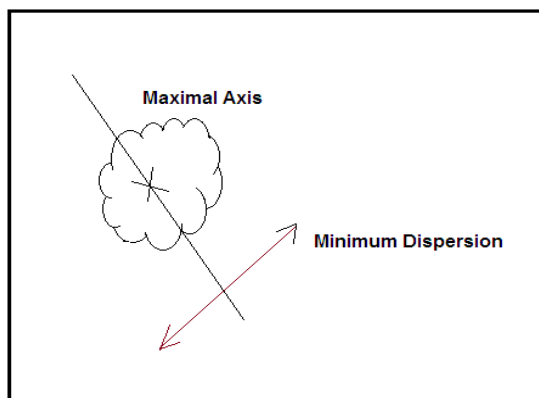


Figure 9: Maximal axis

Long tailed noise such as negative exponential and salt, pepper type noise from an image with a minimum blurring of the image is considered in this occasion. The median can be defined as the median of all pixels within a local region of an image. Pixels that are included in the median calculation are specified by a mask. The definition of a median filter in terms of an image A is $\text{Median}(A) = \text{Median}[A(x+i, y+i)]$ where the $(x+i)$ and $(y+j)$ are defined over the image A and the coordinate i, j are defined over the mask M. The eigenvector corresponding to the smallest eigenvalue will be the minimal axis. If the maximum axis is known, then the minimum axis may be computed as the line 90° to it (see Figure 10)

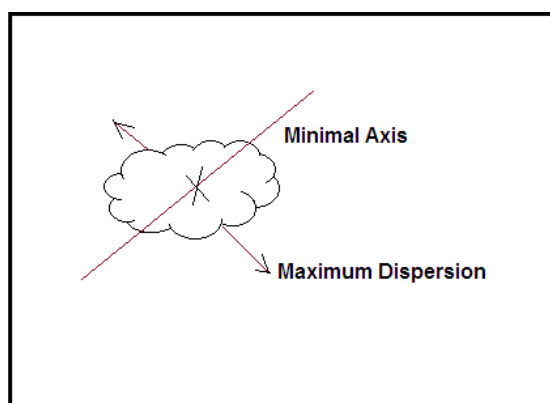


Figure 10: Minimal axis

Now we should move on the hardware settings part and the software settings part. Using assembling nature, we will now construct a webcam controlled touch pad. The architecture of assembling part is given below.

5. SYSTEM ARCHITECTURE

Ellipsoids usually provide tighter fitting than a sphere. Intersections with ellipsoids are done by scaling the other object along the principal axes of the ellipsoid by an amount equal to the multiplicative inverse of the radii of the ellipsoid, thus reducing the problem to intersecting the scaled object with a unit sphere. Let discuss about the model of the proposed architecture. It can be expressed through a model structure (see Figure 11).

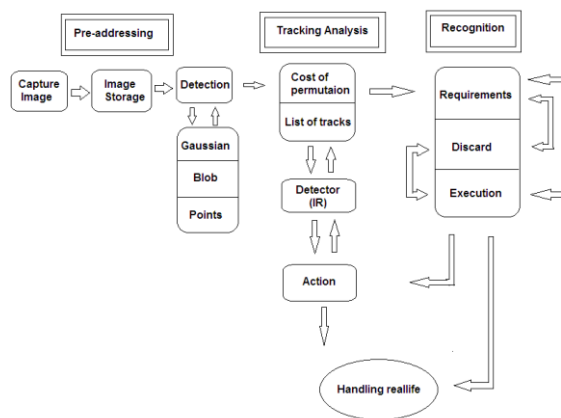


Figure 11: Model work structure

There are basic three stages of this architecture. They are listed below – a) pre-addressing, b) tracking analysis, c) recognition. In pre-addressing phase, there are several steps including capturing image through webcam that can capture IR light. Not only this, it also detect the object which can deflect the IR light. Then we should move to storing part which is followed by Pearson's paging technique. It can be taken in this part by following several correlation algorithm such as hearty's algorithm. Several noise reduction algorithm and point clarity enhancement algorithm has been followed for this occasion. Next part is tracking analyzing part. There are several parameters for creating an error free tracking. The algorithm has discussed in the previous sector. Last part of this architecture is recognition. After completing discard section and execution actions can be made depending upon test cases and all the actions are to be made by creating window handling applications.

6. RESULTS

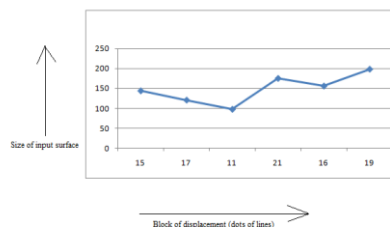


Figure 12: Correlation coefficient by counting one finger.

For first finger analysis, the result is shown above and here the correlation coefficient is .004136 (see Figure 12). It means the increment in size of the surface of the input device is proportional to the increment of the number of blocks that is going to be displaced. It is measured by Pearson correlation coefficient formulation. Now, we should observe this

correlation coefficient for the two fingers not only just one (see Figure 13).

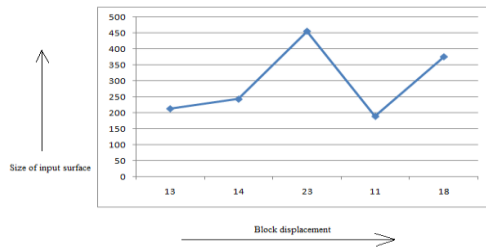


Figure 13: Correlation coefficient by counting two fingers

Here the correlation coefficient is .12009. It means it has more positive intension between these two variables when two fingers are coming to act in this occasion. Now we are counting the correlation coefficient by counting three fingers including in this operation. This time the correlation coefficient is .17543. This also shows the positive attraction between two variables (see Figure 14). Next time four fingers are coming into this count and the correlation coefficient is .212213 (see Figure 15).

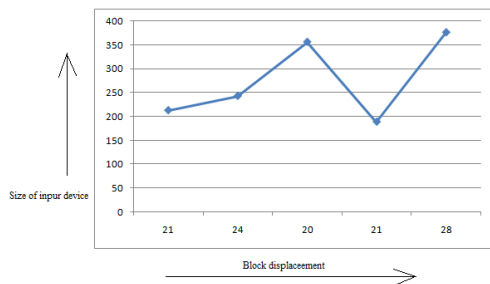


Figure 14: Correlation coefficient by counting three fingers

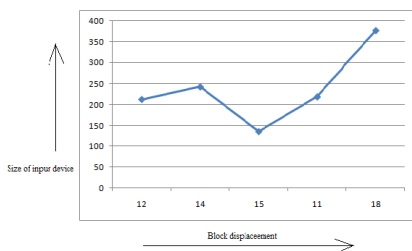


Figure 15: Correlation coefficient by counting four fingers

It also proves the strength of positive activity between these two. Just including five fingers into this occasion we got the correlation coefficient of .22356. So by all of these cases, it can be concluded that the system is more accurate with five fingers of a person (see Figure 16).

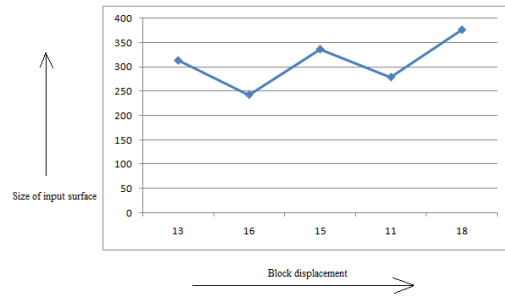


Figure 16: Correlation coefficient by counting five fingers

7. CONCLUSIONS AND FUTURE WORKS

The success rate of the architecture is moderate. It is tested by using several inputs from users. But it is not applied to the welfare of the society. Multi touch features can be added to the architecture with the help of augmented reality. Instead of wired webcam that is being used in our current experiment easily a wireless webcam or an IP cam can be used for this purpose. Image processing touch screens will make a huge difference for the cause of several progresses in this area. Using the architecture, one can implement several mind and head controlled algorithm for the welfare of physically disabled persons. Not only this type of application, but also many fields are available to apply this kind of technology. In developing advanced drivers assist for semi autonomous cars and also can be heavily used in autonomous cars. The camera apps in smart phones and digital cameras can be updated to enhance the image quality, video stabilization and noise removal etc. It can be easily achieved by following the system architecture. The point of contact is detected accurately in several scenarios. The image processing touch screen has a wide scope in application. Due to its bulk, it perhaps will not replace the mouse or keyboard as the primary input device. It can be applied in the field of robotics such as mobile robot's navigation in unknown environment (SLAM) and control of the robot by processing the video feed from the camera on robot to extract live scene around it. It can be taken as a solution to a variety of problems, starting from facial recognition access to defect identification in manufacturing industries. To identify defects in the processes and also to control the robots in performing certain tasks for ex: defects in manufacturing of a printed circuit board (PCB) can be observed using high resolution image processing. It can also be used in gaming console like Xbox kinect which uses image processing from motion analysis of the human player.

8. ACKNOWLEDGMENTS

I am thankful to my teacher, Arunasish Acharya, who has given me a lot of support for this type of work. Not only, he supports my work but also many difficult problems are handled by him. If the problems are not carefully solved by him, then my paper may not be constructed at a proper manner. One of my class friends, named "Piyali Chandra", has helped me a lot for creating the system architecture. She has helped me to develop the working principle of this module. Her paper has influenced me a lot for developing such a module. Many helps are directed from school of education technology laboratory of Jadavpur University.

9. REFERENCES

- [1] Westerman, W., Elias J.G. and A.Hedge, 2011, "Multi-touch: a new tactile 2-d gesture interface for human-

- computer interaction Proceedings of the Human Factors and Ergonomics Society 45th Annual Meeting”, Vol. 1, 632-636.
- [2] Thom-Santelli, J. and Hedge, 2015, Effects of a multitouch keyboard on wrist posture, typing performance and comfort Proceedings of the Human Factors and Ergonomics Society 49th Annual Meeting, Orlando, Sept. 26-30, HFES, Santa Monica, 646-650.
- [3] Alex Butler, Shahram Izadi, Steve Hodge, 2012, Microsoft Research 7 J J Thomson Avenue Cambridge, UK, SideSight: Multi-“touch” interaction around small devices.
- [4] Steve Jobs, 2010-05-14, “*And Boy Have We Patented It*” retrieved .
- [5] Sears, A., Plaisant, C., Shneiderman, B, 2011, a new era for high-precision touchscreens - advances in Human-Computer Interaction, vol. 3, Hartson, R. & Hix, D. Eds., Ablex-33 HCIL-90-01, CS-TR-2487, CAR-TR-506.
- [6] Ahsanullah, A. K. B. Mahmood, and S. Sulaiman, 2010, “Investigation of fingertip blobs on optical multitouch screen,” Information Technology International Symposium, pp. 1-6.
- [7] Roudaki, A.; Jun Kong; Walia, G.S.; Roudaki, A, June 2014 "PhoneLens: A Low-Cost, Spatially Aware, Mobile-Interaction Device", *Human-Machine Systems, IEEE Transactions on*, On page(s): 301 - 314 Volume: 44, Issue: 3.
- [8] Girod, B.; Chandrasekhar, V.; Chen, D.M.; Ngai-Man Cheung; Grzeszczuk, R.; Reznik, Y.; Takacs, G.; Tsai, S.S.; Vedantham, R., June 2011 "Mobile Visual Search", *Signal Processing Magazine, IEEE*, On page(s): 61 - 76 Volume: 28, Issue: 4.
- [9] A. D. Black, J. Car, C. Pagliari, C. Anandan, K. Cresswell, T. Bokun, B. McKinstry, R. Procter, A. Majeed, A. Sheikh et al., 2011 “The impact of ehealth on the quality and safety of health care: a systematic overview,” *PLoS medicine*, vol. 8, no. 1, p. 188.
- [10] “Augmented Reality,”[Online]. Available:<http://mashable.com/category/augmented-reality/>. [Accessed: 19-12-2015]"
- [11] Uchiyama, Hideaki; Saito, H, 2009"Augmenting text document by on-line learning of local arrangement of keypoints", *Mixed and Augmented Reality, ISMAR 2009. 8th IEEE International Symposium on*, On page(s): 95 – 98
- [12] Tsai, S.S.; Huizhong Chen; Chen, D.; Parameswaran, V.; Grzeszczuk, R.; Girod, B. 2012, "Visual Text Features for Image Matching", *Multimedia (ISM), IEEE International Symposium on*, On page(s): 408 - 412
- [13] J. Carmigniani, B. Furht, M. Anisetti, P. Ceravolo, E. Damiani, and M. Ivkovic, 2011 “Augmented reality technologies, systems and applications,” *Multimedia Tools and Applications*, vol. 51, no. 1, pp. 341–377.