Literature Review -3

Tejaswini Gaddam, 01672822

Primary Paper: Computer Vision for Interactive Computer Graphics, by William T. Freeman, David B.

Secondary Paper: 3D position, attitude and shape input using video tracking of hands and lips Andrew Blake1and Michael Isard1 Robotics Research Group, University of Oxford.

INTRODUCTION:

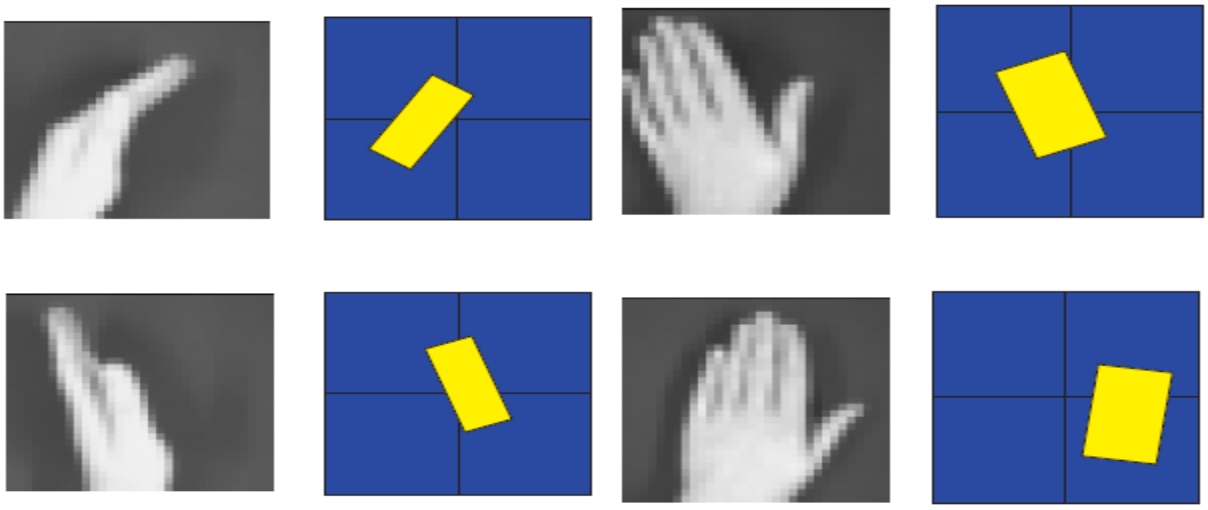
The author of primary paper thinks that rather that pressing buttons, players in a computer game can use gestures which computer can recognize. Author explains that people if use hand gestures to give commands to the machines or appliances that will be a potential benefit to surgeons, soldiers, or disabled parents. These vision-based interactions could make the machine interaction more enjoyable. Author describes various means of measuring these gestures which would reduce the complexity the machine may face. We can also say that the primary paper is the extension or the advanced of the secondary paper. The secondary paper portraits a system that would track accurately the curved outlines of the moving objects for example lips, legs, hands, vehicles, fruit etc. This can be done without using any other hardware other than the video camera. The author in the secondary paper presents 2 algorithms which allows the effective and agile tracking of the curves in live video using a workstation (SUN IPX). Thus, we can note that author of primary paper took secondary paper as the stong base in implementing more advanced features in developing the measurements which would track the gestures.

METHODS DISCUSSED:

Author of primary paper describes various fundamental visual measurements, in order of increasing complexity of the tracking of the images from physical objects. The author of primary paper converted the mathematical calculations used in referenced paper into more generic way. They are large object tracking, shape recognition, motion analysis, and small object tracking. Author used these to make vision-based interfaces for several computer games, plus hand gesture controllers for a toy robot, crane, and television set.

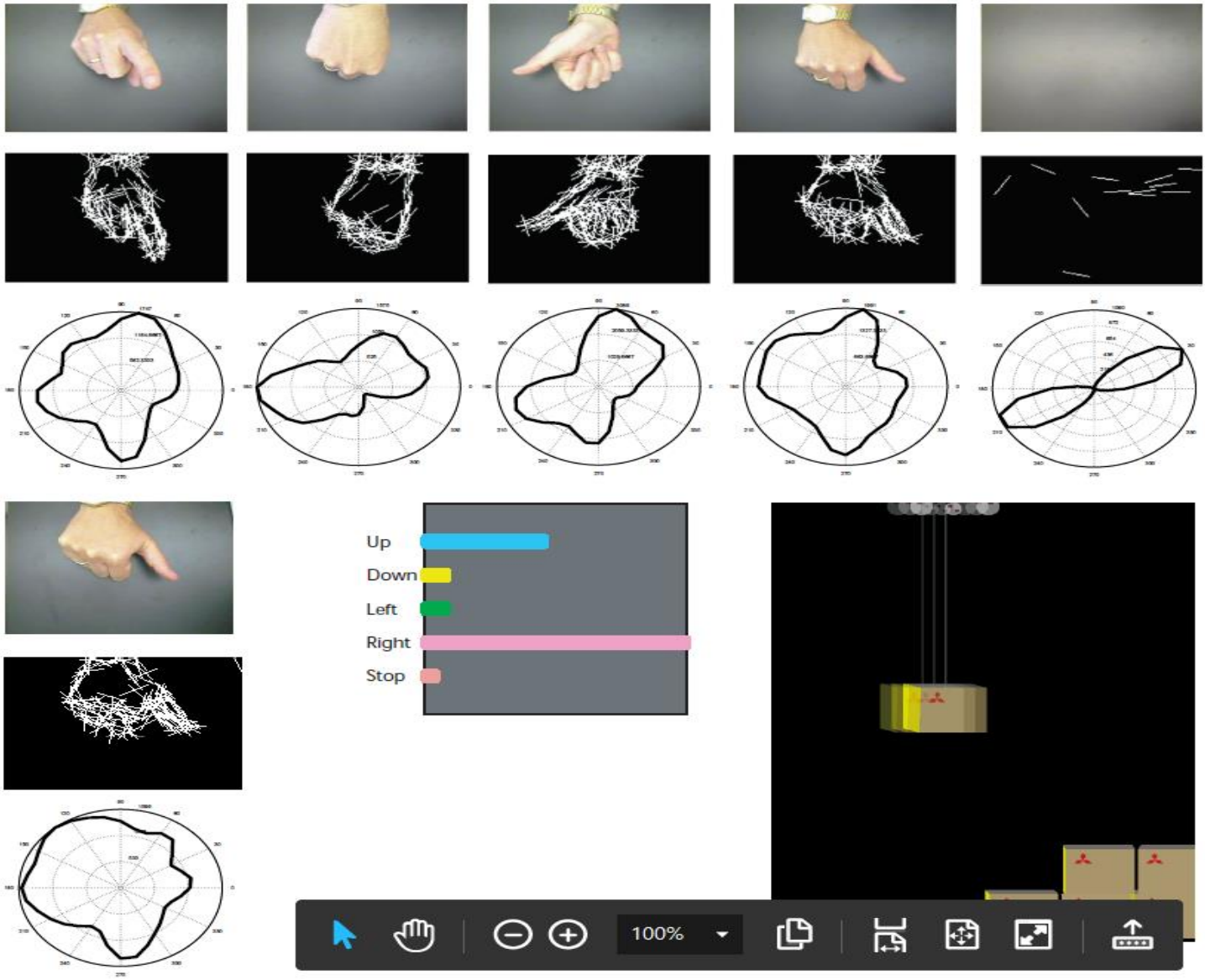
Large object tracking

In some interactive applications the computer needs to track the position or orientation of a body or hand. Relevant applications might include computer games or interactive machine control where the camera's viewing conditions are constrained. In such cases, describing the image's overall properties might suffice. If the camera views a hand on a uniform background, this method can distinguish hand positions and simple pointing gestures.



Author explains that we can calculate moments particularly quickly using a low-cost detector/processor called the artificial retina chip. This chip combines image detection with some low-level image processing (named artificial retina because the human retina combines those same abilities). The chip computes various functions that prove useful in the fast algorithms for interactive graphics applications.

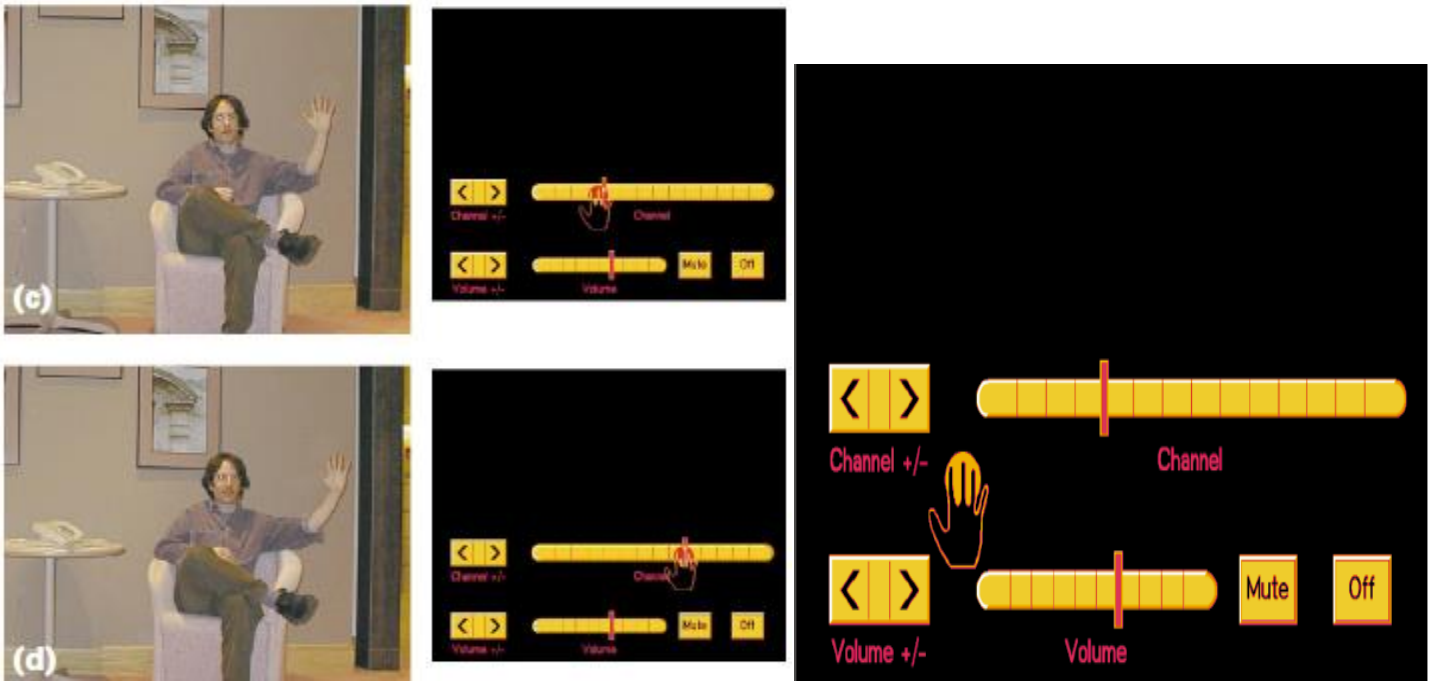
Shape recognition: Most applications, such as recognizing a static hand signal, require a richer description of the input object's shape than image moments provide. He explains that example-based applications involved two phases: training and running. In the training phase, the user shows the system one or more examples of a hand shape. The computer forms and stores the corresponding orientation histograms. In the run phase, the computer compares the current image's orientation histogram with each of the stored templates and either selects the category of the closest match or interpolates between templates, as appropriate.



Motion analysis:

Author says that a person's motion signals the important interface information to the computer. Computer vision methods to analyze "optical flow" can be used to sense movements or gestures. Author applied this to control the Sega Saturn game, *Decathlete*. The game involves the Olympic events of the decathlon. The conventional game interface suffers from the limitations of the handheld control to make the game athlete run faster, the player must press a key faster and faster. Here using motion analysis, the author makes the user runs in a place to make the computer character run.

Small object tracking
Author now shifts to the smaller objects. He explains that many interactive applications also require tracking objects, such as the user's hand, that comprise only a small part of the image. He explains this with an application which controls the television set by hand signals, thus he says that we can replace the remote control.



Author of Secondary paper proposed the concept of Kalman filters. He says that these filters comprise two steps: prediction and measurement assimilation. Prediction employs assumptions about object dynamics to extrapolate past motion from one video frame to the next. Assimilation blends measurements from a given frame with the latest available prediction.

The first algorithm explained, applies such a filter to curves represented by B-splines, to track both rigid and non-rigid motions. The second algorithm is a "system identification algorithm" based on ideas from adaptive control theory and "maximum likelihood estimation". Author says that previous approaches to learning shape variability have used statistical models to represent a family of possible shapes but statically. In contrast the learning method reported here is dynamic, using and modelling *temporal* image sequences. The tracked motion is used as a training set for the new algorithm which estimates the underlying dynamics of the training motion.

Author says that the effectiveness of the algorithms in generating agile trackers which are resistant to distraction from background clutter is demonstrated in this paper.

For rigid motion:
Tracking:
Author says that tracking problem is now to estimate the motion of some curve—in the paper he wrote he mentions that will be the outline of a hand or of lips. The underlying curve the physical truth is assumed to be describable as a B-spline of a certain predefined form with control points, varying over time. The tracker generates *estimates* of those control points. The aim the author says is that those estimates should represent a curve that, at each time-step, matches the underlying curve as closely as possible. The tracker consists, in accordance with standard practice in temporal filtering of two parts: a system model and a measurement model. Broadly, the measurement model specifies the positions along the curve at which measurements are made and how reliable they are. The system model specifies the likely dynamics of the curve over time, relative to some average shape.

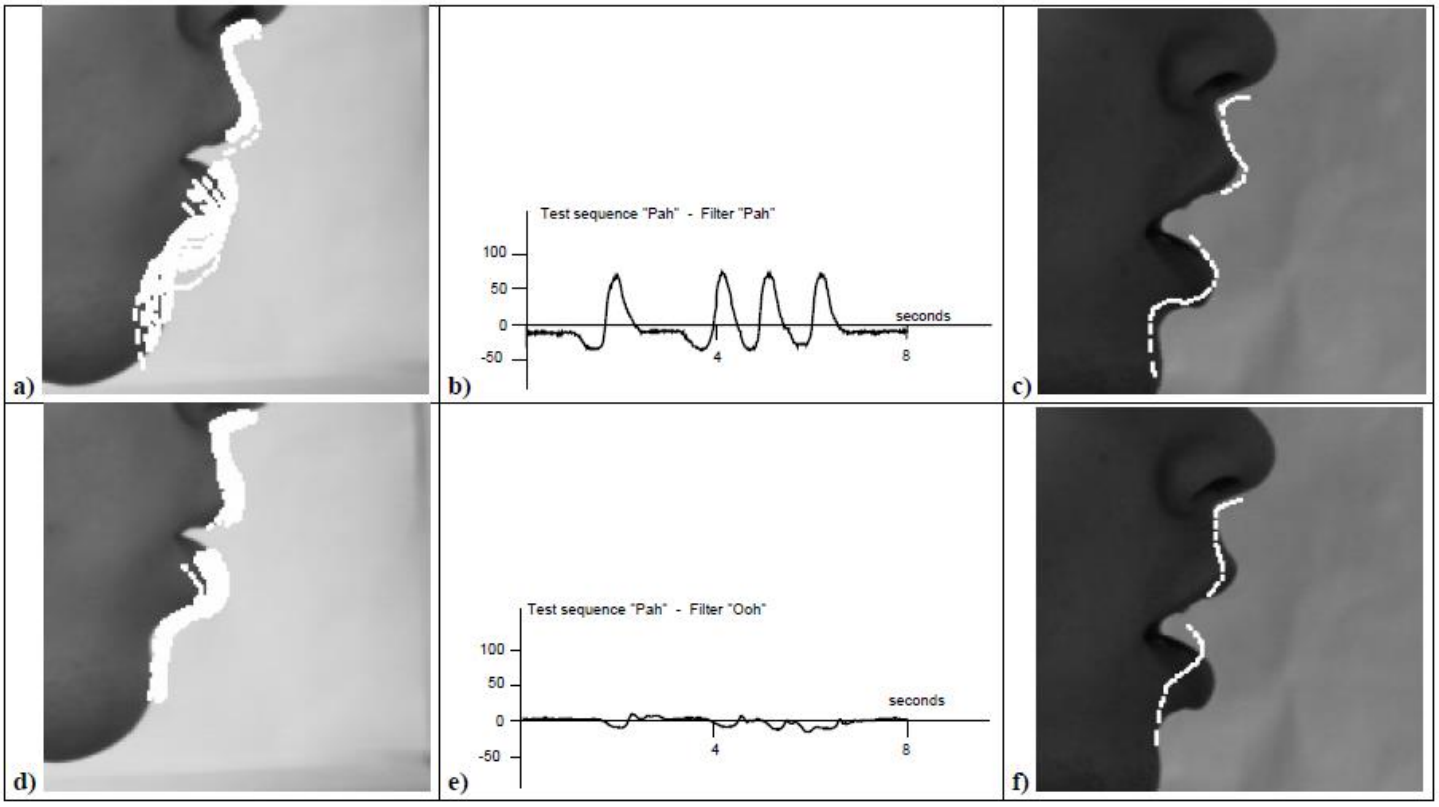
Tracking algorithm:
The tracking algorithm, a standard "steady state Kalman filter", consists of iterating the following equation:

$$\hat{\mathbf{x}}_{n+1} = A\hat{\mathbf{x}}_n + K\begin{pmatrix} \mathbf{z}_{n+1} \\ \mathbf{0} \end{pmatrix}$$

Translation:
Author here in the paper says that the model parameters are inferred from data by the statistical learning algorithm. The resulting tracker is specifically sensitive to a motion, in this case translation. Here the author describes that the learnt motion will also be re synthesized and compared with the original.

Non-rigid Motion:
The methods author explained still now are for learning and tracking rigid motion and these can be extended to non-rigid motion. An example for the non-rigid motion is the movement of the lips.

Lips
Author says that the default tracker should be capable of tracking slow speech and sufficient to gather data for training. As a demonstration of the effect of training, trackers were trained for the lip-motions that accompany the sounds "Pah" and "Ooh" and tested on the sound "Pah". It was clear from these results that the training effect for individual sounds is strong.



PROBLEMS AND SOLUTIONS:

In the primary paper the author also points out that in the large object Tracking where the motion based method was used always requires a stationary background, while the shape-based method requires a uniform background. In the shape recognition as well the system is insensitive to small changes in the size of the hand, is sensitive to changes in hand orientation. For greater robustness, the user may show several training examples, and the computer can use the closest matching example. Author also listed some of the observations where the user was not satisfied with the gesture classification. Figures below of hand two images that users feel should represent the same gesture. However, their orientation histograms are very different, as illustrated by their overlaid histograms.

When coming to the small object tracking author explains that there were problems on both sides. On the human side, the user wanted to give a broad set of commands to the television set by hand signals, yet the user didn't want to require an intensive training period before a user could control the television. On the machine side, author explains recognizing a broad set of hand gestures made within a complex, unpredictable visual scene, such as a living room—proves difficult and remains beyond the realm of current vision algorithms.

Solution to the problem faced:

Author thought that he could adopt a template-based technique, called *normalized correlation.* In this author examines a fit of hand template to every position in the analyzed image. The location of maximum correlation gives the candidate hand's position—the value of that correlation indicates the likelihood that the image region is a hand. The working range of the single template system was 6 to 10 feet from the television. To increase the processing speed, author says that we can restrict the field of view of the television's camera to 15 degrees when initially searching for the hand, and 25 degrees in tracking mode. Author also explains that we can use running temporal average of the image to subtract out stationary objects. Nonetheless, the best results occurred when the background contrasted with the foreground hand.

APPLICATIONS:

The author of secondary paper suggests that there are potentially many applications for real-time position, attitude and shape input using the new algorithms for learning and tracking. He explained two of the uses : the use of a hand as a 3D mouse and the tracking of lips for control of animation.

3D mouse:

Author says that both rigid and nonrigid motion of a hand can be used as a 3D input device. The freedom of movement of the hand is illustrated in figure, with rigid motion picked up to control 3D position and attitude, and nonrigid motion signaling button pressing and "lifting".

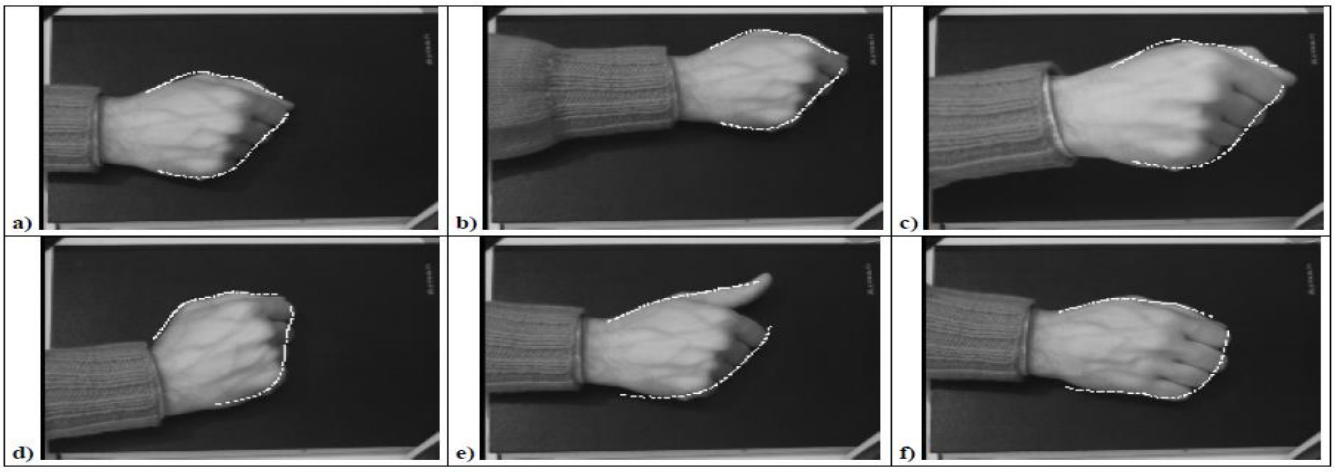


Figure 6: **The unadorned hand as a 3D mouse.** *A hand in its home position (a) can move on the xy-plane of the table (b) to act as a regular mouse but can also rise in the z direction (c) and the zooming effect is picked and used to compute z. Rotation can also be tracked (d). Note that measured affine distortions in the image plane are straightforwardly translated back into 3D displacements and rotations, as the demonstration of hand-controlled 3D motion on the accompanying video shows. Nonrigid motion tracking can be used to pick up signals. For instance (e) signals a button-press and (f) signals the analogue of lifting a conventional mouse to reposition it.*

CONCLUSION:

Here by I can conclude that the author of the primary paper took the ideologies form the secondary paper and implemented an advanced way to detect the gestures and used them for different purposes. The mistakes done in the secondary paper was the author used all the mathematical calculations which drove him to the wrong solution in some of the cases complex cases, but the author of primary paper corrected the mistakes and implemented more other features. He converted the mathematical calculations into more generic. The main aim of the authors of both the papers was to use the vision for more interactive computer graphics. The results of both the papers show that the rigid motion or non-rigid motion or position attitude and shape of the moving objects can be used to control the electronics in an easier way.

References:

Computer Vision for Interactive Computer Graphics, by William T. Freeman, David B.

3D position, attitude and shape input using video tracking of hands and lips Andrew Blake1and Michael Isard1 Robotics Research Group, University of Oxford.