Implementation of Low Cost Human Machine Interface (HMI) using Natural Feature Tracking (NFT)

Krishna Chauhan
Sagar Institute of Reasearch and Technology
Rajiv Gandhi Technical University
Bhopal, Madhya Pradesh
India

Sandeep Singh Padan
Sagar Institute of Reasearch and Technology
Rajiv Gandhi Technical University
Bhopal, Madhya Pradesh
India

ABSTRACT
This paper introduces an altogether a new, easy and user friendly approach for detection and tracking of the natural features such as a human finger in its natural form without the use of any added on artificial reference. The natural features are used to stabilize tracking process against external disturbances, noise or any occlusions. This paper emphasizes on integration of various technological concepts offering users a new and cheap way of studying & experiencing user friendly recognition of natural features without the need of any external and expensive hardware or software.

General Terms
Pattern and Gesture Recognition, Image Processing and Digital Communication

Keywords
Natural Feature Tracking, Natural and Artificial Features

1. INTRODUCTION
Modern day recognition and tracking systems require deployment of some artificial parameters which are used as a standard of reference. In addition to this, the field of gesture recognition, although being applicable in various novel uses worldwide, its technical knowhow is still far from the reach of the common user due to technological complexities and distant from the students community in particular due to the high economic costing involved in the development and application of this technology.

There are many systems designed and developed to trace the six-degree of freedom (6DOF) motion and movements of an object or a human with reference to a set of coordinate points in the surrounding environment for instance, FOXL98 [1] and WELC97 [2]. Such tracking systems comprise a fleet of various technologies as per the requirement of the application.

However, there is a separate class of augmented reality application systems which require a special marking or notation onto them to be tracked and recognized, for example, CAUD92 [3] and FEIN93 [4]. In some cases the object itself is used a point of reference. In many other cases, there arises a need of setting up some artificial references along with the object to be analyzed

However, these practices are not fool proof to ensure that the visible outcomes will be satisfied when it comes to real time application in natural conditions because in real time applications the tracking will depend upon all the features naturally occurring at the scene of recognition

This paper introduces an approach to implement AR tracking by using only natural features. It also highlights the unique combination of components and tools of recognition technology. Here, the idea of integration uncalibrated natural features with other artifacts at the scene in an AR tracking system is unique.

2. TRACKING NATURAL FEATURES
Natural features are features from unprepared scenes. Natural feature tracking (NFT) requires a lot of computationally expensive operations. Most previous natural feature tracking methods include heavy feature extraction and pattern matching procedures for each of the input image frame [5]. Classical NFT approaches repeat the feature extraction and matching procedures for each input frame. They try to match the extracted features from the input frame against each of the registered patterns until a successful pattern is found. The feature extraction and matching procedures require heavy computational cost which is hard to be allowed on low cost mobile devices. Feature tracking is the process of finding corresponding positions on the successive image to image points on the first image. The measure of correspondence is based on the similarity on an image neighbourhood of a fixed size window. A well known method is the feature tracker developed by Lucas and Kanade [6].

Let I and J be two consecutive gray scale images, and their scalar quantities $I(x, y)$ and $J(x, y)$ be the pixel intensities at image coordinates (x, y). The consecutive image frame J contains most parts of the first image frame I. The position (x, y) in I will be moved to a new position $(x + dx, y + dy)$ in J. The tracker must determine the disparity vector (dx, dy) at (x, y) from the intensity similarity of I and J. The similarity criterion is measured from the set of local neighbourhood pixels, denoted by W, centered at the position (x, y). The disparity is commonly computed by minimizing the residual error due to the brightness differences. This approach stably tracks small feature movements in video frames taken in rapid succession, but it is still not reliable enough when the feature movements between frames are farther apart in time.

Feature tracking is a necessary pre-processing step of the problem of structure from motion which finds the 3D structure of captured scenes from images sensed over time. Because the feature matches are the only preliminary information for further vision-based inference, conventional point-based tracking schemes try to seek as many feature points as possible. Most previous schemes of natural feature tracking have focused on the description and matching of features between two consecutive images. Their methods extract a new set of point features from each of the newly appeared image, instead of considering previous matched features. The extraction and matching of a new point set is time-consuming and should be avoided especially when the method is used for a real-time application. Conventional feature tracking approaches find correspondences between two consecutive images, namely I_t and I_{t+1}. They extract an initial set of feature

points from the first frame and track their movements along the consecutive image frames. Contrary to the conventional tracking approaches, the natural feature tracking approach in AR domains finds the correspondences between an on-the-fly image and one of pre-registered patterns. Previous methods of natural feature tracking have been focused on the matching of two point sets and they newly extract feature points on each of the on-the-fly image.

This paper is the outcome of an attempt to utilize the previous matched features to considerably bring down the economical expenses and development complexities of tracking process.

3. PRE REQUISITES & SETUP

Visual Studio 2005 or 2008, or Visual Studio Express Edition, .NET 3.0 or greater and "WebcamLib.dll" are essential as software prerequisites. A standard webcam with 640x480 resolutions was used. The resolution proved more than needed for analyzing gestures of one hand, so frames were captured at half resolution (320x240) to improve processing speed and to make sure that our approach is able to handle low-resolution images. The webcam was placed between one and two feet above a desk, pointing down toward the desk's surface. Various lighting conditions were used during development; the fluorescent ceiling lights in the Laboratories and one, two, or three lights during all times of day and night in a windowed room were tested. Problems encountered. The windowed room had more noise issues. This was not surprising because of the widely varying lighting. It was generally acceptable but very difficult to get a clean image out when the lighting was exceptionally bad. In addition, incandescent light bulbs used in the room produced yellowish light, which made hue-based extraction more difficult.

4. BACKGROUND SUBTRACTION

Background subtraction is the first approach to extracting a finger that was attempted. It is a very simple approach that is highly susceptible to errors from noise, so the first fifty frames of the video feed were accumulated as an entire background sequence to improve robustness over using a single frame. Often a scene will have objects moving through it that are not supposed to be part of the background. By taking the mean of many frames, these objects can be averaged out and only the static background will remain. The mean and variance of each pixel through this initial sequence is calculated and used to find the Mahalanobis distance between the sequence and any new frame. If the distance exceeds a threshold then the new pixel is sufficiently different from the background and is turned on in the final binary finger image.

Problems Encountered

While good images were occasionally possible, generally the shadows overwhelmed the image. This essentially caused two overlapping hands to be extracted, which resulted in a blob that could not easily be analyzed.

5. NEW POINT POSITION ESTIMATION, TIP CLASSIFICATION AND PHOTO RESIZING

Two different recursive filters have been designed and tested to estimate 3D positions of new points based on the camera pose and measurements of the feature image coordinates. The results of synthetic and real data showed that both filters converged and were stable. The EKF (Extended Kalman Filter) is known to have good characteristics under certain conditions, however the RAC (Recursive Average of Co-variances) filter gives comparable results, and it is simpler, operating completely in 3D-world space with 3D lines as measurements. The RAC filter approach eliminates the linearization processes required in the EKF with Jacobian matrices. As the algorithm depends upon curvature calculation from the contour, thus, although the algorithm is immune to rotation, any single peak will always be classified as a thumb. Also, in case the left hand is used the system will treat it like a right hand. In other words the system has no way to determine whether hand is a left or right hand and to distinguish between the fingers. However, the algorithm works very well when there is only one tip in the image. In case the user shows two fingers, disturbance may occur. Also any single finger tip is detected as the thumb and the image is resized according to the motion of that finger. Again if the finger moves too close to any of the corners then the false tips may result. Tracking of an identified tip is an issue. Another constraint is that the hand should not move very swiftly.

6. RESEARCH OUTPUTS AND RESULTS

Figure 1: Multiple recognitions in case of similar features: The colour of my finger and the skin of my face are similar; when kept in close affinity they both are getting recognized

Figure 2: Palm Recognition: The skin colour of my face & the skin colour of my palm are different due to varying brightness & contrast in my surrounding environment. Hence, the only the palm is getting recognized

Figure 3: The Desired Output for Pin Point Accuracy: Use of my Fingertip with my finger being used in its natural form

Figure 4: Information Control & Exchange: Image controlling using only my finger: Use of my finger tip without using any coloured agent (sketch pen) or any artificial material as an external parameter. Resizing of the image is also achieved by the same procedure
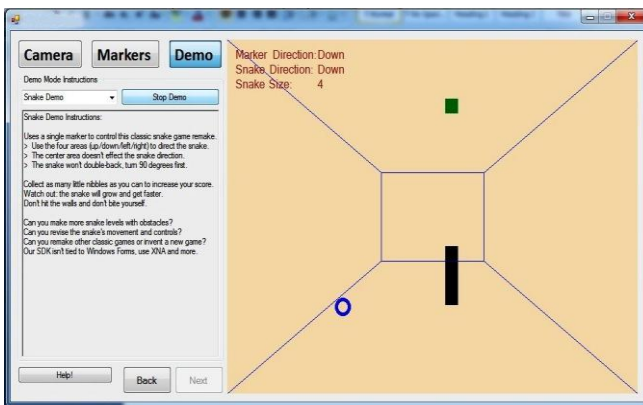

Figure 5: Gaming & Entertainment: The Blue circle is the point of my finger tip movement; Black line is the snake; Green Point as the target

7. FUTURE WORK

Noise remained an issue throughout testing, so finding a better way to extract the hand would be very useful. Specially, a histogram-based HSV extraction could be tried. The classifier lacked precision, so improving peak and valley detection is also a goal. One way of doing this could be through exploring

OpenCV methods for getting the hull around the hand and then detecting convexity defects that correspond to fingers. A larger Hu moment database could also be interesting because current testing has been done with very limited samples. Also, work was mainly done in a fairly constrained setting. Trying to deal with less ideal settings, such as one that could result in partially occluded gestures, could be very interesting.

Implementing a Natural Feature Tracking system is essential for efficient communication and interaction between humans & machines [7][8] since many reviews have been discussed so far [9][10][11][12]. In this paper we have come up with a very lucid and economical form of technology implementation as discussed above. In addition to this our research work is underway in order to develop a system to track multiple geometrical features [13][14] by keeping the basic element of our research intact i.e. no use of any artificial parameter.

8. ACKNOWLEDGEMENT

9. REFERENCES

[1].[FOXL98] E. Foxlin, M. Harrington, G. Pfeifer, "Constellation : A Wide-Range Wireless Motion-Tracking System for Augmented Reality and Virtual Set Applications", Proceedings of Siggraph98, Computer Graphics, pp. 371-378

[2].[WELC97] G. Welch, G. Bishop, "SCAAT: Incremental Tracking with Incomplete Information," Proceedings of Siggraph97, Computer Graphics, pp. 333-344

[3].[CAUD92] T. P. Caudell, D. M. Mizell, "Augmented Reality: An Application of Heads-Up Display Technology to Manual Manufacturing Processes," Proceedings of the Hawaii International Conference on Systems Sciences, pp. 659-669, 1992

[4].[FEIN93] S. Feiner, B. MacIntyre, D. Seligmann, "Knowledge-Based Augmented Reality," Communications of the ACM, Vol. 36, No. 7, pp 52-62, July 1993

[5].Lim, M.J., Jung, H.W., Lee, K.Y.: Game-type recognition rehabilitation system based on augmented reality through object understanding. The Journal of the Institute of Webcasting, Internet and Telecommunication 11(3) (2011) 93–98

[6].Lucas, B., Kanade, T.: An iterative image registration technique with an application to stereo vision. In: Proceedings of the 7th International Joint Conference on Artificial Intelligence. (1981) 674–679

[7] Q. Cai, D. Gallup, C. Zhang, and Z. Zhang. 3d deformable face tracking with a commodity depth camera. In Proc. of IEEE ECCV, 2010.

[8] J. Shotton, A. Fitzgibbon, and etc. Real-time human pose recognition in parts from single depth images. In Proc. Of IEEE CVPR, 2011.

[9]. Rafiqul Zaman Khan1 & Noor Adnan Ibraheem, (2012) "Survey on Gesture Recognition for Hand Image Postures", International Journal of Computer and Information Science, Vol. 5, No. 3, pp 110- 121. doi:10.5539/cis.v5n3p110

[10]. Joseph J. LaViola Jr. (1999) "A Survey of Hand Posture and Gesture Recognition Techniques and Technology", Master Thesis, NSF Science and Technology Center for Computer Graphics and Scientific Visualization, USA.

[11] Thomas B. Moeslund & Erik Granum (2001) "A Survey of Computer Vision-Based Human Motion Capture," Elsevier, Computer Vision and Image Understanding, Vol. 81, pp 231 268.

[12] T. R. Trigo & S. R. M. Pellegrino, (2010) "An Analysis of Features for Hand-Gesture Classification", 17th International Conference on Systems, Signals and Image Processing (IWSSIP 2010), pp 412-415

[13] Luigi Lamberti & Francesco Camastra, (2011) "Real-Time Hand Gesture Recognition Using a Color Glove", Springer 16th international conference on Image analysis and processing: Part I (ICIAP'11), pp 365-373.

[14] Mokhtar M. Hasan, and Pramod K. Mishra, (2012) "Hand Gesture Modeling and Recognition using Geometric Features: A Review", Canadian Journal on Image Processing and Computer Vision Vol. 3, No.1.