Google *Summer of Code 2009*

Project Proposal:

Recognition, Tracking and Association of Hands, Fingers, and Blobs: A Tbeta Upgrade

Mentoring Organization

Author: **Thiago de Freitas Oliveira Araújo**

Contact: thiago.oliveira<at>ee.ufcg.edu.br

Freenode IRC Nick: degekbr

1. Introduction

This Project Proposal consists on the addition of Blob, Hand and Finger Recognition, Tracking and Association as core functionality to the Tbeta for using with DI/DSI multi-touch setups. This functionality may imply on a wide sort of applications that could be imagined and introduced by the multi-touch community using Tbeta. The idea is to try to be as much as possible "setup independent". Moreover, this is an opportunity for introducing not only the author but his lab project to the contributors of NUI Group. That would be an honor and unique opportunity them. Moreover, as can be noticed on figures 2 and 3, the hands/fingers/blobs are visible using the DSI and DI and especially filtering and enhancement steps they become characteristics that can be extracted from the image with its processing. Another possibly, although not previously considered, was found after discussions with *Seth Sandler (*from NUI Group), *that* is that the people involved with FTIR setups with interest on this new Tbeta functionality could make modifications on their hardware, using special materials or light compensation (although it would require a more robust system for background subtraction/image segmentation) in such a way that the hand could become visible for further image processing techniques. That may be clearer after a quick look on Fig. 4 and 5.

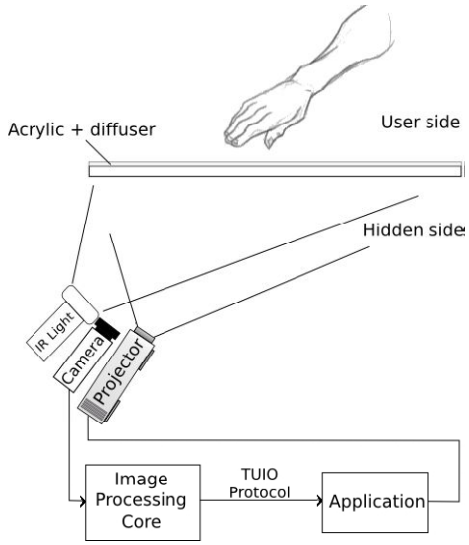

Fig 1. *Current Setup from the Embedded(http://embedded.ufcg.edu.br) Lab.*

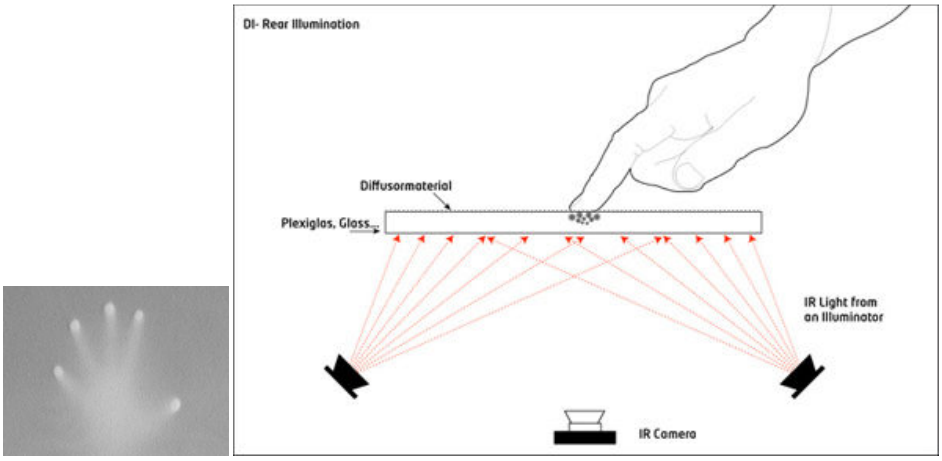

Fig 2. *DI explanation from the NUI Group forum and hand view using this technology.*

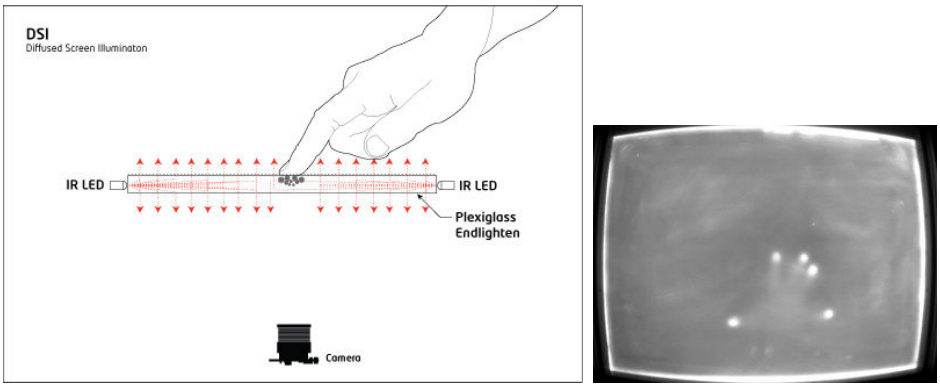

Fig 3. *DI explanation from the NUI Group forum and hand view using this technology.*

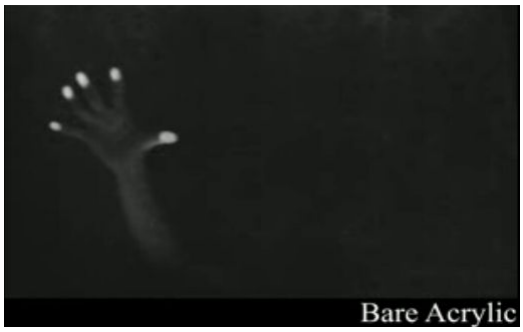

Fig 4. *FTIR possibilities with special materials*

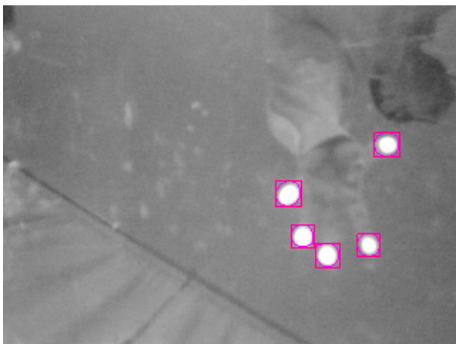

Fig 5. *FTIR possibilities with background light*

2. The Tbeta

As described by its own developers: *"The Beta, **tbeta** for short, is an **open source/cross-platform solution** for **computer vision** and **multi-touch sensing.** It takes a video input stream and outputs tracking data (e.g. coordinates and blob size) and touch events (e.g. finger down, moved and released) that are used in building multi-touch applications. tbeta can interface with various web cameras and video devices as well as connect to various TUIO/OSC enabled applications and supports many multi-touch lighting techniques including: FTIR, DI, DSI, and LLP with expansion planned for the future (custom modules/filters)."* The main aspect of this purpose is then to provide the following implemented algorithms as core functionalities to the Tbeta, in such a way that they are spread and available to every member from the multitouch community to use on its own applications.

3. Blob Recognition

The blobs may be found using a threshold method similarly to the one discussed in [10]. Tbeta already has a blob recognition algorithm implemented on OpenCV what eases the process of algorithms combination.

Fig 6. *Blob Recognition*

4. Finger Recognition

The problem of finger recognition is a relatively well studied topic [16] [17] [18]. Sigurdsson and Wong discuss finger recognition for cropping applications [16], Malik investigates the real-time finger tracking for interaction with computers [17] whereas Letessier and Berardand study specifically the finger tracking problem in interactive displays [18]. Concerning the finger recognition and tracking, some interesting solutions were proposed, such as crevices detection [19] used for tracking fingers on a pianist hand. The missing corner is the dual association between those fingers and the hand, what is necessary for our proposed applications. The current approach that I have implemented restricts the finger recognition to the HROI (Region of Interest where a Hand was found) and so it becomes quite simple, and mainly based on the Sobel [20] Gradient on the Y direction (Figure 3). A(x; y) denotes the digital image taken from the camera which after the gradient transformation is forwarded to the edge detection based on the Canny Algorithm [21]. The Hough lines based algorithm [22] has been evaluated, however in the final implementation the detector was based on the border following algorithm proposed by Suzuki [23]. As a result, the finger positions can be obtained and an ID may be associated, similarly to what has been done for the hands.

Fig 7. *Finger Recognition*

5. Hand Recognition

The hand recognition approach already implemented was based on the state of the art work of Viola and Jones [11] using Haar-like features [12] as shown on Figure 6. The method allows very fast feature evaluation because of the image representation as an integral image [11]. Kölsch and Turk [13] implemented and analyzed the application of the same method on a hand detection system and obtained interesting results, although criticizing the costs for a computationally expensive training method, what prohibits the evaluation of many hand appearances for their suitability to detection. Ong and Bowden [14] state that the training may have an excessive number of weak classifiers for each strong layer. However, on our system the hand positions are not as flexible as the analyzed on those systems, in such a way that the main hand positions may be evaluated by the classifiers and the high accuracy performed by the detector is adequate to justify the "one-time"costs for training. However, there is an ongoing study for these algorithms as all the aspects are wanted to be independent of platform to be more easily integrated to Tbeta. Specially,Hand Geometry recognition according to fingers or blobs positions appears also as a reasonable approach for implementing this functionality to Tbeta. Also there is the approach of constructing a "virtual hand" that could try to "connect" itself to blobs found on the screen and check distances between blobs and between blobs and its "virtual palm" to guarantee that they belong to the same hand, and provide then a simple and fast way for further tracking.

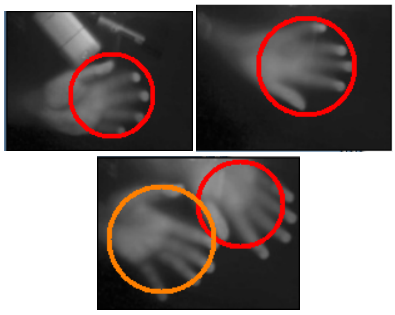

Fig 8. *Hand Recognition using images obtained from our setup after filtering and enhancement steps.*

6. Finger, Blob and Hand Association

This section shows the main ideas behind the purpose. After finding the Hand, the Blob and the fingers, they should be associated, and the blob TUIO would be modified to support the additional information that can be provided to the developer by the Tbeta running server. The figure 10 shows a multi-touch surface where there is the interaction of a Hand and a Blob and a Finger from this Hand. It is also visible that to each entity, an ID is associated, so that this new information may be integrated with

the TUIO string. Furthermore, this adapted string is shown below the figure 9, and where *FID* means finger ID, *xf* the finger x position, *yf* the finger y position, *HID* means the hand ID, and *xH* and *yH* the position of the center of the hand. The right way to do this inclusion would be to create a custom profile on the TUIO protocol and not to extend a 2DObj, however this would consequently require a very consistent definition of this custom profile , what was avoided to guarantee that this can be better discussed and implemented according to the mentors opinion.

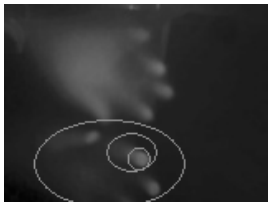

Fig 9. *Finger, Hand and Blob Association*

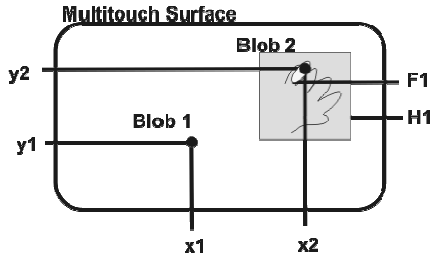

Fig 10. *Finger, Hand and Blob Association Schematics on Surface*

An example of a TUIO string is:

/tuio/2Dobj set s i x y a X Y A m r

The new string would be:

/tuio/2Dobj set s i x y a X Y A m r FID xf yf HID xH yH

7. Tracking

Tracking algorithms need then to be implemented to all the entities found. One idea for fingers is the same of the one used for pedestrian tracking using an Extended Kalman Filter approach, that may be used with Kinematic Models of the entity analyzed or with the Geometry Model. For hands, some algorithms using Mean-Shift Tracking seems to be the most reasonable for achieving such task. For hands, the current most exciting approach is Pseudo Hand-Tracking like the one that seems to be used on the commercial Multi-touch Cell (http://www.multitouch.fi).

Fig 11. *Pedestrian Tracking with EKF*

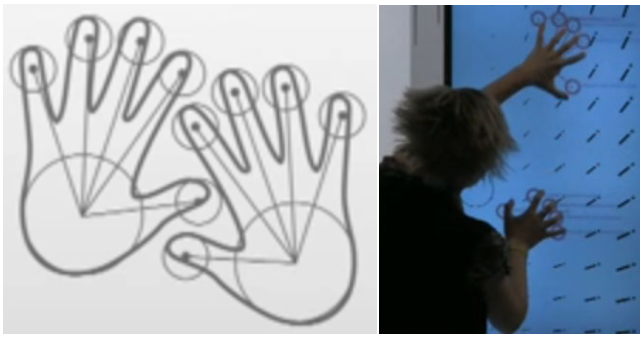

Fig 12. *Pseudo Hand Recognition used on Multitouch.fi*

8. Stereo Vision Analysis

To enjoy the full possibilities of hand tracking on DI/DSI multi-touch setups the system needs to work on the 3D plane, to achieve mechanisms for concluding the real position and situation of the hand , if “really touching”, or “almost touching”. The hand recognition may also be optimized using quadrics geometry model or projecting it to the 2D plane.

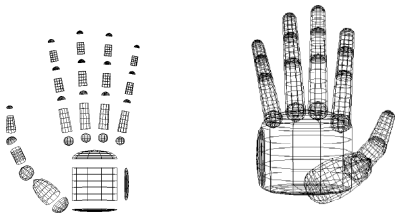

Fig 13. *3D Hand Geometry*

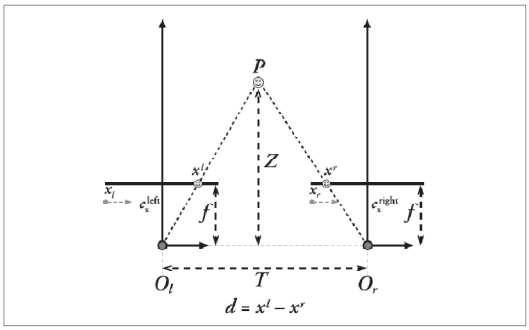

Fig 14. *Stereo Vision sample axes*

9. Schedule

0º SPRINT (20 APRIL – 07 MAY)

Discussion with mentors about the Tbeta architecture and the approach to tackle the challenges proposed. Also, it is time for studying deeply the tracking algorithms.

1º SPRINT (08 MAY – 22 MAY)

Some tests are made implementing the algorithms with OpenCV outside of Tbeta code.

2º SPRINT (23 MAY - 22 JUNE)

This period corresponds to the development of the code for integrating the algorithms to the Tbeta. These are the results to be submitted to mid-term evaluation.

3º SPRINT (23 JUNE – 02 AUGUST)

After mid-term evaluation this period is used for discussing and analyzing the results with the mentors and to establish the priorities for successful completion of the GSOC proposal. The initial idea is to allocate it mainly for stereo vision hand-tracking inclusion on the Tbeta plugin. A beta version may be released for community test of the plugin.

4º SPRINT (03 AUGUST – 17 AUGUST)

Project conclusion. This release includes project packaging through Google Summer of Code and NUI Group sites. Also the code fully documentation is provided within a tutorial, if needed for adapting your setup to be able to enjoy the new functionalities of tbeta.

5º SPRINT (17 AUGUST – …)

Time to enjoy the t-shirt, the community feedback and the money!

10. Conclusions

It seems that a plugin that could implement those functionalities on top of the Tbeta, could enhance and augment considerably the kind of applications the community is working right now. The implementation of the concepts established until the section 7, are the basis of the project purpose. The eight consists more of future work, and what could be done to enhance the system. The development of these techniques should take the 3 months established for the GSOC. After completion, it should be fully documented and added to the Tbeta for the whole community to enjoy the functionality.

11. Author Information

Name: Thiago de Freitas Oliveira Araújo

Email: thiago.oliveira@ee.ufcg.edu.br , thiagodefreitas@gmail.com
Location/Timezone: GMT -3

Website: http://embedded.ufcg.edu.br/~thiagooliveira

Age: 22

Education/Qualifications: Undergraduate Student - Electrical Engineering

Academic and Industry Background: I am an undergraduate student on Electrical Engineering (emphasis on Control and Automation. Currently I am a researcher and developer on the Embedded Systems and Pervasive Computing Laboratory on the Universidade Federal de Campina Grande, on a Multitouch Project, in cooperation with the Instituto Nokia de Tecnologia (Nokia Institute of Technology), where my work is mainly conducted on computer vision, especially with hand, blob and finger recognition and association. I have worked for a computer manufacturer company customizing Linux Distributions for easing their usage for people that are not acquainted with Linux. And also, I am a member of the Brasil-VR team, the Latin American Champion and qualified for the Robocup Worldwide Championship in Graz, Austria 2009, on the Robocup Virtual Robots Rescue Simulation League. My main research interests are Robotics, Computer Vision, SLAM(Simultaneous Localization and Mapping), Probabilistic Robotics, etc.

Open source development experience: My open source experience is mostly from the work cited above. For the project proposal the main experience in my opinion is the

one conducted in C++ for our code to the Robocup. Also, OpenCV experience for the development on the Multitouch Project. Have developed a python Open Source API for interacting Lego Mindstorms NXT with Nokia Linux OS based Devices. And also, a C++ platform for robots interaction with Wireless Networks.

Development Methodologies: The work is going to be conducted within 20 hours each week and continuous contact/ online meetings with the mentors. A project tracking technology is definitely going to be adopted. The project proposal consists on a research and development one, so some part of the time is going to be allocated to the first matter. The implementation will be done on the C++ Language, using the Linux Platform and the CodeBlocks IDE, as the implementation is going to be on the top of Tbeta, and there is a project from this IDE on the Tbeta repository. Also, the algorithms implementation is going to use the OpenCV library on the above mentioned language.

Project Proposal: The proposal is to provide a plugin for Tbeta that would allow the association between hand, blobs and fingers, as well as it tracking, augmenting the opportunities for applications that could make use of this additional information.

The reason for picking specific project:

The Reason for picking the specific project is that it is something that I am currently researching and working on, as well as these additions to Tbeta, would be really useful for my lab's application developers. And also, that would be an amazing opportunity to get even more involved to the nui community and give a small/although really wanted contribution for the enhancement of this wonderful application.

Activity level within the NUI Group Community:

Unfortunately I have not been as involved as I wanted with the NUI Group Community, however the GSOC 09 seems as a wordless opportunity for interacting and contributing with the community, especially with the "gurus", and developers related to the touchlib/tbeta community.

Time working multi-touch, natural user interfaces, HCI and related fields.

Specifically with multi-touch, I have been working for 6 months, but for HCI and related fields I have been working for more than 2, 5 years.

We also allow space for optional details the student sees fit.

Gladly , I am being fully supported by the lab, and by professors from my university, including technical and financial and comprehension support for being able to achieve the things proposed and even negotiating on the top of it for guaranteeing hopefully the acceptance to the GSOC 09, as a student mentored by the NUI Community.

12. References

Part of the text is adapted from the work: “Detecting Hands, Fingers and Blobs for Multitouch Display Applications”, Thiago de F. O. Araújo , Alexsandro J. V. dos Santos, Antonio M. N. Lima, submitted for review to the PAR 2009 , Special Session of the HPCS 2009 conference.

1] L. Muller, “Multi-touch displays: design, applications and performance evaluation,” Universiteit van Amsterdam, GRID-COMPUTING Master Thesis, 2008, 20th June 2008.

[2] M. Kaltenbrunner and R. Bencina, “reactivision: A computer-vision framework for table-based tangible interaction,” Proceedings of the first international conference on ”Tangible and Embedded Interaction” (TEI07). Baton Rouge, Louisiana., 2007.

[3] NuiGroup, “Touchlib: A multi-touch development kit,” http://www.nuigroup.com/touchlib/, 2009, last access on.

[4] ——, “Tbeta,” http://tbeta.nuigroup.com/, 2009, last access on.

[5] G. Bradski and A. Kaehler, Learning OpenCV, Computer Vision with the OpenCV Library, 1st ed. O’Reilly, 2008.

[6] A. Rosenfeld, “Some uses of pyramids in image processing and segmentation,” Proceedings of the DARPA Imaging Understanding Workshop, pp. 112–120, 1980.

[7] C.Tomasi and R.Manduchi, “Bilateral filtering for gray and color images,” Proceedings of the 1998 IEEE International Conference on Computer Vision, Bombay, India, 1998.

[8] J. Serra, “Image analysis and mathematical morphology,” Academic Press, 1984.

[9] R. J. GARDNER, “The brunn-minkowski inequality,” BULLETIN (New Series) OF THE AMERICAN MATHEMATICAL SOCIETY, vol. 39, no. 3, pp. 355–405, 2002.

[10] L. Li, W. Huang, I. Y. H. Gu, and Q. Tian, “Foreground object detection for videos containing complex background,” Proceedings of the Eleventh ACM International Conference on Multimedia, 2003.

[11] P. Viola and M. Jones, “Robust real-time object detection,” International Journal of Computer Vision, 2002.
[12] R. Lienhard and J. Maydt, “An extended set of haar-like features for rapid object detection,” Proceedings of the IEEE International Conference on Image Processing, pp. 900–903, 2002.

[13] M. Kolsch and M. Turk, “Robust hand detection,” IEEE International Conference on Automatic Face and Gesture Recognition, 2004.

[14] E.-J. Ong and R. Bowden, “A boosted classifier tree for hand shape detection,” Proceedings of the Sixth IEEE International Conference on Automatic Face and Gesture Recognition (FGR’04), 2004.

[15] Urtho, “Urtho - Negative face Dataset,” http://face.urtho.net/, Dead Link., 2006, can now be found in http://tutorialhaartraining.googlecode.com/svn/trunk/data/negatives/ , Shared by Naotoshi Seo.

[16] G. Sigurdsson and A. Wong, “Finger tracking for rapid cropping applications,” CS223B Winter Quarter 2008, 2008.

[17] S. Malik, “Real-Time Hand Tracking and Finger Tracking for Interaction,” CSC2503F Project Report, 2003.

[18] J. Letessier and F. Berard, “Visual tracking of bare fingers for interactive surfaces,” Proceedings of the ACM Symposium on User Interface Software and Technology(UIST 2004), 2004.

[19] D. O. Gorodnichy and A. Yogeswaran, “Detection and tracking of pianist hands and fingers,” Proceedings of the 3rd Canadian Conference on Computer and Robot Vision (CRV’06), 2006.

[20] I. Sobel and G. Feldman, “A 3x3 Isotropic Gradient Operator for Image Processing,” Talk at the Stanford Artificial Project, 1968.

[21] J. Canny, “A computational approach for edge detection,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. PAMI-8, no. 6, pp. 679–698, 1998.

[22] J. Matas, C. Galambos, and J. Kittler, “Robust detection of lines using the progressive probabilistic hough transform,” Computer Vision and Image Understanding - Academic Press, vol. 78, no. 1, pp. 119–137, 2000.

[23] S. Suzuki and K. Abe, “Topological structural analysis of digital binary images by border following,” Computer Vision, Graphics and Image Processing, no. 30, pp. 32–46, 1985.

[24] M. Kaltenbrunner, T. Bovermann, R. Bencina, and E. Costanza, “Tuio: A protocol for table-top tangible user interfaces,” Proceedings of the 6th International Workshop on Gesture in Human-Computer Interaction and Simulation (GW 2005), Vannes (France), 2005.

[25] Osama Masoud and Nikolaos P. Papanikolopoulos, “A Novel Method for Tracking and Counting Pedestrians in Real-Time Using a Single Camera” IEEE Transactions on Vehicular Technology, 2001.