Figure 5: Experiment II - Interaction between two participants and two volunteers.

much more recognition attempts have been made in the *M-Asyn* mode (mean = 7.33; STD = 2.16) than in the *S-Syn* model (mean = 4.23; STD = 1.05). Similar results can be found for the poor lighting condition, whereas the gap between the two conditions is even more significant. Apparently, this is a result of opportunistic multi-tasking where the server processes multiple images using multiple CPUs. This should not be interpreted as less efficient because the processing happens in parallel along different threads. Thus, the overall processing time is not increased despite the additional computational load caused by more images. In fact, this is the main factor that contributes to the higher accuracy and less recognition time of the *M-Asyn* mode. By enjoying a larger pool of input images, it increases the chance of capturing an image of better quality amid poor ones. This is particular relevant to a dynamic interaction process in inferior lighting conditions.

Experiment II – Field Test

While Experiment I shows quantitative evidence of the performance advantage of the multi-threaded asynchronous structure, a more important issue is how such performance influences user experience in a practical interaction scenario. In particular, we want to find out what is the desirable performance level in terms of speed and accuracy of face recognition, and if our system can achieve such performance needs.

To do so, we conducted a field test to evaluate the application in a real interaction scenario in the workplace. Twenty subjects participated in the study (7 female; Mean age: 24.6 yrs and standard deviation (STD): 5.2 yrs). In an experiment session, participants wore Google Glass and met with two persons (randomly chosen from the 15 volunteers in the face dataset). The Glass performed face recognition, which is used as a trigger to retrieve the personal information, i.e., upon successful recognition it retrieved the biographical information of the respective persons from the database and showed it on the Glass's prism display. Each experiment session lasted about 15 minutes. The main purpose of the meeting is for the participants to get to know the volunteers in a simulated context where they just started working in a new company (Figure 5). After the experiment session, we interviewed the subjects about their attitudes towards the device. In particular, we reviewed the instances of face recognition and retrieved the time taken for successful face recognition. In case the system fails to recognize a person or retrieves the wrong person's information, a failure is reported. Meanwhile, we collected subjects' perception of the speed of face recognition using a few questionnaire items (e.g. "I think the system is fast in recognizing the person") based on a 7-point Likert scale. (1: not at all, 7: very much so). Note that other aspects of user experience were investigated in the experiment, whereas they are not reported in this paper.

The current system achieves an accuracy of 92% (i.e., 3 failures out of 38 instances), and the speed of recognition range from 1 to 5 seconds in the given context (mean: 2.7s, STD: 0.9s). The statistics shows larger variation as compared to that in Experiment I, mainly due to the uncertainties in a real interaction context. The mean response time is 2.7 seconds (std=0.94s) and the mean acceptance level is 3.4 (std=1.66). Subjects' acceptance level is plotted in Figure 6. There is a rough trend of acceptance decline with an increase of response time. Fitting the data points using a linear trend-line and intersecting it with an acceptance level of '4' (corresponding to *moderate level of acceptance*), an acceptable speed of face recognition is estimated to be 2.3 seconds. Apparently, there is still a sizable portion of instances where the time taken for a successful recognition is above this threshold. Nevertheless, our test happened in an unconstrained environment with varying illumination conditions and motion blur. As such, it is close to supporting practical applications in the wild, which has been barely tackled in legacy systems.

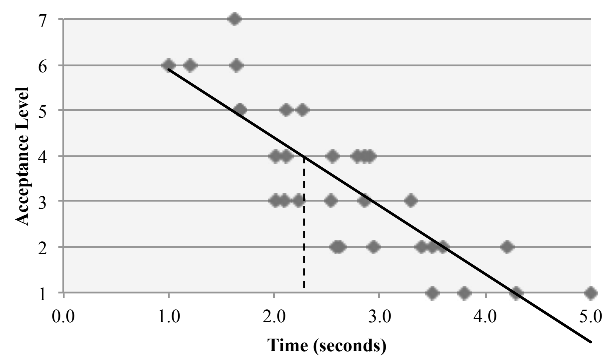

Figure 6: Data plot showing the relationship between user acceptance level and response time of face recognition. Each data point denotes an instance of face recognition. The solid line is a linear fitting of the trend. Acceptance Level '4' is considered to be a moderate level of acceptance.

A notable detrimental factor of user attitude is system failure. It is observed that once a false recognition happened, users' trust of the system plummeted. This is true for all three subjects who experienced the error. Meanwhile, the subjects were dissatisfied by the fact that only frontal face was recognizable. Since pointing the camera (on Google Glass) directly toward a person's face is impolite, it is desirable to recognize the face using the side- or oblique- view images. Some subjects suggested that face recognition be done offline, i.e. to capture a few good quality images when opportunities arise, and carry out the recognition at the back-end. This may mitigate the awkwardness of pointing or aligning the camera towards a person for a prolonged face recognition process.

Discussions

The proposed multi-threaded asynchronous structure consistently outperforms the sequential synchronous structure [5]. Therefore, it is a useful strategy to enhance the system performance in such a scenario. Nevertheless, it is worth mentioning that there is a trade of performance gain and computation overhead. First, one may expect higher computational load in the *M-Asyn* mode due to data transfer (via Bluetooth) and multi-threading. This is caused by (1) the overhead for managing and scheduling multiple threads on the server, and (2) the need to process more images in total. Thus, the structure consumes more CPU resources. Accordingly, it may lead to higher energy consumption, albeit this is not tested in the study. Second, there is an apparent increase of data transfer load (L) due to the number of threads (N). Such an increase is partially reflected by the number of recognition attempts (Figure 4). Since we did not conduct systematic study, we do not know the exact relationship between the number of threads and data transfer volume. A reasonable guess is $L \propto N^b$, where b<1. Second, it may cause a depletion of CPU resource if the devices need to process multiple tasks on top of face recognition. Nevertheless, for a highly dynamic task such as social interaction, the performance factor probably outweighs energy efficiency. Third, the structure is tested on a quad-core structure, so that the results may not be generalizable to other structures with varying number of CPU cores. Therefore, it may be beneficial to consider the complexity of the applications and the server hardware before switching to multi-threaded structure.

Based on the second experiment, we observed stringent demands on the performance of face recognition in a social interaction scenario. The multi-thread asynchronous structure can partially achieve the technical demands as shown in our experimental results. However, we acknowledge that performance alone may not cover all users' concerns. An in-depth study of user needs and attitudes is required [8].

Conclusion

In this paper, we present a multi-threaded asynchronous structure that has been implemented for face recognition application in wearable devices. The experimental results show that this structure improves the application's performance by using multi-threading to process more images and return the recognition results to the client in an asynchronous manner. Although the structure is implemented on a face recognition application, we believe it can be applied to many other similar applications. Moreover, as mobile devices become more powerful, this structure has the potential to reap significant performance improvements with minimal code changes.

Acknowledgements

The work is funded by the Singapore A*STAR JCO VIP – Revers-Engineering Visual Intelligence for Cognitive Enhancement (Project No. 1335H00098).

References

1. Anam, A., Alam, S. and Yeasin, M.. Expression: A dyadic conversation aid using google glass for people with visual impairments. in *Proc. UbiComp'14*, (2014), 211-214.
2. Flautner, K., Uhlig, R., Reinhardt, S. and Mudge, T. Thread-level parallelism and interactive performance of desktop applications, in *Proc. ASPLOS-IX*, (2000),129-138.
3. Ha, K., Chen, Z., Hu, W. Richter, W., Pillai, P. and Satyanarayanan, M.. Towards wearable cognitive assistance, in *Proc. MobiSys'14*, (2014), 68-81.
4. Hwang, A.D. and Peli, E., Augmented edge enhancement for vision impairment using google glass, *SID Symposium Digest of Technical Papers*, 45, (2014), 305-307.
5. Mandal. B., Chia, S.C., Li, L., Chandrasekhar, V., Tan, C. and Lim, J.H., A wearable face recognition system on google glass for assisting social interactions, in *Proc. ACCV'14*, (2014).
6. Wang, X., Zhao, X., Prakash, V., Shi, W. and Gnawali, O., Computerized-eyewear based face recognition system for improving social lives of prosopagnosics, in *Proc. PervasiveHealth'13*, (2013),77-80.
7. Wasik, B. Why wearable tech will be as big as the smartphone, *Wired*, http://www.wired.com/2013/12/wearable-computers/ (last visited 19/06/2015)
8. Xu, Q. Mukawa, M., Li, L., Lim, J-H, Tan, T., Chia, S.C., Gan, T. and Mandal, B., Exploring users' attitudes towards social interaction assistance on google glass, in *Proc. AH'15*, (2015), 9-12.