Asian Institute of Technology

Report of External Examiner
for the Degree of Doctor

Name of Student : Mr. Kan Ouivirach [107830, CS]

Candidate for the Degree of : Doctor of Philosophy

Academic Unit : School of Engineering and Technology

Title of Dissertation:

Incremental Behavior Modeling and Suspicious Activity Detection

The assessment of the External Examiner of the above dissertation is that the dissertation is:

- [ ] satisfactory and meets the normal requirements for a doctoral dissertation (comments attached);
- [x] acceptable after revision according to comments attached;
  - [ ] I would like to see the response of the student on my comments;
  - [x] I request the Program Committee to ensure that my comments on the dissertation are addressed by the student. It is not necessary to send the response to me.
- [ ] to be re-submitted for another review after the attached comments have been answered;
- [ ] not acceptable for reasons attached.

Please attach detailed evaluation.

Prof. James J. Clark
Name

Signature
Date April 25 2013

Dr. Matthew Dailey
Program Committee Chair

Date

Ph.D. Thesis Report

For Kan Ouivirach
Asian Institute of Technology

I found this thesis to be well written and easy to read. I found the technical work rather pedestrian (no pun intended), however. As a development of a well-performing anomaly detection system for single moving objects, the thesis work is acceptable, but it does not provide much in the way of an advance in the state-of-the art. The thesis does a fair job of relating previous work in areas directly related to the proposed approaches, but omits mention of current state-of-the-art research in the area of human activity recognition and detection. Finally, the thesis is deficient in the experimental validation of the proposed approaches. Only a single, rather simple, experimental situation was considered. The generality and robustness of the proposed approach cannot be judged with such a limited range of test scenarios. The important case of multiple objects was not at all considered, not even at a hypothetical level. A discussion of the failure modes of the proposed approach was not provided. This is extremely important as it suggests how the method can be improved, and for what range of applications is appropriate. The experimental scenarios should accordingly include those in which the system will fail.

The thesis was rather short as Ph.D. theses go, and it appears that this is due to a lack of detail in the description of the various algorithms, the aforementioned sparsity of the experimental scenarios, and a lack of informed and insightful discussion of the implications of the experimental results. The discussion sections of each chapter are basically just brief summaries of the results.

The following are some detailed comments regarding the thesis text.

- The statement of the three main challenges for human behavior understanding, on page 2, is too strong. Point 1 states that "It is impossible to store all data in a system'". It is *not* impossible, although it may be impractical. Point 2 states that "Real human behavior is always ambiguous;". This is not the case, human behavior is sometimes unambiguous. Point 3 states that "Unusual behavior is rare and diverse; therefore we cannot learn a model for it.". This is not true either, since we can always learn a model, it just might not be a very good model. Also, it is possible, but probably impractical, to acquire sufficient data for a good model even for infrequent events.

- There should be a more in-depth review of the state-of-the-art in surveillance systems. By restricting the review to commercial and open-source systems the thesis misses out on the most advanced techniques. Commercial and open-source systems always lag behind the state-of-the-art by many years. For examples of such, see the special issue of Machine Vision and Applications, August 2007, Volume 18, Issue 3-4, on "Novel concepts and challenges for the next generation of video surveillance systems". There should also be a review of recent work in human activity analysis, a currently hot area in machine vision research (e.g. see the paper Turaga, P.; Chellappa, R.; Subrahmanian, V. S.; Udrea, O., "Machine Recognition of Human Activities: A Survey," *Circuits and Systems for Video Technology, IEEE Transactions on* , vol.18, no.11, pp.1473,1488, Nov. 2008). An example of recent high-quality work in this area is "Weilong Yang; Yang Wang; Mori, G., "Recognizing human actions from still images with latent poses," *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on* , vol., no., pp.2030,2037, 13-18 June 2010".

- In section 2.3 a simple changed pixel measure is used to detect motion in an image. No details are given, however, such as the choice of the intensity difference threshold value. Such an approach can fail in common surveillance applications where the camera is subject to small vibration. These vibrations cause small shifts of the entire image, which can be sufficient to make a large number of pixels change their values significantly.

- On page 19 the formulae for the dx and dy should be given. If these are normalized (since they are referred to as a unit vector) then it is improper to label them as dx and dy (since that implies an incremental vector, not a unit vector). Also, if they define a unit vector, then it is redundant to use both of them in the feature vector. You should instead use a single angle parameter.

- It is important to say how you improved upon the method of Gharti (2010) for blob-tracking. The reader should know exactly what your contribution was. Just saying that you improved the code is not enough. You have to say what exactly you did.

- You say that you did not use Kalman filters or other approaches (presumably particle-filter methods) in the blob tracking since your data has little noise and the results are being coarsely quantized. However, the difficulties in tracking that are handled by particle-filter type methods are not those caused by noise, but those caused by multiple moving objects, with the attendant occlusions and ambiguity. Presumably these issues are present in your application as well (as you later note in section 2.7).

- You should define all acronyms used in the thesis at their first appearance in the text (for example, HMM should be defined at the start of section 2.6). It is good practice to include a table of acronyms and symbols in the front matter of the thesis.

- On page 22 it is stated that prior to clustering you normalize each feature by a z-score. What is a z-score in this context? Is it based on a statistical model of the feature? Details should be given here, at least provide an equation showing how the z-score is computed and the normalization done.

- The way in which equation 3.1 is written makes it unclear that (as I assume) the three likelihoods P(H_diff), P(S_diff), and P(V_diff) are multiplied together. Perhaps put in a multiplication symbol (* or X). Also, it is improper to call this the “measurement likelihood”, as it is actually the *probability* of measurement given the classification A, whereas it is the *likelihood* of the classification given the measurement, i.e. L(Axy=sh | Mxy) = P(Mxy | Axy=sh).

- How are the Gaussian model parameters estimated (in section 3.2.1)? The distributions for the true object pixels are clearly non-Gaussian so there are many ways that a Gaussian could be fit to these. What approach did you use? Did you just use the sample means and variances? A minimum variance estimator might be better.

- In section 4.2.2 it is said that morphological operations are used. More detail should be given for these. In general, enough detail should be given so as to allow someone to fully implement the proposed approach and replicate the experiments done.

- Section 4.2.2 states that the evaluation is limited to a single moving blob. It is said that this is for simplicity, but it also prevents the approach from being evaluated on the more usual and interesting problem of multiple targets. Saying that the restriction is for simplicity implies that it is not imposed due to the inability of the proposed approach to handle the multiple target case. If the proposed approach does in fact work for multiple targets this should be stated and verified, as it would be a strong advantage of the approach. If the approach does not work in this case, then it is misleading to say that the single blob case was used for simplicity, when in fact it is being imposed due to the approach failing. In either case the proposed approach should be tested on multiple targets to see how it performs and where it might fail.

- How is the empirical "tuning" described in section 4.2.2 actually done? Details are needed!

- The equation on page 36 (which should have an equation number!) defines a (log) likelihood, but the text incorrectly refers to this as the likelihood of the sequence O_i. This indicates a misunderstanding of the meaning of the term –"likelihood". The quantity L_i is the log of the *probability* of the sequence O_i given the HMM M_c, and is also the log of the *likelihood* of the HMM M_c given the sequence O_i. Note the difference!

- Figures 4.4, 5.2 and 6.1 are duplicates of each other. Only one need be given. These figures (or figure) should show one or more examples of the anomalous behaviors. The text should also describe the nature of typical anomalous behaviors that were observed (e.g. a person standing still or dancing, or stopping halfway and turning back in the direction they came).

- In table 4.1 it is evident that the clusters do not separate the Walking In and Cycling In behaviors. This implies that the HMMs are not discriminative enough and perhaps could benefit from additional training examples. Of course, for the purposes of deciding on usual versus anomalous this is acceptable, but that leads to the conclusion that a simpler training of HMMs would work just as well.

- It is misleading to claim that the separation of anomalous and typical behaviors in Table 1 is 100%. It is not, since there are still unseparated behaviors in the single sequence clusters. A manual classification of these single-sequence clusters could result in the anomalous behaviors being classified as one of the typical behaviors.

- The conclusion stated in section 4.3.3 that "our method achieves perfect separation of anomalous and typical behaviors" is completely unwarranted. All that you have done is demonstrated this perfect separation in a single restricted dataset. You would need to repeat this experiment over a very much larger database of examples before anyone would take this conclusion seriously. The limited experimentation is a major weakness of the thesis. Nothing can really be concluded from it.

- Is the θ_z of Algorithm 2 the same as the quantity p_c defined in section 5.2.1?

- Is the testbed described in section 5.4 the same as that used in chapter 4?

- The graphs in figure 5.3 should be enlarged. Perhaps array these as 2 columns by 3 rows, and fill an entire page. Otherwise it is difficult to read the legend and labels of the graphs. The examination copy that I was given did not have any color, so the captions should be altered to remove any reference to color. Markers could be added to the lines on the graph to help distinguish them. The axes need to be labeled!

- Table 5.1 does not show "perfect accuracy" as stated in section 5.4.1, since, as in chapter 4, the one-sequence clusters do not separate the anomalous from typical behaviors. This has to be done manually, and it is not demonstrated that manual separation can be done perfectly.

- The version of the thesis I was given did not have color, so the different curves in figure 5.4 could not be distinguished. These should have different line types, or have distinguishing markers added.

- On page 50 it is stated that the ROC (figure 5.4) reveals that a threshold of -3.259 achieves zero false negatives. However, there are no indications of the threshold values on figure 5.4 so I do not see how this threshold value of -3.259 can be obtained by looking at figure 5.4.

- The acronyms (e.g. TP, FP, TN, etc.) used in Table 5.2 should be defined in the table caption.

- How do the results of chapter 5 vary with changes in the size of the bootstrap set? Is there an optimal size?

Professor James J. Clark

Department of Electrical and Computer Engineering

McGill University

Montreal, Canada

April 24, 2013