Final Report for Period: 03/2006 - 02/2007 **Submitted on:** 03/30/2007
Principal Investigator: Rehg, James . **Award ID:** 0133779
Organization: GA Tech Res Corp - GIT
Title:
CAREER: Motion Capture from Movies: Video-Based Tracking and Modeling of Human Motion

Project Participants

Senior Personnel

Name: Rehg, James
Worked for more than 160 Hours: Yes
Contribution to Project:

Post-doc

Graduate Student

Name: Jie, Sun
Worked for more than 160 Hours: Yes
Contribution to Project:

Name: Wang, Ping
Worked for more than 160 Hours: Yes
Contribution to Project:

Name: Wu, Jianxin
Worked for more than 160 Hours: Yes
Contribution to Project:

Name: Brubaker, Spencer
Worked for more than 160 Hours: Yes
Contribution to Project:

Name: Kim, DongShin
Worked for more than 160 Hours: Yes
Contribution to Project:

Name: Patrick, Hugh
Worked for more than 160 Hours: Yes
Contribution to Project:
H. Alton Patrick was a Masters student who worked on this project, receiving his Masters degree in 2004. He developed a software codebase for figure tracking algorithms and explored the use of particle filter methods for figure tracking. His code base became the basis for our work on comparing particle filter algorithms and is freely available through our project website. In addition to the support provided by this grant, Mr. Patrick was also supported by an NSF Graduate Fellowship.

Undergraduate Student

Name: Banerjee, Sujoy
Worked for more than 160 Hours: Yes
Contribution to Project:

Name: Kalra, Ankur
Worked for more than 160 Hours: Yes
Contribution to Project:

Name: Kumar, Neeraj
Worked for more than 160 Hours: Yes
Contribution to Project:
Neeraj Kumar did an undergraduate research project with me which focused on the development of methods for generating training sets for limb detection in video. He developed a technique for using a software package known as Poser to produce realistic positive training examples that could supplement video training data.

Technician, Programmer

Other Participant

Research Experience for Undergraduates

Organizational Partners

Intel Research Pittsburgh - Intel Corp
We worked with Intel Pittsburgh to develop several novel applications of our cascade classifier techniques to medical image analysis. This work began under the CAREER award and is the basis for an on-going partnership.

Other Collaborators or Contacts

Dr. Jessica Hodgins, School of CS, Carnegie Mellon University
Dr. Alexai Efros, School of CS, Carnegie Mellon University
Dr. Frank Dellaert, College of Computing, Georgia Tech

Contacts in rare event detection:
Dr. Wenke Lee, College of Computing, Georgia Tech

Contacts in insect behavior modeling:
Dr. Tucker Balch, College of Computing, Georgia Tech

Contacts in application of tracking and analysis to activity understanding
Dr. Gregory Abowd, College of Computing, Georgia Tech
Dr. Tanzeem Choudhury, Intel Research Seattle
Dr. Matthai Philippi, Intel Research Seattle

Activities and Findings

Research and Education Activities: (See PDF version submitted by PI at the end of the report)

Findings: (See PDF version submitted by PI at the end of the report)

Training and Development:
This grant supported four PhD students, two Masters students, and

three undergraduates. Of the four supported PhD students, Jie Sun, Ping Wang, Jianxin Wu, and Charlie Brubaker, the first three are all doing their thesis proposals this year and will graduate in an additional year. This group includes one female PhD student. The two Masters students, H. Alton Patrick and Dong Shin Kim successfully defended their Masters theses. Mr. Patrick is currently working in the David Sarnoff Labs in Princeton NJ and Mr. Kim is a PhD student at Georgia Tech working in vision and robotics. All students were provided advanced training in computer vision and machine learning through mentoring and interaction with the PI and other students. The students learned the theoretical and experimental skills which are necessary to solve real-world detection and tracking problems.

The three undergraduate researchers on this grant, Ankur Kalra, Neeraj Kumar, and Sujoy Banerjee, all graduated successfully. Mr. Kalra is currently a PhD student in Computer Science at the Univ. of Illinois at Urbana-Champaign (UIUC) and Mr. Kumar is a PhD student in Computer Science at Columbia University.

Outreach Activities:

During the course of this grant, the PI began an on-going collaboration with the computer graphics group at CMU as a consequence of his work on SLDS models. The PI served on the PhD thesis committee of one of his co-authors and is collaborating with Drs. Hodgins and Efros on several new publications as well as looking for additional funding to develop the work in new directions.

In addition, our work on bee dance modeling has led to new connections between biologists studying insect behavior and computer scientists. Our work directly supports a collaboration which is aimed at compiling statistical evidence to support a novel biological theory of the bee dance.

Journal Publications

J. M. Rehg, D. D. Morris, T. Kanade, "Ambiguities in Visual Tracking of Articulated Objects Using 2-D and 3-D Models", Intl. Journal of Robotics Research, p. 393, vol. 22(6), (2003). Published

J. Wu, J.M. Rehg, and M.D. Mullin, "Learning a Rare Event Detection Cascade by Direct Feature Selection", Proceedings, Conference on Advances in Neural Information Processing Systems 16 (NIPS*2003) by MIT Press, p. 1, vol. 1, (2004). Published

J. Sun, A. Bobick, J.M. Rehg, "Automatic Cascade Training with Perturbation Bias", Proceedings IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2004), Washington DC, p. 276, vol. 2, (2004). Published

J. Wu, M. D. Mullin, J. M. Rehg, "Linear Asymmetric Classifier for Face Detection", Proceedings, Intl. Conf. on Machine Learning (ICML 05), Graz, Austria, p. 993, vol. 1, (2005). Published

S. Oh, A. Ranganathan, J. M. Rehg, and F. Dellaert, "A Variational Inference Method for Switching Linear Dynamic Systems", GVU Technical Report GIT-GVU-05-16, Georgia Institute of Technology, p. 1, vol. 1, (2005). Published

L. Ren, A. Patrick, A. Efros, J. Hodgins and J. M. Rehg, "A Data-Driven Approach to Quantifying Natural Human Motion", ACM Transactions on Graphics, Special Issue: Proceedings SIGGRAPH 2005, p. 1090, vol. 24(3), (2005). Published

S. M. Oh, J. M. Rehg, T. Balch and F. Dellaert, "Learning and Inference in Parametric Switching Linear Dynamic System", Proceedings, International Conference on Computer Vision (ICCV 05), Beijing, China, p. 1161, vol. 2, (2005). Published

S. M. Oh, J. M. Rehg, T. Balch, and F. Dellaert, "Data-Driven MCMC for Learning and Inference in Switching Linear Dynamic Systems", Proceedings, Twentieth National Conferecn on Artificial Intelligence (AAI 05), Pittsburgh, PA, p. 944, vol. 1, (2005). Published

S. Charles Brubaker, Matthew D. Mullin, and James M. Rehg, "Towards optimal training of cascaded detectors", Proceedings, European Conf. on Computer Vision (ECCV), Graz, Austria, p. 325, vol. 1, (2006). Published

S. Charles Brubaker, JianxinWu, Jie Sun,Matthew D.Mullin, and JamesM. Rehg, "On the Design of Cascades of Boosted Ensembles for Face Detection", International Journal of Computer Vision, p. , vol. 1, (2007). Accepted

Sang Min Oh, James M. Rehg, Tucker R. Balch, and Frank Dellaert, "Learning and Inferring Motion Patterns Using Parametric Segmental Switching Linear Dynamic Systems", International Journal of Computer Vision, p. , vol. , (2007). Accepted

Sang Min Oh, James M. Rehg, and Frank Dellaert, "Parameterized duration modeling for switching linear dynamic systems", Proceedings, IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), New York, NY, p. 1694, vol. 2, (2006). Published

PingWang and James M. Rehg, "A Modular Approach to the Analysis and Evaluation of Particle Filters for Figure Tracking", Proceedings, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), New York, NY, p. 790, vol. 1, (2006). Published

Jianxin Wu, S. Charles Brubaker, Matthew D. Mullin, and James M. Rehg, "Fast Asymmetric Learning for Cascade Face Detection", IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI), p. , vol. , (2007). Accepted

Books or Other One-time Publications

H.A. Patrick, "Figure Tracking with Switching Linear System Models", (2004). Thesis, Published
Collection: Masters Thesis
Bibliography: College of Computing, Georgia Institute of Technology

Web/Internet Site

URL(s):

http://www.cc.gatech.edu/cpl/figuretracking/

Description:

Complete data sets, testing videos, software, and scoring methodologies for comparing figure tracking methods.

Other Specific Products

Contributions

Contributions within Discipline:

Video remains an open frontier for computer vision. If vision techniques could extract meaningful information from general video data, then the nature of multimedia data would change fundamentally. With the exception of success in limited contexts such as face detection, however, vision results on general video data are scarce. We believe that figure tracking is the next frontier, in the sense that it will soon be possible to reliably track figures in general video. If we are successful, then it would be possible to reliably analyze people in video for the first time. Since people are the single most interesting category of video content, this would be a significant enabling technology for a wide range of applications.

Our work in the project has quantified the state of the art in figure tracking methods through our comparative study of particle filter algorithms. Moreover, our work on cascade classifiers has set the stage for a detection-based approach to tracking. Our work on SLDS models has paved the way for the development of a general-purpose motion prior for tracking applications.

Contributions to Other Disciplines:

The process of searching for faces and figures in video is an instance of a general problem of rare event detection. From the standpoint of the bit-rate of the video stream, any single object in a video frame is a rare event. We believe that the technology we are developing for rare event detection will be applicable in many other domains besides video analysis. In fact, as data sizes grow, it will increasingly be the case that any

event of interest to a user at a particular time will be rare. We are currently exploring connections between rare event detection and intrusion detection problems with our collaborator Wenke Lee.

Our collaboration with Jessica Hodgins at CMU resulted in the application of our SLDS modeling techniques to a novel application in computer animation. The SIGGRAPH paper that results has brought these models to the attention of the animation community and has helped to popularize the use of data-driven techniques in computer graphics and animation research.

Contributions to Human Resource Development:

This project contributed to the training of four graduate students, two masters students, and three undergradate students. This list includes a female PhD student. All students received extensive training in computer vision and machine learning techniques. In addition, the courses that were developed and taught as a consequence of this grant have contributed to the development of a population of graduate students who are well-versed in these emerging research techniques.

Contributions to Resources for Research and Education:

This work resulted in a database and testing methodology for comparing figure tracking methods and associated particle filter algorithms. All of the data, software, testing video, and scoring methods are freely-available for download from the project web site. This will be a resource for on-going research in figure tracking within the computer vision community.

Contributions Beyond Science and Engineering:

If our technology enables more effective access to multimedia content, then this could have broad impact on the public welfare. Something like 'video google' could have a major impact on the consumption and development of multimedia content.

Categories for which nothing is reported:

Any Product

November 30, 2004

Re: Continuation – CAREER Proposal James Rehg

I have read, and I continue to endorse, James Rehg's career development plan. I attest that the PI's career-development plan is supported by and integrated into the educational and research goals of the College of Computing and the Georgia Institute of Technology. I personally commit to the support and professional development of the Principal Investigator.

Sincerely,

Rich DeMillo
Dean

/jcm

NSF CAREER 0133779: Final Report on Research Findings

James M. Rehg, PI

1 Review of Project Goals

Our findings from this project can be organized into three major topic areas:

- Comparative analysis and development of particle filter algorithms for figure tracking
- Development of fast and effective cascade classifier architectures for figure detection
- Improved learning methods for Switching Linear Dynamic System models and their application to motion analysis

In the sections that follow we present the in-depth findings from each project area.

2 Particle Filter Algorithms for Figure Tracking

There were three types of findings from the effort on systematically comparing particle filter algorithms for figure tracking:

- Categorization of particle filter algorithms
- Assessment of the relative benefits of algorithm components
- Novel particle filter algorithm based on a combination of components

We identified two basic strategies to improve particle filter (PF) performance. These strategies partition the space of algorithms that we examined:

- *Increase sample importance.* This category includes three basic algorithm types: Unscented PF (UPF), Optimized PF (OPF), and Optimized Unscented PF (OUPF).
- *Increase sample diversity.* This category includes Regularized PF (RPF), Resample Move (RM), and Adaptive Resampling (ADR).

The OUPF is a novel PF algorithm which we invented during the course of our research. It is based on the observation that two existing PF methods, OPF and UPF, could be combined in a beneficial manner. The construction of OUPF illustrates the benefit of our modular approach to PF algorithm analysis, which makes it easy to identify the essential elements in existing PF algorithms and combine them.

We introduced the notion of *computationally equivalent* sample sets to address the fact that different algorithms do varying amounts of work per sample. All of our experiments were based on using computationally equivalent sample sets for each method, with the result that each method uses the same amount of computational resources in analyzing the testing data. Our experiments demonstrated an ordering in the performance of the importance-based algorithms as follows: OUPF $>$ OPF $>$ UPF $>$ PF. Methods that address sample impoverishment can be

combined with the importance methods. We examined the combinations: OUPF+RM, OUPF+ADR, OPF+RM, and OPF+ADR. We can eliminate UPF and the basic PF in combination with ADR or RM on the basis of the relatively poor performance of these methods. We did not consider RPF because it's computational cost was far too high relative to the benefit it provided, therefore it was not competitive. Among the competitive methods we found that OUPF was superior to all combinations of ADR and RM with OPF and OUPF. We conclude that the strategy of increasing sample diversity is less effective in general than the strategy of increasing sample importance. Strategies like ADR and RM do improve the performance of any PF algorithm they are combined with, but the benefit is not competitive as a result of the computational cost of these methods.

In summary, we have developed a modular strategy for analyzing and evaluating particle filter algorithms and demonstrated that a novel algorithm OUPF out-performs all competing methods. These results appeared in [10], and a journal version is in preparation [11]. All of the software, testing video sequences, and raw data from our work is available from [9].

3 Cascade Classifier Learning for Figure Detection

We have results in three main areas:

- Efficient Cascade Learning
- Optimal Global Methods for Cascade Learning
- Asymmetric Classifiers

3.1 Efficient Cascade Training

We developed a novel cascade learning algorithm based on Forward Feature Selection (FFS). This work was interesting for two reasons. First, it demonstrated that while ensemble classifiers clearly were an effective choice for face detection, the manner in which the ensemble is formed is not that important. The evidence of this statement is that FFS, which is strictly less powerful than Adaboost as a boosting method, can result in cascades whose overall performance is within a delta of the Viola-Jones results. This was an interesting finding because it came at a time when most of the cascade classifier community was focused on the development of different forms of boosting algorithms to construct the stage ensembles.

The second benefit of our work was that it resulted in an algorithm which was two orders of magnitude faster than the original Viola-Jones method in training classifiers. This was a tremendous practical benefit, and it caused us to look more closely at the reason why the Viola-Jones method was so much slower. We discovered (simultaneously with many others as it turned out) that a similar caching scheme could yield the same level of speed-up for the original Adaboost approach, in comparison to the naive implementation. The paper describing this work [14] was published in NIPS, and is part of a journal submission [13].

3.2 Optimal Cascade Learning

The appeal of cascade classifiers is that they are both fast and accurate. Moreover, the speed of the cascade classifier is a property of the cascade architecture, and not merely the result of engineering a traditional pattern recognition method to run fast. However the original Viola-Jones algorithm does not explicitly address the challenge of automatically designing a cascade which is both fast and accurate, and in fact the performance of the best cascade classifiers is the result of manually tuning the number of features in the early stages of the cascade so as to achieve both high speed and accuracy.

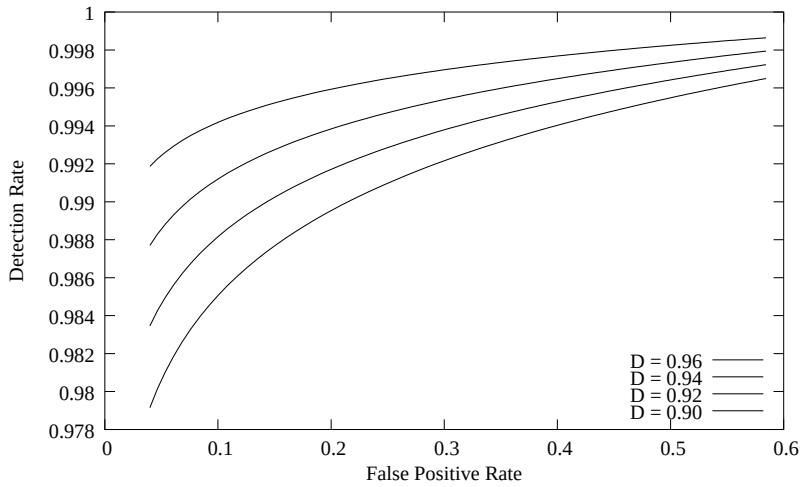


Figure 1: Indifference curves can be identified by their detection rate D at a fixed false positive rate F_g. Each curve is a plot of the equations $d = D^a$ and $f = F_g^a$ for $a > 0$.

In our work we formulated cascade training as a constrained optimization problem in which the goal is to achieve a certain goal false positive rate (F_g) and detection rate (D_g), while minimizing the computational cost of the classifier. We train cascade stages in sequence beginning with the first stage and going until the end. The key decisions in meeting the global performance objective with minimum computation are to decide in each stage how many features to add to the ensemble and how to tune the ensemble performance. Our work on this problem has resulted in a number of important findings:

- We showed that implicit in all of the cascade training algorithms is an assumption about the ability of the classifier to achieve a certain ROC operating point in future stages, given the ability to achieve the operating point at a current stage. We call this the *repeatability assumption*.

- We showed that under the repeatability assumption there is a family of curves in the ROC space, which we call *cascade indifference curves*, that link local decisions about stage performance to global predictions of cascade performance. These curves are illustrated in Figure 1. Every performance point for a given stage lies on a unique indifference curve, and the target point for the cascade lies on a cascade indifference curve. The question of whether an operating point for a particular stage classifier is admissible or not can then be easily reduced to a question about the distance between indifference curves.

- A consequence of the cascade indifference curve is that the space of admissible stage operating points, which we call the *admissible region*, is actually much larger than was considered by the original Viola-Jones learning method. This point is illustrated in Figure 2.

- The original Viola-Jones learning approach proposed to construct a cascade by repeating the same operating point for N stages. Given the expanded admissible region in Figure 2, it is natural to ask whether combining multiple points (e.g. assigning different operating points to different stages) in constructing the cascade could lead to better performance. We show that in many situations it is advantageous to use pairs of points (e.g. a point from Region 3 in Figure 2 in conjunction with a point from Region 2). We prove that using more than two points will not provide any additional benefit.

- We developed two other innovations in cascade training which lead to improved performance. The first is a method we call *recycling* for reusing ensemble scores from previous stages to provide a head-start in

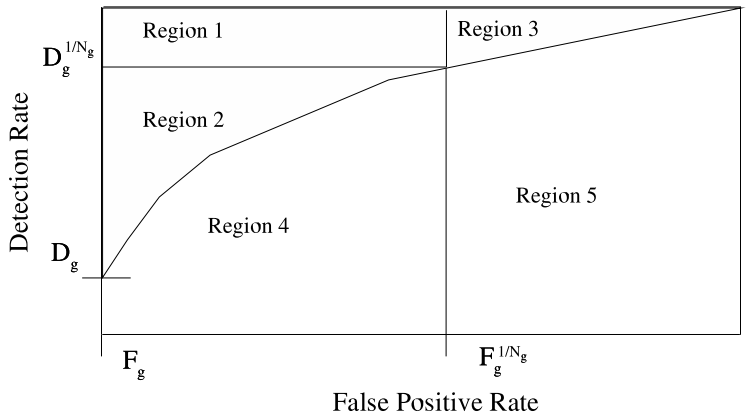


Figure 2: The cascade feasible region defined by the indifference curve consists of regions 1 and 2. The original Viola-Jones approach only considered points in region 1, which is an unnecessary restriction. Points in region 3 are inadmissible under 1-point planning because they cannot achieve the cascade false positive rate in N stages. But under 2-point planning, points in Region 3 can be combined with points from Regions 2 or 4 to obtain improved performance. Points in Region 5 offer no benefit. D_g and F_g represent the goal detection and false positive rates for the cascade.

training a new stage. The second is a method called *retracing* which analyzes previously-trained cascades and inserts additional decision points, leading to improved speed with little impact on accuracy.

- We present a *two-point algorithm* for training cascade stages which optimizes for speed and accuracy. In conjunction with recycling and retracing, this method leads to automatically-trained cascades which are much faster than those produced by the standard Viola-Jones algorithm. This is illustrated in Figure 3. The average number of features/window used by our automatically-trained cascade is 8, which is comparable to the best hand-tuned result of Viola-Jones.

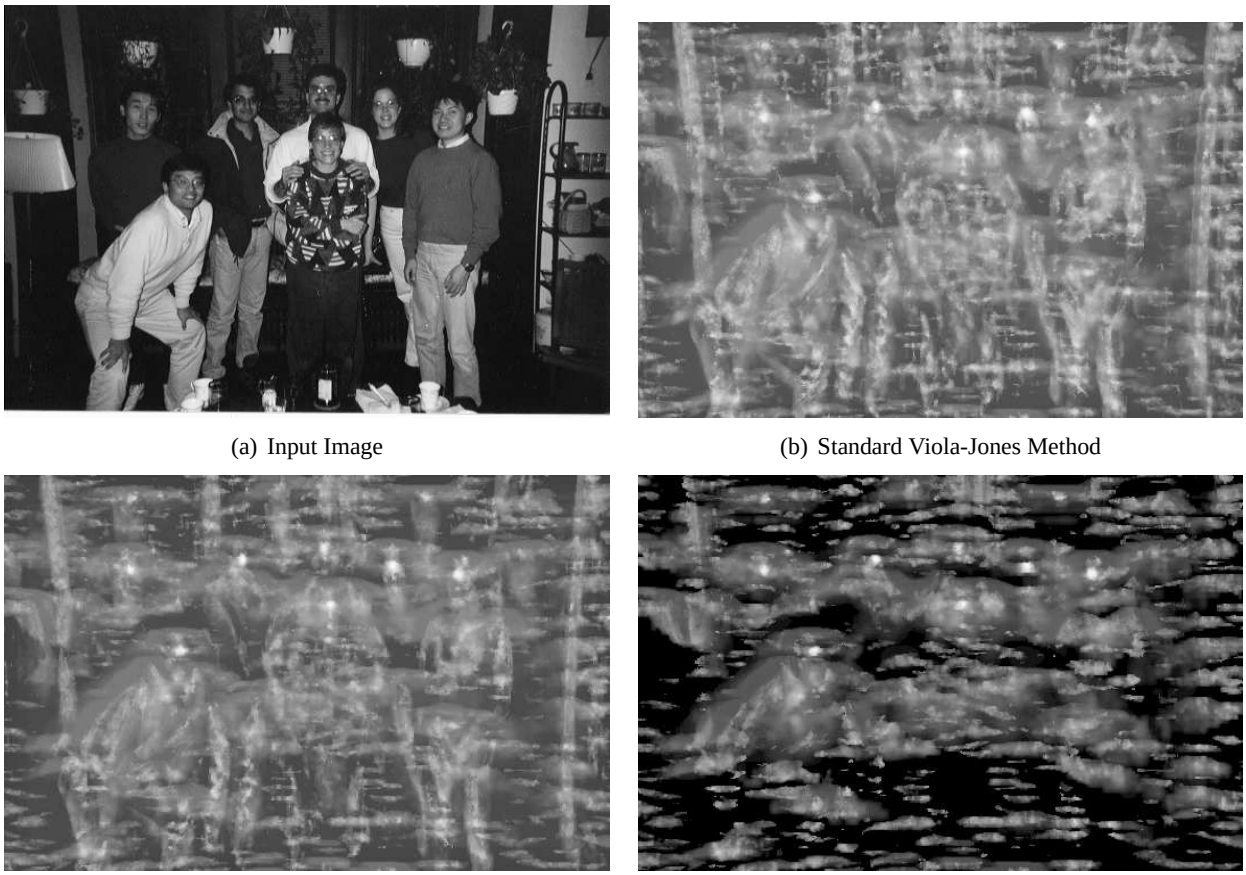
The cascade indifference curve was initially described in [8]. An initial formulation of a global model for cascade training was presented in [1]. The majority of our results on optimal cascade training will appear in [2].

3.3 Asymmetric Classifiers

We have developed a novel algorithm for forming an ensemble classifier from the features which are used in a given cascade stage. The ensemble weights are derived as the solution to a constrained optimization problem of maximizing the detection rate of the stage while achieving a specified false positive rate. This is a natural learning goal for the cascade classifier, as every stage must reject some fraction of the nontarget instances (i.e. must achieve a desired false positive rate) while correctly classifying all of the target instances (i.e. must achieve the maximum detection rate).

We develop an approximate solution to this problem under the case where the false positive rate is 50% (a standard choice in practice for cascades). We show that our method, which we call the *Linear Asymmetric Classifier*, yields ensemble classifiers with better performance than standard AdaBoost and AsymBoost, an asymmetric generalization. The resulting algorithm is related to both Linear Discriminant Analysis and recent machine learning work on Minimax Probability Machine.

An initial version of this work appeared in [12]. The complete exposition is contained in [13].



(a) Input Image (b) Standard Viola-Jones Method

(c) 2-Point Planning (d) 2-Point Planning & Recycling & Retracing

Figure 3: (a) Input image (b)-(c) Computational load images for automatically-trained cascades. The intensity of each pixel is proportional to the log of the computation spent at that location in image (a). Notice that the most intensive computation is reserved for the locations that most resemble faces.

4 Switching Linear Dynamic Systems for Motion Analysis

Our findings in SLDS models for motion analysis consist of two topic areas: bee dance analysis and human motion analysis.

4.1 SLDS Models for Bee Dance Analysis

We developed three major extensions of the basis SLDS modeling paradigm to address issues that arose in bee dance analysis:

- We developed a data-driven Markov Chain Monte Carlo (DD-MCMC) inference procedure which made it possible to obtain greater accuracy in segmenting the bee dance motion into phases, at the cost of increased computation. We demonstrated that DD-MCMC is capable of removing over-segmentation errors that result from simpler SLDS inference methods.
- We added global parameters to the SLDS model that explicitly encode elements of the bee dance such as the angle and duration of the waggle phase. We demonstrated that fitting the global parameters to the data

during inference leads to more accurate estimation of the global parameters and better segmentation of the bee dance into phases.

- We added explicit duration models to the SLDS discrete state to solve the problem of inaccurate state durations, which stem from the geometric duration distribution that is induced by the Markov transition model. We demonstrated that learning the durations from the data led to better segmentation performance.

We tested our new SLDS modeling framework on a set of noisy bee dance tracks and quantified the performance improvements due to our new method. Our DD-MCMC work appears in [3]. SLDS models with global parameters were introduced in [4]. Our results for duration modeling were published in [6]. A comprehensive set of results and experiments will appear in [5].

4.2 SLDS Models for Motion Naturalness

Human motion analysis represents a more significant challenge than bee dances due to the high dimensionality of human movement and the complexity of human motion. Given this complexity and the lack of a systematic characterization of all human movements, we chose to focus on a specific classification task for our first application of SLDS models to human motion. In collaboration with Jessica Hodgins' group at CMU, we addressed the problem of classifying natural human motion using learned motion models.

One aspect of this problem is that while data for natural motions is abundant (the mocap database collected at CMU contains hours of movement data), data for unnatural motions is rare. Unnatural motions are the result of mistakes during motion editing, marker clean-up, and other tasks in computer animation applications. We therefore focused on modeling natural motion data with generative probability models and then used a small corpus of unnatural motion data to tune the threshold on a likelihood-based classifier and validate our method's performance. We found that our SLDS models did particularly well in detecting certain types of unnatural motions such as badly-edited motions and bad motion graph transitions, and less well with unnatural motions that resulted from adding noise to clean motion data. The results of project appeared in [7].

References

[1] S. Charles Brubaker, Matthew D. Mullin, and James M. Rehg. Towards optimal training of cascaded detectors. In *Proc. European Conf. on Computer Vision (ECCV)*, volume 1, pages 325–337, Graz, Austria, 2006.

[2] S. Charles Brubaker, Jianxin Wu, Jie Sun, Matthew D. Mullin, and James M. Rehg. On the design of cascades of boosted ensembles for face detection. *Proc. Intl. Journal on Computer Vision (IJCV), Special Issue on Learning for Vision*, 2007. Accepted for publication.

[3] Sang Min Oh, James M. Rehg, Tucker R. Balch, and Frank Dellaert. Data-driven MCMC for learning and inference in switching linear dynamic systems. In *Proc. Twentieth Natl. Conf. on Artificial Intelligence (AAAI 05)*, pages 944–949, Pittsburgh, PA, July 2005.

[4] Sang Min Oh, James M. Rehg, Tucker R. Balch, and Frank Dellaert. Learning and inference in parametric switching linear dynamical systems. In *Proc. Intl. Conf. on Computer Vision (ICCV)*, pages 1161–1168, Beijing, China, October 2005.

[5] Sang Min Oh, James M. Rehg, Tucker R. Balch, and Frank Dellaert. Learning and inferring motion patterns using parametric segmental switching linear dynamic systems. *Proc. Intl. Journal on Computer Vision (IJCV), Special Issue on Learning for Vision*, 2007. Accepted for publication.

[6] Sang Min Oh, James M. Rehg, and Frank Dellaert. Parameterized duration mmodeling for switching linear dynamic systems. In *Proc. Conf. on Computer Vision and Pattern Recognition (CVPR)*, volume 2, pages 1694–1700, New York, NY, 2006.

[7] Liu Ren, Alton Patrick, Alexei A. Efros, Jessica K. Hodgins, and James M. Rehg. A data-driven approach to quantifying natural human motion. *ACM Transactions on Graphics, Special Issue: Proc. of 2005 SIGGRAPH Conf.*, 24(3):1090–1097, August 2005.

[8] Jie Sun, James M. Rehg, and Aaron Bobick. Automatic cascade training with perturbation bias. In *Proc. Conf. on Computer Vision and Pattern Recognition (CVPR)*, volume 2, pages 276–283, 2004.

[9] Ping Wang and James M. Rehg. Figure tracking project web site. `http://www.cc.gatech.edu/cpl/figuretracking`, 2006.

[10] Ping Wang and James M. Rehg. A modular approach to the analysis and evaluation of particle filters for figure tracking. In *Proc. Conf. on Computer Vision and Pattern Recognition (CVPR)*, volume 1, pages 790–797, New York, NY, June 2006.

[11] Ping Wang and James M. Rehg. A modular approach to the analysis and evaluation of particle filters for figure tracking. *Proc. IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI)*, 2008. In submission.

[12] J. Wu, M. D. Mullin, and J. M. Rehg. Linear asymmetric classifier for cascade detectors. In *Proc. 22nd International Conference on Machine Learning*, pages 993–1000, 2005.

[13] Jianxin Wu, S. Charles Brubaker, Matthew D. Mullin, and James M. Rehg. Fast asymmetric learning for cascade face detection. *Proc. IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI)*, 2007. Accepted for publication.

[14] Jianxin Wu, James M. Rehg, and Matthew D. Mullin. Learning a rare event detection cascade by direct feature selection. In *NIPS 16*, 2004.

NSF CAREER 0133779: Final Report on Research and Education Activities

James M. Rehg, PI

1 Review of Research Goals

The goal of this NSF CAREER project was to develop and analyze a new class of algorithms for figure tracking and explore the application of techniques from figure tracking to other domains such as graphics and animation. After an initial exploratory period of research, we defined three main areas of activity which continued throughout the duration of the project:

- Comparative analysis and development of particle filter algorithms for figure tracking
- Development of fast and effective cascade classifier architectures for figure detection
- Improved learning methods for Switching Linear Dynamic System models and their application to motion analysis

In the sections that follow we motivate each of these topics and provide a detailed description of our research activities in each area.

2 Particle Filter Algorithms for Figure Tracking

Figure tracking is framed as a sequential density estimation problem. Given an input video sequence, frames are analyzed in temporal order and the result of processing each frame is an updated posterior distribution over the possible locations and poses of the figure. Uncertainty in the configuration of the figure comes from two major sources. The first is variability in the appearance of figure due to lighting changes, occlusions, and nonrigid motion. The second is background clutter, resulting in background regions whose appearance is similar to the target. The problem of background clutter is particularly important, because it results in a multi-modal posterior distribution. We address the 2-D localization of the figure in the image plane, but all of the issues that are addressed in our work arise equally in the case of figure tracking in 3-D.

Modern figure tracking methods use particle filters to address the multimodal nature of the posterior pdf. A particle filter represents the posterior pdf as a collection of weighted particles. The filter algorithm propagates the particles and updates their weights, resulting in an incremental on-line tracking algorithm. A major limitation of the particle filter is the fact that the number of particles which is required to obtain an accurate density estimate grows exponentially with the dimension of the state space. As a consequence of the large number of degrees of freedom of the human figure, straight-forward particle filter algorithms would require an intractable number of samples. This problem is compounded by the issue of sample impoverishment, in which samples tend to collapse toward the modes of the posterior distribution over time, resulting an effective sample size which is often much less than the total number of particles.

Existing figure tracking algorithms employ a wide range of modifications to the basic particle filter algorithm in an attempt to address the issues of reducing the necessary sample size and avoiding sample impoverishment.

However, the lack of a standardized testing set for figure tracking algorithms has made it very difficult to understand the relative merits of these different algorithms or to ascertain which modifications of the particle filter are essential for good figure tracking performance. The goal of this research activity was the development of a systematic approach to comparing the performance of particle filtering algorithms, along with a standardized testing set. A goal of this comparative analysis was to identify the components of particle filtering algorithms which are essential for good performance.

With my student Ping Wang, we began this activity by surveying the existing particle filter literature, both in the statistics community and in the computer vision community. We developed a taxonomy for these algorithms which identifies their essential characteristics and algorithmic differences. With this taxonomy in place, we designed an evaluation strategy for comparing particle filter algorithms with three major elements:

- Method for acquiring ground truth data for figure pose and location
- Video data set that covers relevant sources of variability in figure tracking in a controlled manner
- Scoring method for measuring the difference between estimated and ground-truth tracking performance

Our video data came from two sources: archival video footage which was captured prior to our use of it, and footage we captured ourselves in our laboratory. In the case of archival data, we manually generated ground truth data by marking the joint centers of the figure in each image. For our own captured footage, we developed a system which allowed us to jointly capture video and motion capture data in our laboratory setting. We were aided during this part of the project by the efforts of two undergraduate researchers who were funded under this grant: Ankur Kalra and Sujoy Banerjee. The major challenge in building this acquisition system was ensuring that the captured video was time-synchronized with the commercial motion capture system which provides the marker data. We built some special hardware to accomplish the synchronization, with the help of a visiting Korean professor, Dr. Young Ki Ryu (not supported on this grant). With this system we could capture video footage and then use the marker data to obtain reprojected joint centers on each video frame automatically. With these reprojected joint centers we can obtain automatic ground truth for the projected configuration of the arms and legs. This provided the basis for an automatic scoring system which removes the need for human labor in evaluating the accuracy of algorithms on our video footage.

We designed a set of videos for evaluation purposes which covered a variety of subjects and types of motions. This dataset addresses a variety of issues in tracking, including appearance variations in the target, degree of self-occlusion, and angle of the direction of motion relative to the camera pose. We developed two scoring systems, one based on the squared error at the joint centers, and another based on the overlap between the projected and ground-truth limb positions. The overlap metric was designed to address the fact that human perception of tracking performance is based more upon whether the overlay of the tracker model seems to overlap with the figure in the image than with joint angle error.

The final element that was needed to conduct the evaluation experiments was a modular software architecture for figure tracking that makes it possible to change the details of the particle filter algorithm while keeping all other elements of the figure tracker fixed. A Master's student, Hugh Alton Patrick, developed the needed software framework for his Master's thesis. Alton was partially supported by the grant and also by an NSF Graduate Fellowship. We created a web site for the evaluation project which makes the software, data sets, and evaluation protocols publicly-available for use by others in the figure-tracking community.

3 Cascade Classifier Learning for Figure Detection

The multimodality of the posterior distribution of the figure is the key challenge in reliable and accurate figure tracking, and this multimodality is the direct result of impoverished appearance models for the figure. Traditional

appearance modeling has relied on edges or templates, with the result that many structures in the background of the scene are similar in appearance to some part of the figure. My approach to this problem was to turn the appearance modeling problem into a discrimination problem: the goal is to learn discriminative models which are focused on distinguishing the figure from the background scene. From this perspective, the motion pattern of the figure is its most distinctive element. So this part of the project addresses the question of detecting human figures in video based on their motion pattern. In order to cope with the complex DOF's of the figure, we target a subproblem of detecting moving limbs. The idea is that if we can detect the limbs with some reliability, than we can group them to obtain hypotheses about the location of the figure. There are a variety of ways that the fusion of detection events could be accomplished, so the key challenge is the development of a fast and accurate limb detector.

We approached the limb detection problem by building on top of the cascade classifier approach which was popularized by Viola and Jones for the task of face detection. Cascade classifiers are appealing because of their efficient testing performance, which will be critical for a practically successful figure tracking system. There was also evidence from the pedestrian detection work of Viola, Jones, and Snow that suggested that cascade classifiers could be effective for motion pattern detection. As we began to investigate the cascade classifier framework, we discovered that there were several inadequacies in the existing formulation and we began a research effort to explore optimal learning of cascade classifiers. This ended up becoming an important subthread within the project, and we now have a deep understanding of this research area and have obtained results that define the state-of-the-art for this problem.

3.1 Efficient Cascade Training

Our initial work in this area, with my student Jianxin Wu, addressed the computational requirements of the learning algorithm for cascade classifiers. A naive implementation required days of training time to learn a classifier for face detection, a result which meant that the required training time for limb detection would be completely untenable, given that it would involve orders of magnitude more data and concept complexity. At that time, the cascade research community was mostly focused on proposing new boosting algorithms for the ensemble classifiers that are used within the stages of the cascade. We explored a very simple and fast voting classifier based on forward feature selection and found that it gave surprisingly good performance. This approach lent itself to an efficient caching scheme, resulting in orders of magnitude less computational cost. We later discovered that the same caching scheme could be employed with the original Adaboost algorithm as well, resulting in tractable training times.

3.2 Optimal Cascade Learning

Working with my students Jie Sun and Charlie Brubaker, I then focused on the issue of making better decisions during stage classifier training to improve the detection performance. Cascade learning is a greedy sequential algorithm in which stages are trained in sequence, beginning with the initial stage and progressing until the cascade is complete. The key decision in each stage is how many features to add to the ensemble classifier and how to tune the stage performance (in other words, what detection rate and false positive rate to choose at each stage). The original cascade classifier work addressed this problem through a simple decomposition of the target detection rate and false positive rate for the cascade into stage-wise targets. The disadvantage of this approach is that it ignores the fundamental product structure of the cascade itself (e.g. the fact that two stages at a detection rate d and false positive rate f are equivalent to a single stage with performance d^2 and f^2). We introduced the concept of a cascade indifference curve that captured this basic fact and led to a new learning algorithm with improved detection performance. Jie also developed a method to exploit symmetries in detected patterns (the fact that faces have bilateral symmetry) to improve the detection accuracy.

Another issue in training the stage classifiers is the fact that the Adaboost algorithm which is used to form

the ensemble classifiers is optimizing a balanced loss function, which assumes that patterns from both classes have equal representation in the dataset. This is problematic because the cascade is in fact solving a highly imbalanced problem in which faces are extremely rare and nonfaces are plentiful. This mismatch implies that better performance could be obtained if a more appropriate loss function could be employed. Working with my student Jianxin Wu we formulated a constrained optimization problem for the unbalanced case and derived an approximate solution method, which we called the Linear Asymmetric Classifier (LAC). This resulted in improved stage-wise classification performance.

Note that a research scientist in my group named Matthew Mullin was collaborator on much of this cascade classifier work. He was not supported under this grant.

3.3 Limb Detection

Now that we have put the problem of learning cascade classifiers from data on solid footing, we have begun to apply this machinery in earnest to the problem of limb detection in video. An initial study was done early in the life of the grant by a Masters student Dongshin Kim, working with an undergraduate researcher in my lab named Neeraj Kumar. Dongshin obtained some promising initial results in limb detection, but was hampered by the large computational cost of the training problem and the need to use large numbers of features in each stage (resulting in slow classification performance) in order to get reasonable detection results. With our improved framework for cascade learning in place, I believe we are now poised to attack the limb detection problem directly. This bulk of this work began after the CAREER award expired, and is being conducted by a new student Gallagher Pryor who is partially-supported by an NSF Graduate Fellowship. This work is one of several examples of on-going research threads which stemmed from the results of this CAREER award.

4 Switching Linear Dynamic Systems for Motion Analysis

Another major issue in figure tracking algorithms is the question of what model to use for the figure dynamics. Historically, tracking algorithms have tended to use very simple dynamic models such as constant velocity or acceleration. These models do not provide very strong constraints on the possible motion of the figure over time. In principle, more accurate dynamic models could provide a better constraint on the possible movement of the figure, leading to more reliable tracking. Unfortunately it is very difficult to derive an accurate dynamic model that is computationally tractable from first principles (or from the biomechanics literature) that can cover the full complexity of human movement.

One possible approach to obtaining an accurate dynamic model is to learn the dynamics of the figure from a corpus of representative motion data. The existence of large libraries of general-purpose motion capture data have recently made this approach feasible. A basic question is what parametric model form should be used in learning. In past work I have explored a model class known as Switching Linear Dynamic Systems (SLDS) which has several attractive properties. In the course of this grant I developed several extensions of the SLDS modeling framework and applied it to a novel application in animation and computer graphics. From a figure tracking standpoint, there is no point in developing complex dynamic models until the basic elements of the tracking approach are solid. As a consequence, I addressed the SLDS element of this work in other problem domains, in preparation for the time in which the work on particle filtering and limb detection would be ready for the development of an integrated system.

4.1 SLDS Models for the Honeybee Dance

The first opportunity to extend the SLDS models arose in the context of another NSF-funded project on Biotracking, which was led by Prof. Tucker Balch and Prof. Frank Dellaert at Georgia Tech. This project focused on the

use of video tracking technology to record the motion of honeybees under realistic hive settings. The goal of this project was to understand the motion pattern of the bees and quantify the classical honeybee dance which is used to communicate the direction and distance to food sources.

Working with a student Sangmin Oh, who I co-advised with Frank Dellaert, we applied the SLDS modeling framework to the analysis of honeybee dance data. Note that Sangmin was fully-supported by a Samsung Lee Kun Hee Fellowship and did not require support from this grant. We addressed two deficiencies of the standard SLDS framework (from my earlier work) which were limitations in its application to the bee dance analysis. The first limitation was the inflexible tradeoffs between computational cost and accuracy provided by standard approximate inference algorithms for SLDS models. The second limitation was the lack of a principled mechanism for incorporating global parametric variations into the SLDS model.

The SLDS model can be viewed as a hybrid between an standard HMM model and a convention Linear Dynamic System (LDS). As in an HMM, there is a discrete state variable governed by discrete Markov dynamics (a state transition matrix) which evolves in time. In contrast to an HMM, the discrete variable switches between continuous LDS models instead of static mixture densities. Each LDS model has its own continuous linear dynamics which in turn generates the observation sequence. Thus an SLDS describes complex system dynamics by switching within a family of LDS models over time. While the SLDS has greater descriptive power than either an HMM or an LDS, exact inference in an SLDS is intractable. As a consequence, approximate inference techniques such as Viterbi approximations or structured variational inference are required in order to estimate the hidden state variables (the discrete and continuous states) from the measurement sequence.

In earlier work, I developed a variety of approximate inference techniques, but each one represented a fixed way of achieving a certain level of accuracy for a given amount of computational work. In biology applications where the learned models may provide insights into animal behavior, it would be advantageous to spend additional computational resources to obtain a greater level of accuracy. Sangmin, Frank, and I developed a new data-driven MCMC method which makes it possible to flexibly trade computation for increased accuracy. This new method takes the result from a standard approximation method as a starting condition and improves it via sampling. It uses a specially-designed proposal distribution to speed convergence of the MC sampler, and Rao-Blackwellization to ensure that the samples are used efficiently.

The second limitation of SLDS models stemmed from the fact that the parameters in these models which were of interest to biologists were actually global parameters of the dance, such as the angle of the dancer bee relative to the direction of gravity or the duration of the waggle phase of the dance. These parameters were encoded only implicitly, with the result that their estimates were often noisy. We developed an extension of the basic SLDS model which included global parameterizations of both the switching process and the observation process. As a result of this explicit global parameterization, it is possible to estimate the angle and duration parameters jointly with the hidden state, resulting in more accurate estimates and better segmentations of the bee dances into phases.

4.2 SLDS Models for Naturalness Analysis

In parallel with bee dance modeling work, we began a collaboration with Prof. Jessica Hodgins' group at Carnegie Mellon University to explore the use of SLDS models in analyzing the naturalness of human motion. The goal of the project was to develop automatic methods for discriminating between natural human motion and the many unnatural motions that arise by mistake during production work in computer animation. The idea was to leverage a large corpus of motion capture data and learn models that encode natural motion. Given a new motion sequence, a likelihood score can be computed which captures the likelihood that the input sequence could be explained by the natural motion model. Given a few examples of unnatural motion, it is possible to set a threshold on the likelihood score, resulting in a computational test for natural motion. The project investigated a variety of representations for human motion data including HMM models, LDS, and SLDS representations. My Master's student Alton Patrick and I developed a naturalness classifier based on SLDS models which proved to be very competitive with standard

HMM's and other techniques.

5 Education Activities

5.1 Teaching and Curriculum Development

During the course of this grant the PI designed and taught two new graduate level courses:

- *CS 7636 Computational Perception* This course introduced advanced undergraduates and graduate students to the use of computer vision techniques to sense humans and their environments. Large sections of the course were devoted to the detection and tracking problems that were the focus of this project. This course provided a vehicle to disseminate the latest results in this field to interested students.

- *CS 8803 Introduction to Probabilistic Graphical Models* This course provided an overview of current research and methods in probabilistic graphical models, a branch of machine learning that studies graph-based representations and algorithms for probabilistic inference and learning.

Both offered courses met a significant student need, as they addressed popular research topic areas for which there are currently no good textbooks. I taught 7636 three times with a total enrollment of 73 students. I taught 8803 twice with a total of 60 students. The course material in both cases was drawn extensively from my research activities on this grant and I was able to use several of our research results to motivate the need to address various topics in the class. In addition to developing new mathematical skills, my students in these classes have learned experimental techniques, in particular effective approaches to working with large datasets.

I also taught an undergraduate pattern recognition course on two occasions and I was able to incorporate our latest research results on cascade classifiers into my lectures.

5.2 Training of Graduate and Undergraduate Students

This grant supported four PhD students, two Masters students, and three undergraduates. Of the four supported PhD students, Jie Sun, Ping Wang, Jianxin Wu, and Charlie Brubaker, the first three are all doing their thesis proposals this year and will graduate in an additional year. This group includes one female PhD student. The two Masters students, H. Alton Patrick and Dong Shin Kim successfully defended their Masters theses. Mr. Patrick is currently working in the David Sarnoff Labs in Princeton NJ and Mr. Kim is a PhD student at Georgia Tech working in vision and robotics. All students were provided advanced training in computer vision and machine learning through mentoring and interaction with the PI and other students. The students learned the theoretical and experimental skills which are necessary to solve real-world detection and tracking problems.

The three undergraduate researchers on this grant, Ankur Kalra, Neeraj Kumar, and Sujoy Banerjee, all graduated successfully. Mr. Kalra is currently a PhD student in Computer Science at the Univ. of Illinois at Urbana-Champaign (UIUC) and Mr. Kumar is a PhD student in Computer Science at Columbia University.