Classroom Analytics: Measuring Student Engagement with Automated Gaze Tracking

Technical Report · November 2011

DOI: 10.13140/RG.2.1.4865.6242

Jon Bidwell
Georgia Institute of Technology

Classroom Analytics: Measuring Student Engagement with Automated Gaze Tracking

Jonathan Bidwell, Henry Fuchs
University of North Carolina at Chapel Hill
Department of Computer, Chapel Hill, NC
bidwej@unc.edu

Figure 1 ***Multiple color and depth-sensing cameras are used to record lessons in a third grade classroom. These recording are processed to extract student gaze direction over-time and coupled with expert observations to train an automated student engagement classifier.***

Abstract

We present a video recording and behavior analysis framework as a first step towards an automated teacher feedback tool for measuring student engagement. The long-term goal of this work is to help teachers identify student behaviors and behavioral trends that may otherwise be unobserved during classroom activities. Initially experts train a student engagement classifier by making observations during recorded lessons. Face tracking software extracts student gaze targets from recorded footage. Expert observations are paired with sequences of student gaze targets to build a Hidden Markov Model (HMM) based student engagement classifier for each child. This paper details initial experiments and a session recorded in a third grade classroom with three expert evaluators and their minute-by-minute assessment of selected students. The resulting HMM uses expert made observations as a gold standards for classifying student engagement on a second-by-second basis without requiring expert assistance. To the best of our knowledge we are the first to explore automatic gaze-target tracking as an indicator of student engagement in classrooms.

1. Introduction

Expert analysis of recorded footage of has long been used as feedback mechanism for improving human performance. In sports, coaches and players review past games to improve their strategies and skills. In the military, military instructors review recorded footage with trainees following simulated combat exercises. Expert analysis of this kind is often manageable when the duration of recordings and number of cameras are limited; however, hundreds of cameras and many hours of recorded footage can prohibit this kind of comprehensive expert analysis.

Researchers and military instructors at Camp Pendleton and TwentyNine Palms, CA are experimenting with automated behavior analysis software for analyzing hundreds of camera feeds following training exercises. The Behavioral Analysis and Synthesis for Intelligent Training (BASE-IT) program identifies errors that were made during training, such as soldiers grouping too close together or a group of soldiers not scanning a 360-degree area while on patrol [6].

In this paper we take this idea of manual vs. automatic behavior analysis into the classroom as a feedback tool for teachers and educational researchers. Many states require teachers to participate in professional development programs. These professional development programs could be better informed by automated feedback to address the specific learning needs of students in a given classroom. Next, teachers could use this feedback for better designing lessons. “I want to do the best for all of my students” says Donna Rand, 2010 Connecticut Science Teacher of the Year, “[Having] feed-back on the engagement level of every student in my classroom would allow me to plan activities and lessons for helping to enable children understand concepts. I find it difficult to effectively monitor the individual engagement of twenty-five active, students in every part of a lesson.” Teachscape is commercial product that teachers can use for recording classroom lessons [1,5]. The downsides to these types of products are that teachers must review each hour of recoded footage after class; this review process can be time-prohibitive with busy teaching schedules. Here annotating the footage with automated behavioral feedback could reduce this time constraint and in doing so make such tools more practical use on a day-to-day basis. For example teachers could could observe a given student’s behavioral trend at a glance over the course of a month or jump and then jump to a specific section recorded video from two days earlier to recall the context of a particular set of student behavior measurements.

Meanwhile, educational researchers and school psychologists continue to make traditional in-person observations for diagnosing and modifying student behaviors. These observations require training, are expensive and labor intensive, infrequent and often limited in duration to 20-30 minutes. Student observations by school specialists are usually reserved for diagnosing clinical behavioral disorders such as ADHD and autism. Here we could imagine democratizing student behavior observation. Having installed low-cost webcams in classrooms behavioral analysis software could be upgraded over time to support specialized educator and research needs and reach a large audience. Here we present our initial steps in the direction of developing an educator feedback tool for measuring individual student behaviors on a second-by-second basis in an active, real-world classroom.

In this paper we introduce “Classroom Analytics” to describe an automated behavioral analysis framework. The basic premise of our approach is to automate the process of manually observing, recording and tabulating behavioral observations in classrooms. Much in the same way Google Analytics provides companies with website traffic statistics for improving online marketing, Classroom Analytics seeks to provide teachers with behavioral insights for creating more engaging lesson plans. Instead of simply measuring events like the number of times a student raises his or her hand during a lesson we seek to relate automated measurements to higher-level student behaviors over-time. In short, we believe that there is value in making student behavioral observations available to teachers and researchers on a second-by-second basis. Hours of classroom footage would be analyzed by software instead of busy teachers. This software would extract measurements from the video on a frame-by-frame basis, and the pattern of these measurements would be matched to expert identified representative behavior examples.

- We expect that making this information available to teachers and researchers will expand what is currently possible in terms of enabling teachers to individualize learning and help students to succeed in school. Individualized student behavioral feedback over weeks or months could help a teacher recognize that a particular student learns better in the morning than the afternoon or is more engaged during lessons that involve kinesthetic learning, small group, individual or large group instruction.

- The use of commercial off the shelf hardware would also make this proposed framework affordable; expanding research possibilities for evaluating curriculums on a large scale with hundreds of classrooms without the need for hundreds of dedicated researchers for conducting observations in classrooms.

The premise of our approach is to measure individual student behaviors given patterns of student gaze-targets overtime. Much like recognizing the tune of a song from the radio; our goal is to recognize the signature of specific behaviors student gaze-target measurements. Multiple cameras are used to record classroom lessons. Each student’s head position and orientation is extracted from the footage to compute the target of each student’s gaze over time. The resulting gaze-target patterns are then classified as belonging to one of a discrete set of behavioral categories.

Measuring student engagement is a challenging problem. Each of us expresses engagement differently. For example, one student may appear to be daydreaming but actually be concentrating on solving a difficult problem, while another student may be passively listening to a lecture. Even expert observers can have trouble distinguishing these inner states. Instead, this work focuses on measuring coarse indicators of attention. For example, is a student looking at the teacher or a friend?

In this study we adopted a time-interval sampling classroom observational behavioral coding method for establishing baseline student behavioral observations. The premise of this method is that student behavior can be understood and interpreted through statistical sampling. For example suppose a researcher is interested in measuring a student's "attending" behavior to understand the effectiveness of a proposed teaching curriculum. He or she would sit in on a lesson with a clipboard and stopwatch and mark one of two possible behavior categories "attending" or "not attending" at every 15-second time interval for 10 minutes. Each behavior would be described before hand. For example "attending" might be defined as anytime a student follows along with the lesson and does not leave his or her chair at an inappropriate time. The researcher's resulting baseline observations could then be compared with a set of observations from a modified lesson. The resulting observations could be compared and to better inform curriculum changes or provide beginning teachers' with constructive feedback.

This behavioral analysis approach offers two advantages for automated behavioral analysis in classrooms.

- First, predefined behavior categories support making comparisons between classrooms. For example empirical field study observations are typically less structured and driven by tacit knowledge. Having formal behavior definitions can make observations less susceptible to individual interpretation and provide a standard representation for comparing observations between observers and studies.
- Second, percent agreement between two or more observers can be readily computed for examining subjective bias. In this study, three observers established in advance of our observation sessions. Establishing observer reliability enabled us to consolidate multiple sets of observer data that would otherwise be time-prohibitive to collect with a single observer. The typical accepted percent agreement in the field of applied behavioral analysis is 75-80% [15]. Establishing the same reliability can be more difficult or even impractical for similar ethnographic methods. For example in a diary-study researchers could keep separate observational behavior accounts during a lesson. The free-form structure of these accounts confounding factors such as individual writing styles would present problems for associating researcher observations with particular students.

Finally, an inherit drawback of any structured coding method is that certain behaviors may go unaccounted for due to our selection of coarse behavioral categories. Here we embrace these distinctions for the simple reason that we need a starting point for automation. Its important to note that our automated measurement capabilities are much more limited than those of an experienced human observer. Next, while we do not expect our automated behavior classifications to agree 100% with those of human observers the field of applied behavioral psychology tells us that certain behavior indicators are more prevalent than others in classroom settings. For example hand raising and writing are each considered indicators for engaged student behavior. Here we adopted an observational behavioral coding method from Wasik et al [13,14] and evaluated behavior indicators as candidates for behavioral measurement automation.

The Wasik et al observational student behavioral coding method defines 8 discrete behaviors categories for encoding structured, time interval-based student behavioral observations. Each behavior category is defined by a set of observation criteria, listed in Table 1. These criteria were used to catalog student behaviors during observations. Each observation is encoded at the end of a fixed time interval as 1 of 8 possible behavior categories. The following table details each of these categories and is reproduced with permission with permission [27].

Table 1 In-person behavioral observation categories

1. Engaged Behavior: The time a child is behaving appropriately and his or her overt behavior suggests the child is actively involved in and paying attention to work. This includes productive group or independent activities such a reading, writing, painting, constructing, or working with a teaching device; assertive work such as asking for help or support, contributing information and ideas; cooperative behavior such as talking to or working with peers; appropriate dependent activity such as answering direct questions and following requests. Listening while being read to, and showing interest by asking questions, pointing, turning pages. Includes appropriate play behavior.

2. Passively Attending: The time a child is behaving appropriately, and whose overt behavior is characterized by visual attending, such as by observing another child at work or play or looking towards the teacher, but without any active involvement, or listening to instructions. Listening to directions or an assignment. Listening while being read to, but

without any other actively engaged behavior.

3. Transition: Examples of a child's behavior that could be classified as appropriate but not as attending are arranging materials for work, waiting for help from a teacher while sitting quietly in his or her chair but showing no overt productive behavior. Mainly characterized as preparing for an activity or waiting for help or instructions. Includes moving from one activity to another and cleaning up materials.

4. Nonproductive: Not engaged in appropriate or other specific inappropriate behavior. Examples include looking around, repetitive physical movements such as rocking in a chair, swaying back and forth, fidgeting, aimless wandering. Usually a solitary activity. Withdrawal.

5. Inappropriate: All behaviors in category I which are being performed outside the time limits or in the wrong setting. Examples include continuing with one activity when it is time for another to begin, mot in the assigned place while carrying out work, speaking out of turn, interrupting another person. Includes inappropriate social interactions.

6. Attention Seeking: Behaviors result in social attention from adults or peers. Examples include bothering or annoying or criticizing other students, noise making, loud talking, clowning, excessive or unnecessary requests for assistance, excessive hand raising and temper tantrums.

7. Resistive: Noncompliance. Physically or verbally resisting instructions or directions. Examples include saying "I won't do it," leaving the room, delaying.

8. Aggressive: Physical or verbal aggression. Examples include direct attacks on other children or the teacher, grabbling, pushing, hitting, kicking, destroying property, name-calling, cursing and other verbally abusive language.

Manual observations were made during a lesson in a third-grade classroom. Meanwhile, multiple cameras recorded the lesson. The recorded footage was processed to extract the target of each student’s gaze on a second-by-second basis. Manual, student observations were then paired with the respective student-gaze tracking patterns over time to train a classifier. The classifier mapped patterns of student gaze-targets to behavior categories. The resulting classifier and student-gaze target extraction framework was evaluated by comparing automatic student behavior predictions to manual behavior observations during a recorded lesson.

Here we elected to measure student gaze-targets as an indicator of student engagement. In this case we assumed that a student tends to turn his or her head in the direction that they are attending or focusing. Next, we elected to measure course head position and orientation instead in order to facilitate unencumbered tracking in a classroom. Head pose estimation and numerous other measurements can be extracted from recorded footage without encumbering students with sensors, such as a wristband for detecting hand raising. Like Stiefelhagen et al [16] and Doug et al [17] we use multiple 2D cameras to infer visual focus of attention by extrapolating what each student is looking at based on his or her head pose. Second, cognitive science research suggests a strong link between focus of attention and the direction of people’s direction of eye gaze and head orientation [17-23]. Stiefelhagen et al [17] measured 4 people’s eye and head poses independently during a meeting. The group found that each person’s eye gaze agreed 87% of the time with the direction of his or her head. This agreement led us to adopt head pose estimation as an indicator of student engagement.

Human eye gaze has been correlated with focus of attention in [17-23]. Eye gaze tended to agree with head direction in Stiefelhagen et al’s study [17]. Here we also compute head pose with the goal of inferring each student’s respective focus of attention from the object of his or her gaze. Having extracted each students gaze target from a recorded lesson we seek to make gaze-target comparisons between students as an indicator of engagement. Here our current analytics are computed on a per-student basis; however analysis could be expanded in the future to incorporate relationships between the gaze-targets of multiple students in the classroom. For example if a student is not looking at what the other students are looking at “Bob is looking out the window or Bob is looking at Sally”; this discrepancy could serve as an additional measurement for indicating engagement. However as a starting point our classification efforts have focused on relating individual student gaze-targets to behavioral categories without comparing gaze-target between students.

Here we made several assumptions regarding gaze-targets. First we assume that that there is a strong correlation between head gaze direction measurements and student engagement. Next, we assumed that the pattern of individual student gaze-targets will tend to change as students transition between behavior categories, and finally, that these pattern changes can be associated with the specific behavior categories from Wasik et al [13,14].

Figure 2 presents an outline of our Classroom Analytics framework.

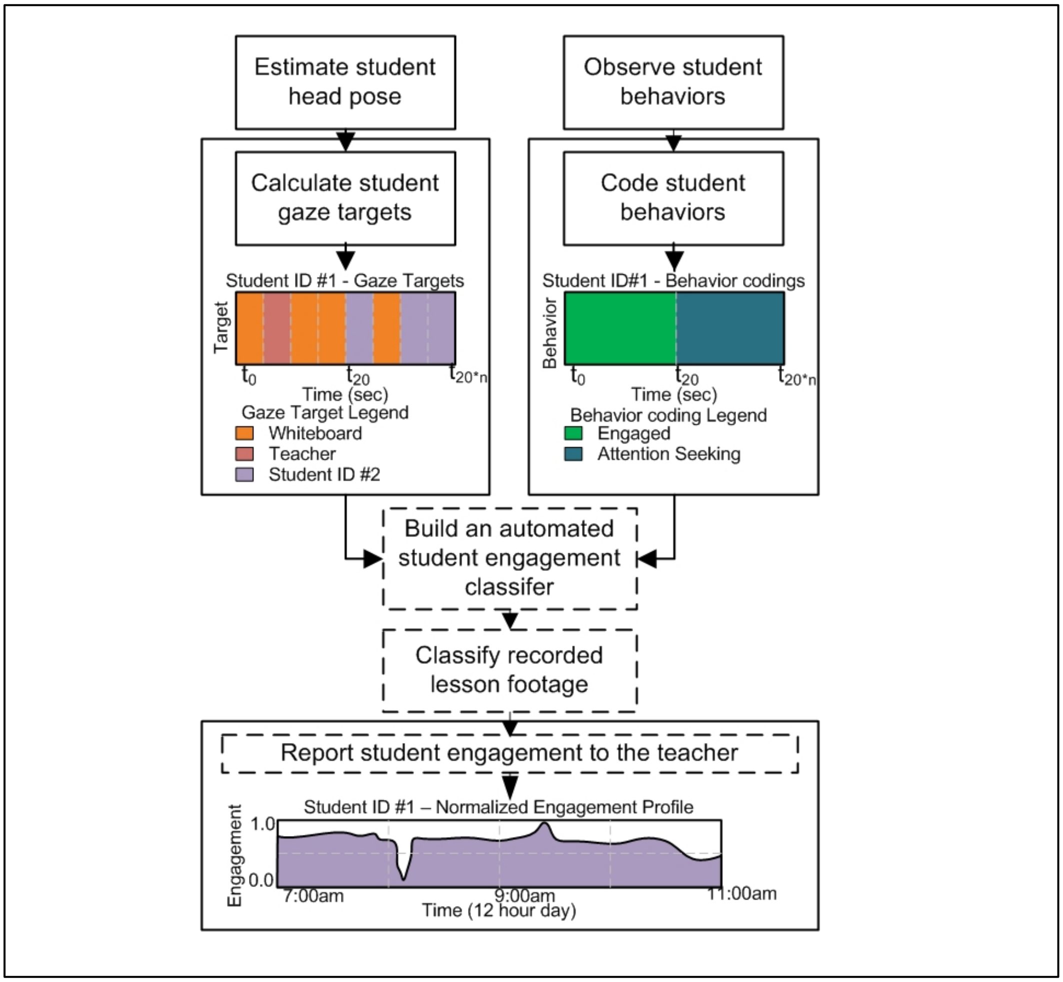

Figure 2 Classroom Analytics video analysis and analysis framework

2. Overview of Related Work

The motivation for this behavioral analysis approach comes from the field of applied behavioral sciences and recent advances in military-training and smart-office research. US military personnel receive training simulation to prepare them for conditions and situations that they may encounter in the field. Effective training requires effective feedback. Having trainers assigned to each patrol can present logistical challenges. The Office of Naval Research (ONR) sponsored Behavioral Analysis and Synthesis for Intelligent Training (BASE-IT) program uses automated behavioral analysis metrics to develop short-video segments that soldiers can view following exercises as a part of their After Action Review [6].

Hundreds of cameras are installed in an urban combat training environment at Camp Pendleton, CA. Behavioral analysis metrics are used to identify and annotate key problems, such as soldiers grouping too closely together, or pointing their rifle at a fellow soldier. Marines are each equipped with GPS receivers and accelerometers to observe his or her movement. Rifles and helmets are outfitted with Intersense inertial cubes to sense weapon acceleration and gaze direction and 37 pantilt-zoom (PTZ) and over 500 static cameras are configured to record exercises. Like BASE-IT, Classroom Analytics also focuses on maximizing the utility of training time but differs in terms of technical challenges.

Head pose estimation is required for tracking where each student is looking in the classroom. Multi-person head pose estimation presents technical challenges in a large classroom setting. Image based head pose estimation in similar office sized environments. Like Chutorian et al [9] and Stiefelhagen [16] we use 2D image based feature detection & tracking to estimate the head position and orientation of 4 people during meetings at medium sized conference room tables. Image based approaches fail if a person's eyes, nose and mouth are occluded and can no longer be tracked between frames. In practice we have observed that common student behaviors such as hand raising and chin resting are problematic for 2D facial feature tracking and that partial facial occlusions are common in busy classroom environments. Having started with 2D feature tracking our group is in the process of integrating multiple depth sensing cameras and 3D image-feature tracking for head pose estimation that is more robust to occlusions.

Measuring student gaze targets in a classroom present similar challenges to those presented in recent work dealing with head pose estimation in office-sized environments [7,8,16]. Here we employ ray casting to compute each student's focus of attention or target of gaze. Like Stiefelhagen [16] gaze targets are determined by intersecting a ray from the child's head and gaze direction with manually defined bounding boxes objects in the room.

Existing commercial products require considerable teacher intervention. Teachscape is a commercial data collection, analysis and reporting tool for evaluating student engagement in classrooms [1,2,5], which requires considerable manpower. Trained evaluators are tasked with manual data reporting that limits scalability, is labor intensive and does not reflect individual student needs. Observations made on Personal Digital Assistants (PDA) by evaluators are then tabulated and made available to teachers and the school principal on a secure website.

Finally, engagement coding has been practiced for decades; but training and coding reliability continue to present bottlenecks for the widespread use of these metrics in classrooms. Teachers must be trained in a particular coding method. Next, the rule of thumb for applied behavioral psychology research publications is that independent observations of the same student from multiple coders should agree to within or above 80% [15]. Efforts have been made in recent years to streamline data-entry [10], but these training and reliability hurdles still remain. There have been no efforts that we know of to automate behavioral analysis in a classroom setting. To the best of our knowledge we are the first to explore automatic gaze-target tracking as an indicator of student engagement in classrooms.

3. Problem Statement

The problem addressed in this paper is to model and classify student engagement from sequences of student gaze targets (what students are looking at) over-time. Multiple cameras are installed in a third grade classroom. Three expert observers classify individual student behaviors into discrete categories during a recorded lesson. Face tracking software is used to extract individual student gaze directions for the footage and calculate what each student is looking at on a second-by-second basis as shown below.

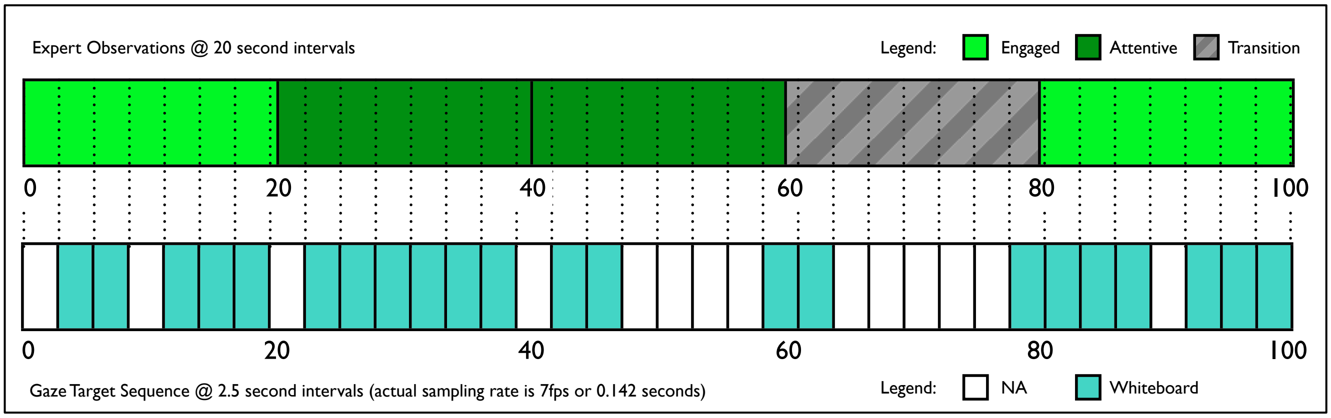

Figure 3 Inputs (Top) expert student behavior observations for a given student (Bottom) gaze target measurements for the same student during a recorded classroom lesson.

In this paper we used HMMs to classify extracted sequences of individual student gaze targets from recorded lessons into behavioral categories [25]. Hidden Markov Models (HMM) are probabilistic frameworks that define a conditional probability distribution over a set of labeled sequences given a particular observation sequence. HMMs are considered a good fit for our behavior classification application for three reasons:

- First, each of our behavior categories are independent. For example according to the coding scheme students can be engaged or attentive but not both at the same time. HMMs require that these class labels be independent by definition.
- Second, measured gaze-target observations tend to be long continuous sequences. HMMs are well suited for making predictions based on long-sequences of observations and can predict class labels at each timestamp without having to iterate over the entire set of preceding gaze-targets. Here we averaged 72,020 gaze targets per student during 30 minute recorded lessons. HMMs served as a tractable behavior classification method.
- Third, HMMs can easily be expanded to account for additional types of automated observations. For example automated measurements such as student hand raising or the teacher's proximity to each student could be extracted from recorded lessons as additional observation sequences for predicting student engagement.

Hidden Markov Models (HMM) was selected as an approach for modeling the relationships between student behavior types and automated gaze target sequences. HMMs are commonly used to infer the most likely hidden state of a system given a probabilistic distribution over a sequence of observation tokens and have long been used for voice pattern and gesture recognition. HMMs are thought to be a good fit for modeling student engagement. First, the approach is tractable for long sequences. In this study a 31-minute recording produced an average of 10,620 gaze targets per student. Second, HMMs lend themselves well to incorporating additional types of observations. For example, we are incorporating student hand-raising and teacher-student proximity as additional observation types. The assumption is that certain sequences of student gaze targets occur with a higher or lower probability for certain behaviors. For example engaged children tend to look at the teacher when he or she is giving instructions before a lesson.

The disadvantage to this approach is well-behaved students. Low frequency student behaviors are difficult and time-consuming to collect in a classroom context. For example in this paper we happened to observe a student whose behavior was more or less consistent during each of our 31-minute observation sessions. The student transitioned between engaged and attentive behaviors. No explicit disengaged behaviors were observed. HMMs typically require more training samples than other methods such as Support Vector Machines and Principal Component Analysis [28]. In addition, HMMs are often less suitable for modeling higher-level trends. In this paper, the Viterbi Algorithm is used to calculate the most likely sequence of hidden behavior states given a historic sequence of student gaze target observations. The best length for these historic sequences (how far back in time to compute joint hidden state and observation probabilities) is limited by processing requirements and difficult to know in advance. For example, perhaps students become less disengaged before certain class periods or before long weekends. The remainder of this paper will detail the group's methodology and present initial results.

4. Methodology

Here we present a methodology for exploring the application of behavioral analysis in a real-classroom setting. The researchers recruited a teacher and her third grade class as participants and received IRB support.

Figure 4 Expert observers (shown right) coded student behaviors during recording lessons as future training input for constructing a student engagement classifier.

Multiple cameras were installed in a third grade classroom. Five color cameras and four Microsoft Kinect depth-sensing cameras were used for these experiments. The five color cameras focused on the front row of the students for image-based face tracking, Figure 1. Two Kinect cameras captured front-row seated students at a distance of ~10' from the front of the room and the two remaining Kinect cameras were mounted from the ceiling captured the same students from a top-down perspective. Each camera was then calibrated using OpenCV [26] to estimate the camera's position, orientation, focal length and lens distortion coefficients, see Figure 5.

Figure 5 (Left) ceiling mounted depth sensing camera (Right) wall mounted color and depth sensing cameras

Three graduate students from the field of educational psychology observed student behavior during the recording sessions to code behaviors into 8 discrete behavioral categories, see Table 1 for behavior descriptions. The footage was then analyzed using face-tracking software to estimate where each student was looking on a second by second basis. The resulting gaze patterns were correlated with expert observer observations. We then used these behaviors and gaze target sequences to train an automated HMM based student engagement classifier for a single student. The HMMs was trained using two types of input 1) manual student behavioral observations and 2) automated gaze-tracking measurements from face tracking software. The following section describes how these inputs were collected during an experiment.

5. Experiment

This section describes a pilot study recording and observation session with a third grade teacher and her students at the Union Independent School in Durham NC. The goal of the experiment was to collect expert student behavior observation data as input for training and evaluating an automated student engagement classifier. Existing datasets of coded student behaviors from video are not available at this time. The experts followed a coding schema that is being developed at the Frank Porter Graham (FPG) child development institute at UNC, NC. Educational psychologists can later follow this schema to make similar observations. This experiment explored the use of a Hidden Markov Model can serve as a first approximation for classify student engagement based on these observations and automated gaze target measurements.

5.1. Hidden States - Expert student behavior observations

Three expert observers, PhD students at the UNC Frank Porter Graham Institute, coded a student's engagement during three recording sessions. The observers were trained under Dr Barbara Wasik, a nationally recognized expert on coding child behaviors in classrooms. In total three observation sessions were run. Each recording session ran for a total of 10 minutes. The observers used a coding schema for classifying an individual student's behavior during a recorded lesson. The coding scheme classified student behaviors as belonging to one of eight discrete behavior types: engaged, attending, transitioning, non-productive, inappropriate, attention seeking, resistive and aggressive. Behavior observations were logged using an iPod application that annotated with a timestamp for later comparison with recorded classroom video footage. Figure 6 shows a screenshot of the iPod recording application; Figure 7 shows an example of a student's behavior coding's from two observers.

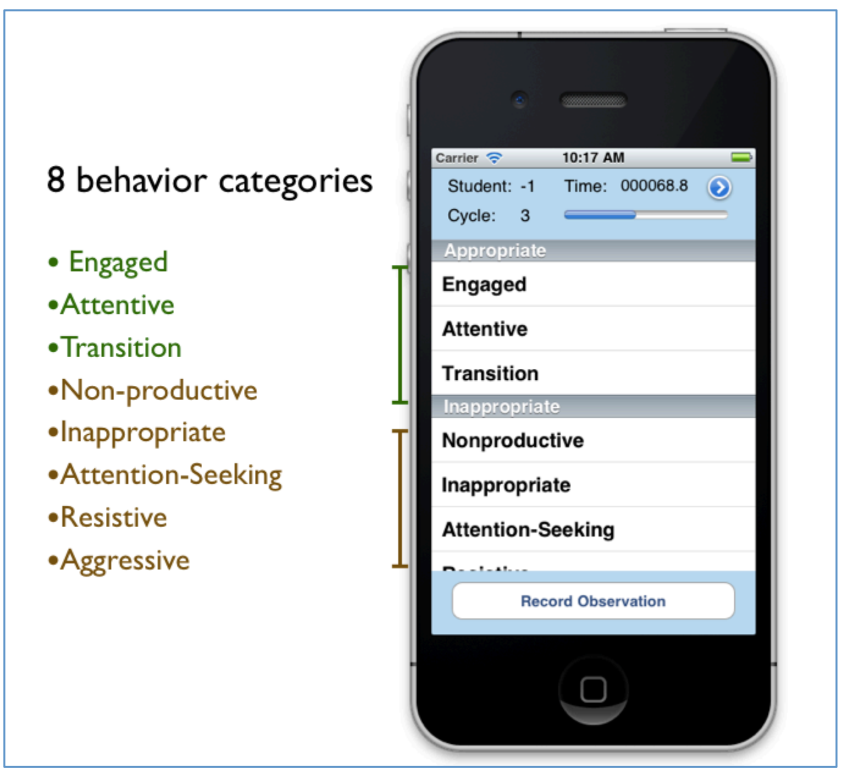

Figure 6 iPod coding application and coding categories

Each expert observed the same student and logged his or her behavior during the lesson. These observations were then resampled to coincide with the respective student's automated gaze target measurements and used as training input for a HMM. The key assumption in this sampling is that behaviors are consistent during time-intervals. For example, student behavior can change considerably in a period of 20 seconds. If we choose time intervals that are too small, expert observations will be accurate but a limited number of gaze target samples will be available for training. If instead we choose an interval that is too large then student behaviors would generalize over time but more gaze tracking samples will be available for training. In this case, the we assume that student behavior is consistent between expert observations and each student gaze targets is included during this interval; see Figure 7 for a comparison of these two time-series.

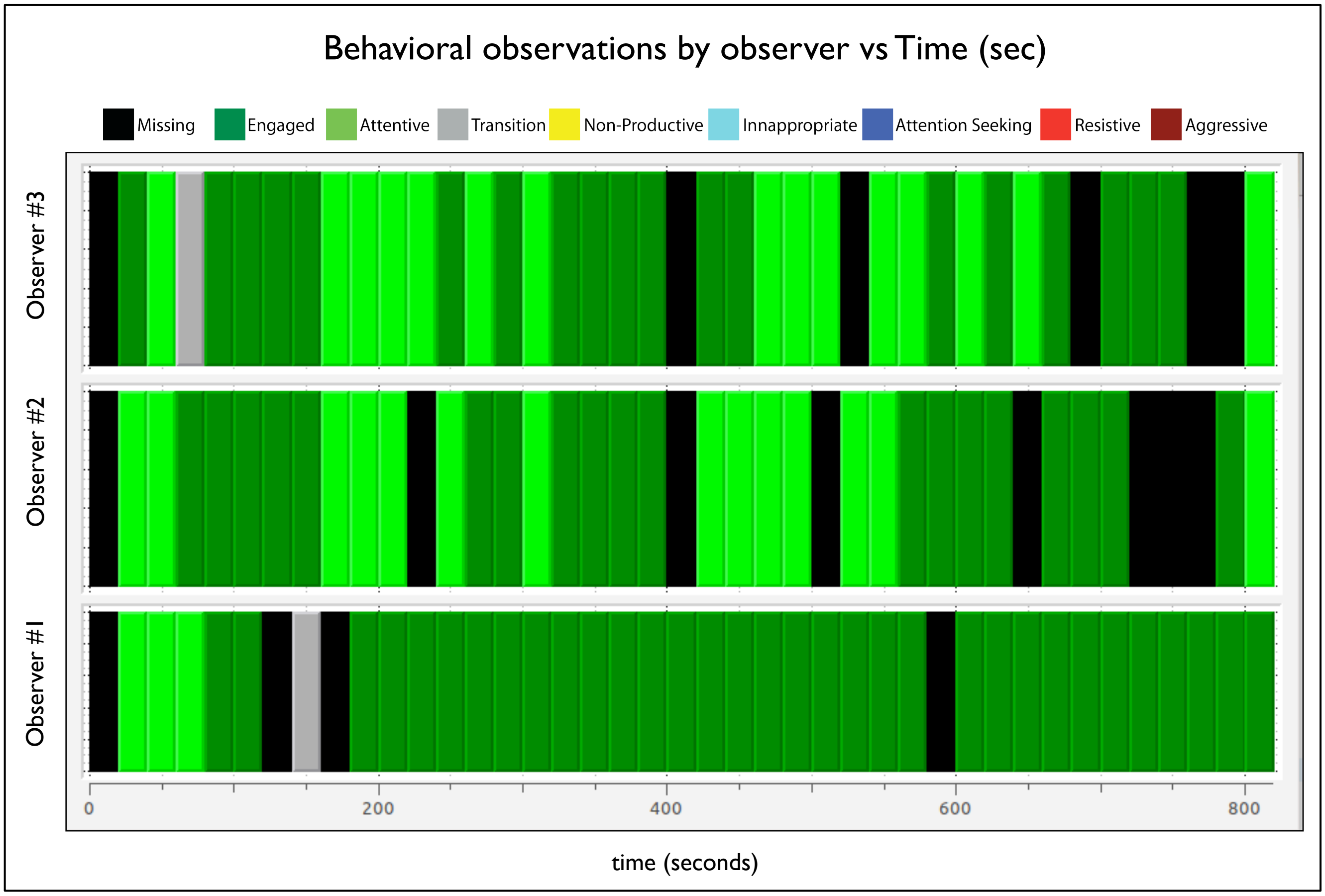

Figure 7 iPod recorded behavior observations from two observers (top and bottom rows)

5.2. Observations - Automated student gaze target sequences

Marker-less head pose estimation was performed using commercial face tracking software [8]. The Pittsburgh Pattern Recognition (PittPatt) SDK is used to extract individual student gaze directions for one or more students appearing in the recorded footage, Figure 5. Image features such as the eyes, nose and mouth are tracked between frames to estimate student head pose [10]. The resulting sequence of time-stamped head pose measurements (position and orientation) was recorded for comparison with time-stamped expert behavioral observations.

Each student gaze student targets were calculated by intersecting his or her gaze direction with manually defined objects in the classroom. First, a 3D model of the classroom was constructed by plotting matching triangulated image features between six hundred classroom photos using Bundler [11, 12]. The resulting 3D model was manually aligned with the calibrated location and orientation of the recordings cameras to established relationship between gaze directions and physical objects in the classroom. Manually aligned bounding box objects were added to the 3D model to define discrete gaze targets for measuring where students were looking at during the lesson. Nine bounding boxes were defined in total, one for a classroom whiteboard and eight for each student in the front row.

The teacher's head pose was not measured due to our tracking limitations during the time of the experiment. The teacher walked around he classroom to address individual student questions during the recorded lesson. This movement prohibited the use of a similar fixed head assumption used for estimating student head locations. The team has since upgraded our software following the experiment for tracking the teacher's head position using Microsoft Kinect depth sensing cameras, see the future work section for more details.

Figure 8 PittPatt SDK head pose estimation screenshot

Next, each student's head location was to a position above his or her desk in the 3D classroom model. Head orientations were computed using the PittPatt SDK. Each student's gaze direction was transformed into a common world coordinate system. Finally, each student's gaze target was measured on a frame-by-frame basis by casting a ray from the student's head position in the direction of their gaze direction and intersecting one or more of these bounding box objects. If no bounding boxes were intersected then no the measurement was recorded. The resulting gaze target sequences were used as training input for a HMM. Figure 9 shows student head locations visualized as red cylinders.

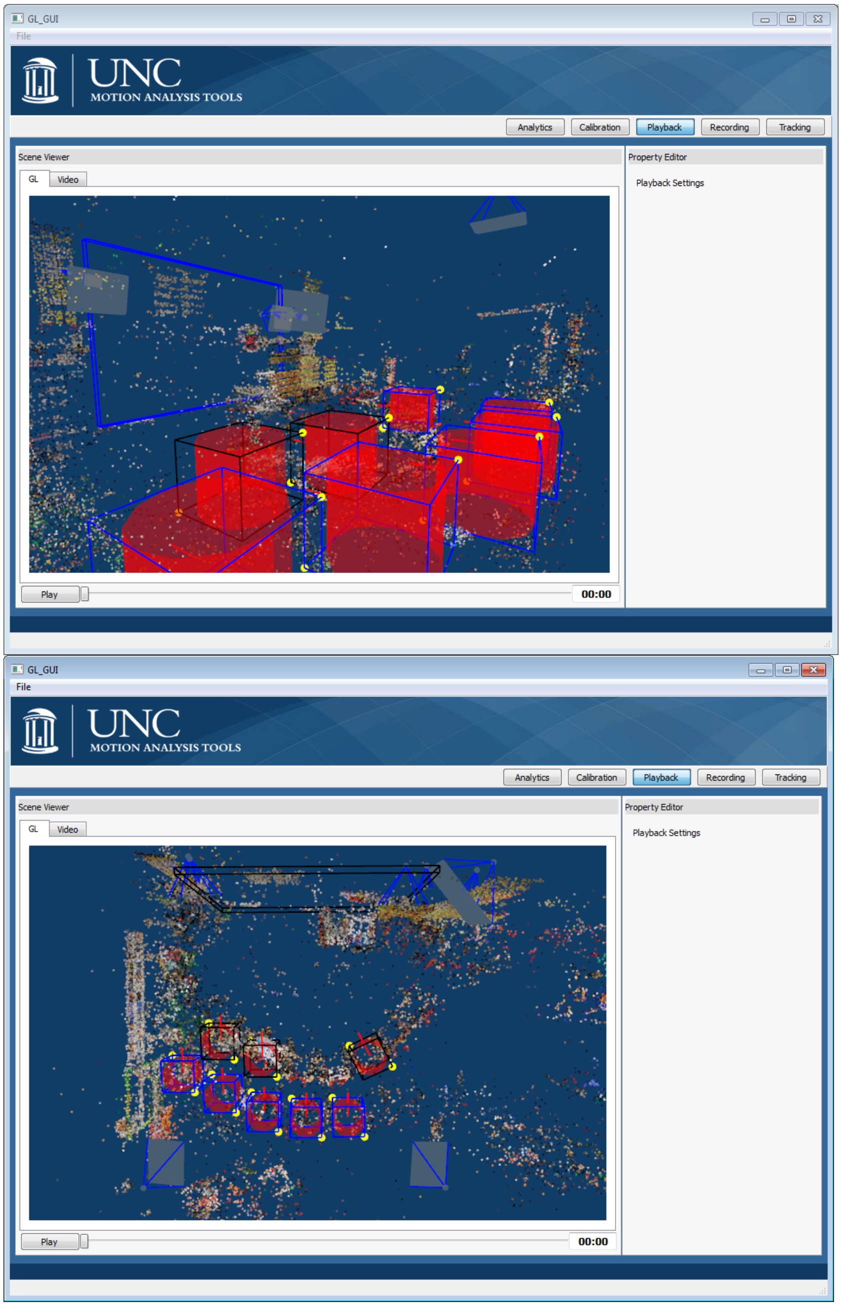

Figure 9 Example student gaze-target intersections. Intersected objects are highlighted by black colored bounding boxes. These bounding boxes reflect objects that are seen by one or more students.

Each student's gaze direction was used to infer a gaze target ie: "what each student was looking at", on a second by second basis. To do this, we first projected a ray from each student's head location along his or her direction of gaze. Manually aligned bounding boxes were modeled to coincide with physical objects in the classroom. Nine such bounding box objects were defined in total. These objects included a classroom whiteboard and desk location of eight students.

Each student's gaze target was then assigned as either the name of the first object to intersect his or "no target" if no object was intersected. Figure 6 highlights gaze-targeted objects with a black bounding box. These gaze targets form patterns over time. If no bounding boxes were intersected then the measurement was skipped. The resulting gaze target sequences were coupled with expert observed behaviors to train a HMM for identifying probable student behaviors without the need for expert intervention.

6. Initial Results

This section describes training and evaluating a single-student, Hidden Markov Model (HMM) student engagement classifier. Here our goal was to classify student gaze-target sequences into separate behavioral categories. This process starts with pairing our manual observations with our automatic student gaze targets.

In this study 3 graduate students categorized a student's behavior at 20-second intervals during three ten-minute sessions. Here we assume that student behavior remains consistent between each of these intervals. For example suppose that one of the observers records "attentive" at 20 seconds, "engaged" at 40 seconds and "attentive" at 60 seconds. The student's gaze gargets are grouped into respective behavioral categories during each interval. This grouping produces representative sets of gaze sequences for each behavior category, Figure 3. In this case, student was well behaved during the lesson and only 3 of the 8 behavior types appeared, engaged, attentive and transition, see Table 1 for behavior descriptions.

The following figure shows initial HMM state transition probabilities used for this study. Each state was assigned an 80% self-state transition probability ie: by default a given student is 80% likely to remain in his or her current behavior state. The probability of a given state being the initial state was set to be uniform.

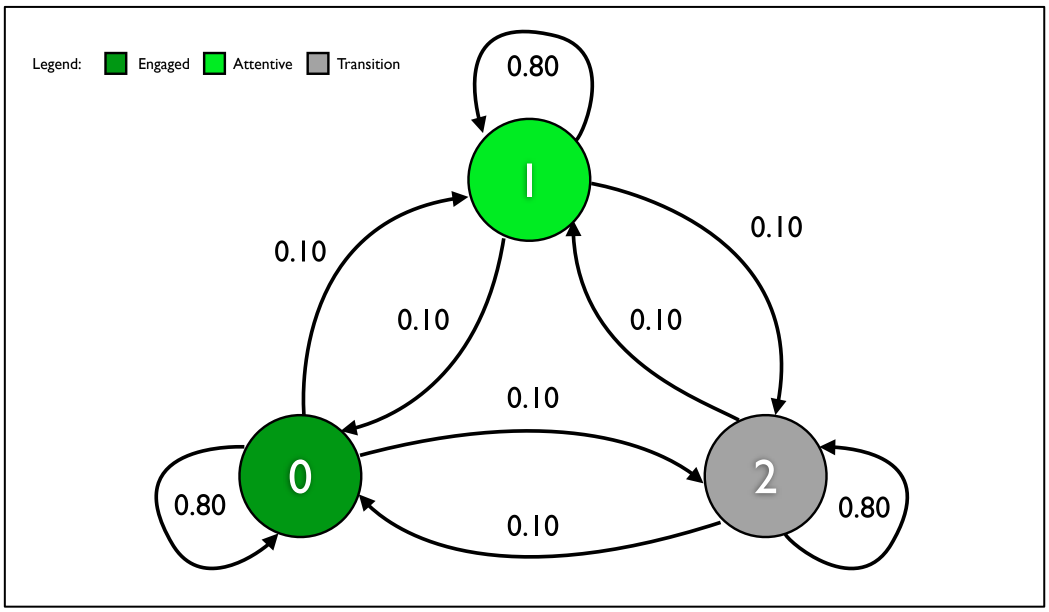

Figure 10 Initial Hidden Markov Model state-transition probabilities between discrete behavior types

The HMM state transition probabilities and state probability distribution functions were trained using Baum Welch. The Baum Welch algorithm assigns initial values to these HMM's parameters. The algorithm then applies maximum likelihood estimation to adjust these parameters such that the probability that the model assigns to the training set is maximized.

Measured gaze target sequences were separated into training and testing groups for evaluating the classifier. The test set consisted of 10% of the gaze sequences. The remaining 90% of the sequences were used for training the HMM. These gaze targets were randomly selected for each behavior type.

The following table shows the total number of gaze target instances for each behavior type given these sequences.

Table 2 Test Dataset - #Gaze Target Instances per Behavior Type

Behavior Type	# Gaze Targets Instances
Engaged	1,019
Attentive	550
Transition	8,705

The distribution of these measured behaviors is non-uniform. The number of gaze target instances for "transition" and "attentive" behaviors were much lower than "engaged" behavior. Ideally we would like to have an even distribution of representative samples from each category to prevent unstable classifications due to over-fitting.[1]

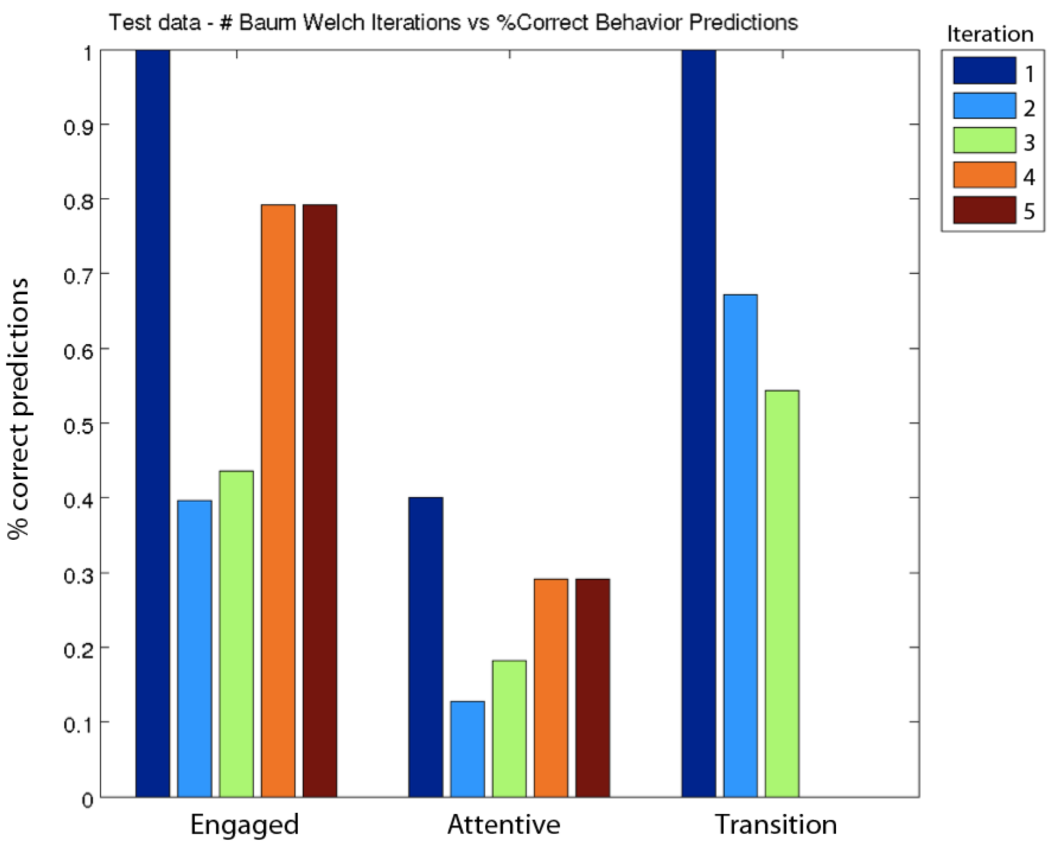

Figure 11 Hidden Markov model (HMM) – test dataset without using K-Nearest Neighbors (KNN) for establishing initial state transition probabilities and state probability distribution functions.

[1] Note: There is also minimal variance between engaged and attentive behaviors.

The same HMM percent correct steps were run using 100% of the training data as test data. The expectation is that this will produce the best possible matching score as a result of over-fitting. The results further confirm the hypotheses that a greater number of behavior observations and gaze target data instances are required for classifying engagement. Engaged behavior is predicted 80% of the time ie: whenever a student is looking at a peer or whiteboard, while attentive behavior is predicted less than 40% of the time, worse than chance, suggesting that additional behavior measurements are needed given minimal gaze target variations in the overall data set.

Table 3 Training Dataset - - #Gaze Target Instances per Behavior Type

Behavior Type	# Gaze Targets Instances
Engaged	6,585
Attentive	3,565
Transition	54,470

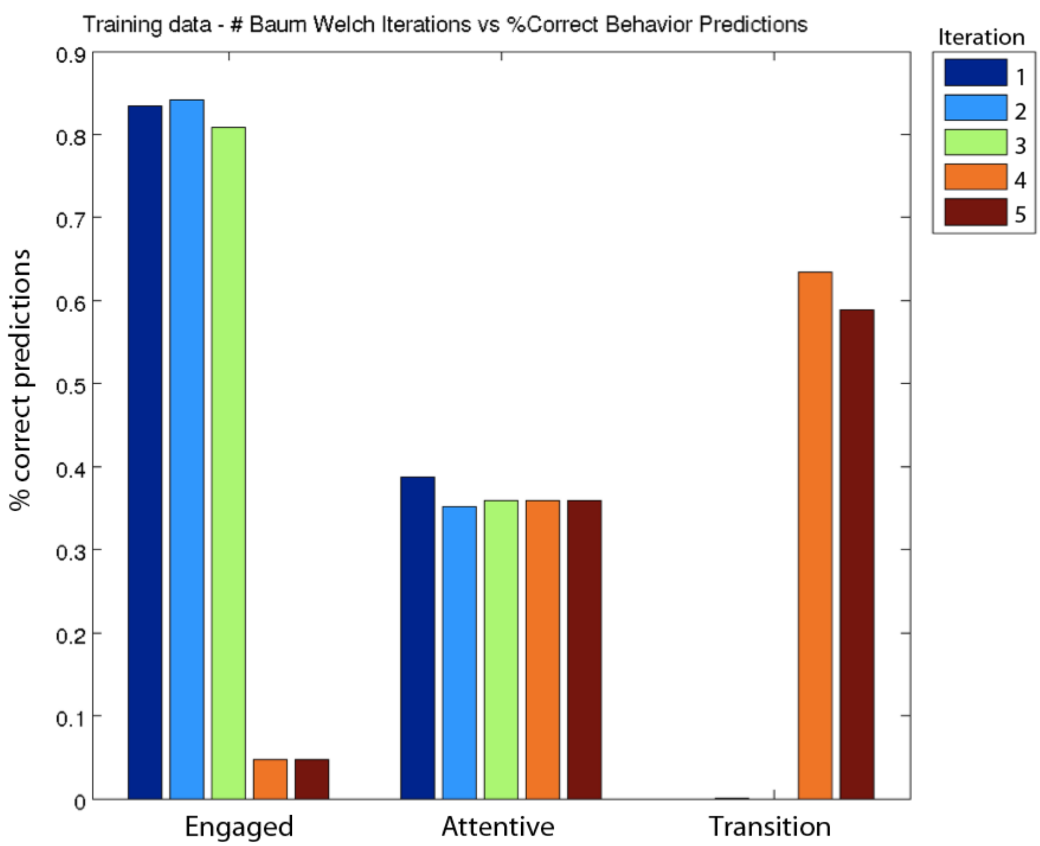

Figure 12 Hidden Markov model (HMM) – training dataset without using K-Nearest Neighbors (KNN) for establishing an initial state transition probabilities and state probability distribution functions.

The poor HMM classification results suggest a need for additional behavioral observation data. For example "attentive" behavior was predicted less than 40% of the time in Figure 12. If the behavior categories are indeed independent than additional training data should bring these results to close to 100%. Next, the states could be redesigned. For example, we could define additional HMM states that are more specific to a given situation and would therefore encode additional variability. For example, a gaze tracking state that includes the teacher's proximity to the student at a given time.

7. Conclusion

This paper introduces Classroom Analytics, a term for an automated behavioral analysis framework, for enabling teachers to efficiently review student behaviors. Here we present initial results by exploring the application of behavioral analysis in a real-classroom setting. The researchers recruited a teacher and her third grade class as participants. Multiple cameras were installed in the classroom to record lessons. Expert observers from educational psychology observed student behavior during the recording sessions to code behaviors into discrete categories such as engaged, attentive and resistive. The footage was then analyzed using commercial face tracking software to estimate where each student was looking on a second by second basis. The resulting gaze patterns were correlated with observer made observations to train a HMM classifier. The poor classification results from this HMM classifier suggest a need for additional behavioral observation data and a revised set of states. The group has since simplified the HMM classifier to “engaged” and “not engaged” instead of containing separate states for each of the 8 categories.

The work represents a first step in automating the measurement student engagement as a teacher feedback tool for personalizing lessons and identifying ways to better help students succeed in school. The group is developing a student engagement classifier and head pose tracker using depth-sensing cameras. The research sets the groundwork for continued work to improve the capability of an automated system in a real-world classroom setting. For example, work will be done to distinguish one student from another when two children in an active classroom work closely together. The connected component algorithm will connect adjoining pixel regions; additional steps such as implementing a particle filter and detecting merge and split will be implemented to better address this situation.

8. Future Work

In the near term we will complete building the presented prototype system and finish developing an automated classifier for measuring the engagement of students in a classroom.

Multiple 3D depth-sensing cameras have since been installed in a classroom to track the head locations of a teacher and his or her students. Here we are experimenting with depth image segmentation and connected components labeling algorithms for tracking head locations and recently implemented a random regression forest based head tracking algorithm for more robust head pose estimation in the presence of partial facial occlusions.

Next, we have upgraded our iPod app with an alarm clock feature for scheduling synchronized in-person observations. In the past our observers pressed a start-record button on the count of three. The upgraded iPod app lets set a common starting time for observations for example “start recording at 9:30am”; as a result we anticipate that that subsequent in-person observations will better synchronized and less susceptible to human error.

Finally, the group will then train a student engagement classifier using student gaze target patterns from 1 hour of recorded footage and evaluate classification results against an additional 30 minutes of in-person observations. In the longer term we will develop a recording and playback database that will enable a teacher to view a survey of each student’s engagement at a glance and explore recorded lessons at a click by to better understand the dynamics of a particular moment in his or her classroom.

References

[1] “Teachscape Classroom Walkthrough” Teachscape.com Web. 10.20.2010.http://www.teachscape.com/classroom-walkthrough/index.html

[2] Derry, S. J., Siegel, M., Stampen, J., & STEP Team (2002). “The STEP system for collaborative case-based teacher education: Design, evaluation & future directions.” In G. Stahl (Ed.), Proceedings of Computer Support for Collaborative Learning (CSCL) 2002

[3] RJ Marzano. “Classroom instruction that works: Research-based strategies for increasing student achievement” McRel. 2001

[4] “School of One. NYC Department of Education”. NYC Department of Education. Web. Online 10. 20.20010. http://schools.nyc.gov/community/innovation/SchoolofOne/default.htm

[5] “Teachscape Classroom Walkthrough Research Base” Teachscape.com Whitepaper. Web. 10.20.2010. http://www.teachscape.com/classroom-walkthrough/pdf/CWT4pager_web.pdf

[6] Amela Sadagic, Greg Welch, Chumki Basu, Chris Darken, Rakesh Kumar, Henry Fuchs, Hui Cheng, Jan-Michael Frahm, Mathias Kolsch, Neil Rowe, Herman Towles, Juan Wachs, and Anselmo Lastra. New Generation of Instrumented Ranges: Enabling Automated Performance Analysis. In Proceedings of the 2009 Interservice/Industry Training, Simulation, and Education Conference

[7] Voit, M., Stiefelhagen, R. 3D user-perspective, voxel-based estimation of visual focus of attention in dynamic meeting scenarios ICMI-MLMI 2010

[8] Ba, S.O.; Hung, H.; Odobez, J.-M.; , "Visual activity context for focus of attention estimation in dynamic meetings," Multimedia and Expo, 2009. ICME 2009. IEEE International Conference on , vol., no., pp.1424-1427, June 28 2009-July 3 2009

[9] Murphy-Chutorian, E.; Trivedi, M.M.; , "3D tracking and dynamic analysis of human head movements and attentional targets," Distributed Smart Cameras, 2008. ICDSC 2008. Second ACM/IEEE International Conference on , vol., no., pp.1-8, 7-11 Sept. 2008
[10] PittsburghPatternRecognition (SDK). Online 4.15.11 http://www.pittpatt.com/
[11] Noah Snavely, Steven M. Seitz, Richard Szeliski. Photo Tourism: Exploring image collections in 3D. ACM Transactions on Graphics (Proceedings of SIGGRAPH 2006), 2006.
[12] Noah Snavely, Steven M. Seitz, Richard Szeliski. Modeling the World from Internet Photo Collections. International Journal of Computer Vision, 2007.
[13] Coding procedure for teachers. Kubany, Edward S.; Sloggett, Barbara B. Journal of Applied Behavior Analysis, Vol 6(2), Sum 1973, 339-344.
[14] Behavior modification with culturally deprived school children: Two case studies. Wasik, Barbara H.; Senn, Kathryn; Welch, Roberta H.; Cooper, Barbara R. Journal of Applied Behavior Analysis, Vol 2(3), 1969
[15] Fisher, Wayne W, Cathleen C. Piazza, and Henry S. Roane. *Handbook of Applied Behavior Analysis*. New York: Guilford Press, 2011. Print
[16] Stiefelhagen, R., "Tracking focus of attention in meetings," *Multimodal Interfaces, 2002. Proceedings. Fourth IEEE International Conference on* , vol., no., pp. 273- 280, 2002
[17] N. Emery. The eyes have it: the neuroethology, function and evolution of social gaze. Neuroscience and Biobehavioral Reviews, 24:581–604, 2000.
[18] M. von Cranach. The role of orienting behaviour in human interaction. In A. H. Esser, editor, Environmental Space and Behaviour. Plenum Press, New York, 1971.
[19] M. Argyle and M. Cook. Gaze and Mutual Gaze. Cambridge University Press, 1976.
[20] Conty, Laurence, Grèzes, Julie. Look at me, I'll remember you: The Perception of Self-Relevant Social Cues Enhances Memory and Right Hippocampal Activity. Human Brain Mapping. Wiley Subscription Services. 2011
[21] George N, Conty L (2008): Facing the gaze of others. Neurophysiol Clin 38:197–207.
[22] Stephen R.H. Langton, Roger J. Watt, Vicki Bruce, Do the eyes have it? Cues to the direction of social attention, Trends in Cognitive Sciences, 2000
[23] Ligeng Dong, Huijun Di, Linmi Tao, Guangyou Xu and Patrick Oliver. Visual Focus of Attention Recognition in the Ambient Kitchen. Computer Vision – ACCV 2009
[24] Milton Chen. 2003. Visualizing the pulse of a classroom. In Proceedings of the eleventh ACM international conference on Multimedia (MULTIMEDIA '03). ACM, New York, NY, USA, 555-561. DOI=10.1145/957013.957130 http://doi.acm.org/10.1145/957013.957130
[25] Juang, B.-H. 2003. Hidden Markov Models. Encyclopedia of Telecommunications. Print.
[26] OpenCV. http://opencv.willowgarage.com/wiki/. Online 11.20.11
[27] B. Wasik. CHILDREN'S CLASSROOM BEHAVIOR "Cheat sheet". University of North Carolina at Chapel Hill. Frank Porter Graham (FPG) Child Development Center. 11.20.11. Document
[28] Cristianini, N and Shawe-Taylor, J (2000) An introduction to Support Vector Machines. Cambridge University Press. Print