UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

A Hidden Markov Model for Analyzing Eye-Tracking of Moving Object

Permalink

https://escholarship.org/uc/item/92r3s94h

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 40(0)

Authors

Kim, Jaeah
Singh, Shashank
Velde, Anna Vande
et al.

Publication Date

2018

eScholarship.org Powered by the California Digital Library
University of California

A Hidden Markov Model for Analyzing Eye-Tracking of Moving Objects

Jaeah Kim[1*] **(jaeahk@andrew.cmu.edu)**
Shashank Singh[2*] **(sss1@cs.cmu.edu)**
Anna Vande Velde[1] **(avandeve@andrew.cmu.edu)**
Erik D. Thiessen[1] **(thiessen@andrew.cmu.edu)**
Anna V. Fisher[1] **(fisher49@andrew.cmu.edu)**
[1] Carnegie Mellon University, Department of Psychology, Pittsburgh, PA 15213, USA
[2] Carnegie Mellon University, Machine Learning Department, Pittsburgh, PA 15213, USA

Abstract

Eye-tracking provides an opportunity to generate and analyze high-density data relevant to understanding cognition. However, while objects in the real world are often dynamic, eye-tracking paradigms are typically limited to assessing gaze toward static objects. In this study, we propose a generative framework, based on a hidden Markov model, for using eye-tracking data to analyze behavior in the context of multiple moving objects of interest. We apply this framework to analyze data from a recent visual object tracking task paradigm, TrackIt, for studying selective sustained attention in children. We also present a novel 'supervised' variant of TrackIt that we use to tune and validate our model, while providing insights into the visual object tracking abilities of children and adults.

Keywords: eye-tracking; visual object tracking; hidden Markov model; TrackIt; selective sustained attention

Introduction

Eye-tracking provides temporally rich behavioral data (gaze) that is closely linked to many cognitive functions. It has been widely used to study cognition, in diverse research areas including category learning (Rehder & Hoffman, 2005), visual attention (Doran, Hoffman, & Scholl, 2009), sports expertise (Smuc, Mayr, & Windhager, 2010), visual perception (Gegenfurtner, Lehtinen, & Säljö, 2011), implicit bias and stereotype (Pyykkönen, Hyönä, & van Gompel, 2009), language processing (Barr, 2008) and psychological disorders such as schizophrenia (Holzman et al., 1974). Beyond psychology, eye-tracking applications include safety evaluation in driving (Palinko, Kun, Shyrokov, & Heeman, 2010), usability studies in human-computer interaction (Jacob & Karn, 2003), and diagnosis of Alzheimer's disease (Biondi, Fernandez, Castro, & Agamenonni, 2017).

Most of these applications rely on the extensive work that has been done based on two important components of gaze: *fixation* (maintenance of gaze on a single location) and *saccade* (quick movement of gaze between two fixations) (Cassin, Solomon, & Rubin, 1984). There exist well-documented standards for identifying and analyzing fixations and saccades in eye-tracking data (Duchowski, 2017), and meta-analysis has shown that the most commonly used eye-tracking measures are number of fixations, mean fixation duration, and gaze duration (a function of multiple fixations) (Jacob & Karn, 2003). These have been incorporated into user-friendly analysis software built into commercial eye-trackers, and there also exists open-source software for fixation- and saccade-based analyses of generic eye-tracking data (e.g., Dink and Ferguson (2015)). This has facilitated adoption of fixation- and saccade-based eye-tracking methods as standard tools for investigating cognition and behavior.

While fixations and saccades describe most human eye movement in response to stationary or rapidly moving visual stimuli, tracking of smoothly moving stimuli obeys a different dynamic, namely *smooth pursuit* – slow eye movement that maintains the image of a moving object on the fovea (Cassin et al., 1984). Far less work with eye-tracking has studied smooth pursuit, in part due to a relative lack of analysis tools. A recent 400-page review of eye-tracking methodology mentions smooth pursuits only thrice and notes that 'a robust and generic algorithm for their detection is currently an open research problem' (Duchowski, 2017, p. 176).

In this paper, we propose a novel hidden Markov model (HMM) approach to analyzing eye-tracking data in the context of multiple moving objects of interest. Given continuous gaze data collected from a subject tracking moving objects with known positions over time, our model can accurately determine the object being tracked at each time point. Since smooth pursuit is intimately tied to tracking smoothly moving objects, this model effectively provides a way of analyzing smooth-pursuit movement. We anticipate that our model may be useful for researchers in cognitive science and related areas and have made a Python implementation freely available.

A Hidden Markov Model Approach

Hidden Markov Models (HMMs) are a popular generative model for time series data, in which observed data is assumed to be drawn, at each time point, from a distribution depending on an unobserved *hidden state*. An HMM is a natural fit for the problem at hand; at each time point t, the subject is looking at *something* $S(t)$ (the hidden state), and we observe eye-tracking data $X(t)$ that is primarily a function of $S(t)$ and random noise. Unlike simpler models that consider data at each time point independently, an HMM mitigates noise and easily handles complex scenarios such as object collisions (when multiple objects briefly occupy the same space), without losing the fine temporal resolution of eye-tracking data.

Selective Sustained Attention and TrackIt

Selective sustained attention (SSA) is an important cognitive process that enables everyday functioning and task performance by allowing us to: 1) choose components of our en-

[*]These authors contributed equally to this work.

vironment to process at the exclusion of others and 2) maintain focus on those components over time. SSA relies on both endogenous factors (e.g., internal goals) and exogenous factors (e.g., stimulus salience), and studying how these factors develop and interact in guiding attention during childhood is of special interest for SSA development research (O'Connor, Manly, Robertson, Hevenor, & Levine, 2004).

TrackIt is a child-appropriate visual object-tracking task recently developed to measure SSA, that can capture differential contribution of exogenous and endogenous control of attention and allow flexible assessment over a range of developmental years (including pre-school years, for which there is a relative lack of appropriate SSA measures), with parameters for adjusting difficulty with age (Fisher, Thiessen, Dickerson, & Erickson, 2013; Fisher, Thiessen, Godwin, Kloos, & Dickerson, 2013; Kim, Vande Velde, Thiessen, & Fisher, 2017). In the TrackIt task, participants visually track a single target object moving about on a grid, among other moving distractor objects. At the end of each such trial, all objects vanish from the grid, and participants are asked to identify the final grid cell location the target occupied before vanishing.

Prior studies using TrackIt have measured task performance mainly in terms of this final response – whether the final grid cell was correctly identified. However, this measure has several limitations. For example, Kim et al. (2017) suggested that many behavioral 'errors' may be attributable to subjects' limited visual resolution when identifying the final grid cell location of the target (thereby clicking an adjacent cell). Also, this measurement is made *after* task and only indirectly tells us what participants do *during* task.

To address these limitations of data available directly from TrackIt, we began collecting eye-tracking data. Analyzing these rich data, however, involves a non-trivial technical challenge, namely that of robustly identifying the object a participant is tracking from noisy eye-tracking data, even when objects are moving, crowded, and potentially overlapping. This problem motivated the methods proposed in this paper, which we present in the belief that they may be useful for analyzing eye-tracking data in more general experimental contexts.

Related Work

There has been prior work on analyzing eye-tracking data from behavioral studies using HMMs. Kärrsgård and Lindholm (2003) used HMMs for an eye-typing application (in which users form words by fixating on characters on a display). More recently, Haji-Abolhassani and Clark (2013, 2014) used HMMs to predict the visual tasks being performed by subjects viewing a painting. Finally, a substantial line of work has used HMMs to study eye movement patterns involved in face recognition (Chuk, Chan, & Hsiao, 2014, 2015; Chuk, Chan, Shimojo, & Hsiao, 2016; Chuk, Crookes, Hayward, Chan, & Hsiao, 2017; Chuk, Chan, & Hsiao, 2017). A MATLAB toolbox has also been published implementing these analyses (Coutrot, Hsiao, & Chan, 2017).

These studies share several related features that contrast from the current study. First, the stimuli presented are static images. While Coutrot et al. (2017) used conversational video stimuli, the regions of interest, which were the faces of speakers, were essentially stationary relative to the display. In contrast, our stimuli are videos of moving objects, and so the parameters of our HMMs evolve over time as objects move. Second, these prior analyses are all based on first identifying fixations and then modeling these fixations using HMMs, while our HMMs directly model continuous eye-tracking data; the latter is more appropriate for measuring smooth pursuit, which is not composed of fixations. Finally, these prior studies use repetitive tasks (e.g., face recognition with aligned face stimuli) or identical tasks performed by different subjects, so that many identically distributed samples can be combined (across stimuli or across subjects) to learn a single HMM. This was important because these studies were studying *where* most humans gaze when presented with certain stimuli. In our case, object trajectories are randomly generated before each trial, and we are interested in studying broad patterns behavior, independent of the exact stimuli presented. As a result, each trial is distinct, and an HMM must be fit for each trial *using data from only that trial*. Fortunately, positions of objects of interest over time are known, and we can build an HMM around this fact.

Contributions The contributions of this work are: 1) We propose a novel HMM approach for analyzing eye-tracking data in the presence of moving visual stimuli. 2) We validate our model on data from a variant of TrackIt (called *supervised TrackIt*). 3) We apply the HMM to analyze data from the original TrackIt experiment (which we call *unsupervised TrackIt*) and show that it provides a robust analysis method.

Methods

Source Code and Reproducibility A TrackIt executable (including supervised and unsupervised variants) and its source code are freely available at http://www.psy.cmu.edu/~trackit/. Python scripts for reproducing our analyses, results, and figures, as well as the eye-tracking and TrackIt data used, are available at https://github.com/sss1/eyetracking. The eye-tracking analysis accepts a generic CSV data format containing timestamped (x,y) coordinates, making it compatible with any standard eye-tracker. Also included is a Python executable for collecting data in this format using the SMI RED-250 mobile eye tracker. Finally, videos of example unsupervised and supervised TrackIt trials can be found at https://github.com/sss1/eyetracking/tree/master/videos.

Hidden Markov Model Specification

Overview We model the subject as being, at each time point, in one of N states $\mathcal{S} = \{s_1, \ldots, s_N\}$, corresponding to the N visible objects of interest; state s_j indicates the subject tracking the j^{th} object. When in the state s_j, we model the subject's eye-tracking data with a Gaussian emission distribution centered at the center of the j^{th} object. In the case of TrackIt, if N_D denotes the number of distractors (in our stud-

ies, $N_D = 4$), $N = N_D + 1$ (1 target, N_D distractors). Figure 1 illustrates the components of our model in this context.

Notation Spatial coordinates are measured in pixels ($\approx 0.02^\circ$ of visual field) with $(0,0)$ denoting the bottom left corner of the display. $x_{min}, x_{max}, y_{min}$, and y_{max} respectively denote the minimum and maximum horizontal and vertical coordinates observable by the eye-tracker. The observable region $R := [x_{min}, x_{max}] \times [y_{min}, y_{max}]$ is a rectangle including the entire grid traversable by TrackIt objects. Within the context of any particular trial, T denotes the trial length (in $60Hz$ frames), and $t \in [T] := \{1, 2, ..., T\}$ indexes individual frames.

Hidden State Model The sequence of underlying hidden states is modeled as a Markov chain with a fixed initial distribution $\pi \in [0,1]^{\mathcal{S}}$ (such that $\sum_{S \in \mathcal{S}} \pi_S = 1$) and transition matrix $\Pi \in [0,1]^{\mathcal{S} \times \mathcal{S}}$ (such that, for each $S \in \mathcal{S}$, $\sum_{S' \in \mathcal{S}} \pi_{S,S'} = 1$). Since, in this study, we are interested in using our model to classify participants' behavioral states over time, to avoid biasing the model, π is constrained to be uniform (i.e., $\pi_{s_1} = \cdots = \pi_{s_N}$), and Π is constrained to have identical diagonal values c_1 and identical off-diagonal values c_2; i.e.,

$$\Pi = \begin{bmatrix} c_1 & c_2 & \cdots & c_2 \\ c_2 & c_1 & \cdots & c_2 \\ \vdots & \vdots & \ddots & \vdots \\ c_2 & c_2 & \cdots & c_1 \end{bmatrix}.$$

We set $c_1 = \frac{599}{600}$ and $c_2 = (1 - c_1)/N$, corresponding to an average of 1 uniformly random transition per 600 frames ($\approx 10\,\text{s}$); this choice is due to the tuning procedure used to learn the model hyperparameters (see 'Supervised TrackIt').

Emission Distributions Let $S^* : [T] \to \mathcal{S}$ denote the sequence of states assumed by the subject. At each time point, if the subject is in the state corresponding to tracking the object s, the model assumes the eye-tracking data of the subject is distributed according to an isotropic Gaussian centered at the center of S; that is, for each $t \in [T]$ and $s \in \mathcal{S}$,

$$E(t) | S^*(t) = s \sim \mathcal{N}\left(X_s(t), \sigma^2 I_2\right),$$

where $E : [T] \to R$ denotes the eye-tracker trajectory, and, for each $S \in \mathcal{S}$, $X_S : [T] \to R$ denotes the trajectory of the object corresponding to state S. The spherical standard deviation σ, which we model as common across objects, is an important hyperparameter whose selection is discussed below.

Model Fitting Because, when analyzing eye-tracking data from TrackIt, we do not *a priori* know the true state sequence S^*, the model is trained in an unsupervised manner, via maximum likelihood estimation (MLE); that is, the estimated sequence of states is that which maximizes the likelihood of the observed eye-tracking data. An HMM's MLE can be efficiently computed using the Viterbi algorithm (Forney, 1973).

Parameter Selection The most influential parameter in the model is the spherical standard deviation σ of the Gaussian emission distributions. To calibrate σ, we used data from a novel 'supervised' variant of TrackIt (described below), in which we are confident about the true state at most time points and can hence estimate model performance. We tuned σ by grid-search over 50 logarithmically spaced values of σ between 10 and 10^4 pixels ($\approx 0.2^\circ$-24° of visual field), selecting the value that maximized empirical accuracy of predicting the true state. We tuned σ separately for adults and children, as we expect tracking precision to improve with development.

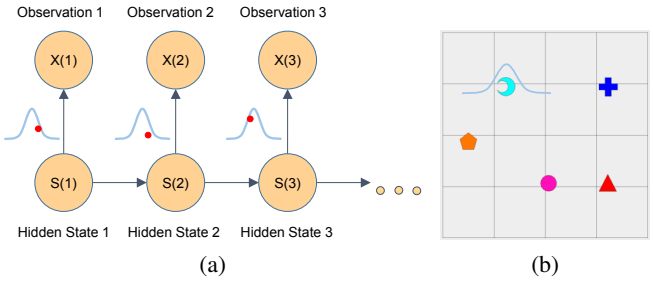


Figure 1: (a) graphical model schematic of HMM (b) example conditional distribution of $X(t)$ given $S(t) =$ "Blue Moon".

Unsupervised TrackIt

In the unsupervised (original) TrackIt task, participants visually track a single target object as it moves on a 4×4 grid, among 4 moving distractor objects. For each trial, the target and distractor objects are randomly selected without replacement from a set of unique objects spanning 9 different shapes with 9 different color possibilities (81 possible objects). See Figure 2 for an example. At the beginning of each trial, objects appear on a grid, centered in random, distinct grid cells, and the target object is indicated by a red circle around it.

Upon starting the trial (by button press), the red circle disappears, and the objects begin to move in piecewise-linear trajectories from grid cell to grid cell at a constant speed (500 pixels, or 10°, per second). At the end of each trial, all objects vanish, and the participant is asked to indicate the grid cell the target object last occupied before disappearing.

The path of each object is randomized, with the constraint that the target has to be in the center of a grid cell at the end of the trial, to reduce ambiguity for the participant in determining its final location. Due to this constraint, trial length is not fixed, but varies slightly between trials (to allow the target to reach the center of a grid cell), with a minimum of 10 s.

Grid size, object speed, number of distractors, minimum trial length, etc. are experimenter-selected TrackIt parameters; the above values were suggested by prior work as appropriate for young children (Kim et al., 2017).

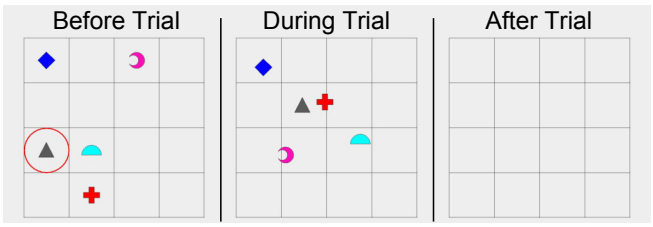


Figure 2: The unsupervised TrackIt task.

Supervised TrackIt

To tune the parameter σ and evaluate model performance, we designed a 'supervised' variant of TrackIt, in which we know, with relatively high confidence, what object the participant is looking at (i.e., the 'true state') at most time points. To do this, we made the target flash white repeatedly (for 100 ms, separated by 200 ms) during the entire trial, making it salient and easy to track (without relying on endogenous SSA). Participants were instructed to follow the flashing object with their eyes. Rather than using a single target for the entire trial, the flashing target changed at random intervals (uniformly between 5 s and 15 s). To allow multiple target changes, trials were lengthened to a minimum of 30 s (from 10 s in unsupervised TrackIt). Changing the target within trials was essential to ensure the fitted model could accurately detect transitions between objects; without this, the model would learn to always estimate a single most likely target during each trial (i.e., the selected σ would be too large). As in unsupervised TrackIt, the target was circled in red and flashed before trial start, so participants could begin the trial tracking the correct object. Other parameters and preprocessing steps were identical to the unsupervised TrackIt setup. TrackIt recorded the flashing target's identity in each frame, allowing us to compare model predictions to this 'ground truth'. Some error is introduced by the delay with which participants transition after the blinking object changes. Better results might be obtained by ignoring a few frames after each change when measuring error, but our results are robust without doing this.

Experimental Procedure and Data

Subjects For supervised TrackIt, 15 healthy adult volunteers and 15 typically developing 5-year-olds each performed 12 trials, including 2 initial practice trials during which the experimenter explained the task. Practice trials were not analyzed, giving 10 usable trials/subject. For unsupervised TrackIt, 10 healthy adult volunteers each performed 5 trials and 10 typically developing 3-year-olds each performed 3.

Materials and Apparatus Stimuli were presented on a Lenovo laptop screen with physical dimensions 19.1 cm $\times$ 34.2 cm and pixel dimensions 1080×1920 pixels (approximately $22^\circ \times 40^\circ$ of visual field). Subjects were seated at a desk facing the screen with their heads about 0.5 m away from the screen. The SMI RED-250 mobile eye tracker was used to record continuous gaze positions at 60 Hz during TrackIt trials. After using SMI's iView X software to calibrate the eye-tracker, we used a custom Python script to collect eye-tracking data synchronized with TrackIt.

Data Preprocessing Child eye-tracking data contains a large proportion of missing values (due to children looking away from task or moving excessively), and so we preprocessed data to mitigate this. Whenever a short interval of at most ≤ 10 consecutive frames (≈ 16.7 ms) of eye-tracking data was missing, we linearly interpolated gaze during those frames from non-missing data immediately before and after that interval. Next, we discarded trials missing eye-tracking data for more than 50% of frames (53 child trials and 5 adult trials). Finally, we discarded data from subjects for whom more than 50% (> 5 trials) had been discarded (3 children). After preprocessing, 86 child trials and 145 adult trials remained. Even after these steps, intervals of (> 10 frames of) eye-tracking data may still be missing. For these frames, the HMM automatically assigns a 'null' state, and the frames before and after each such interval are fit independently by the Viterbi algorithm. When evaluating model performance, we report results both treating these frames as incorrect classifications (giving a conservative 'worst-case' lower bound on performance) and ignoring these frames (giving a less conservative 'average-case' performance estimate).

Results

Model Validation (Supervised Data)

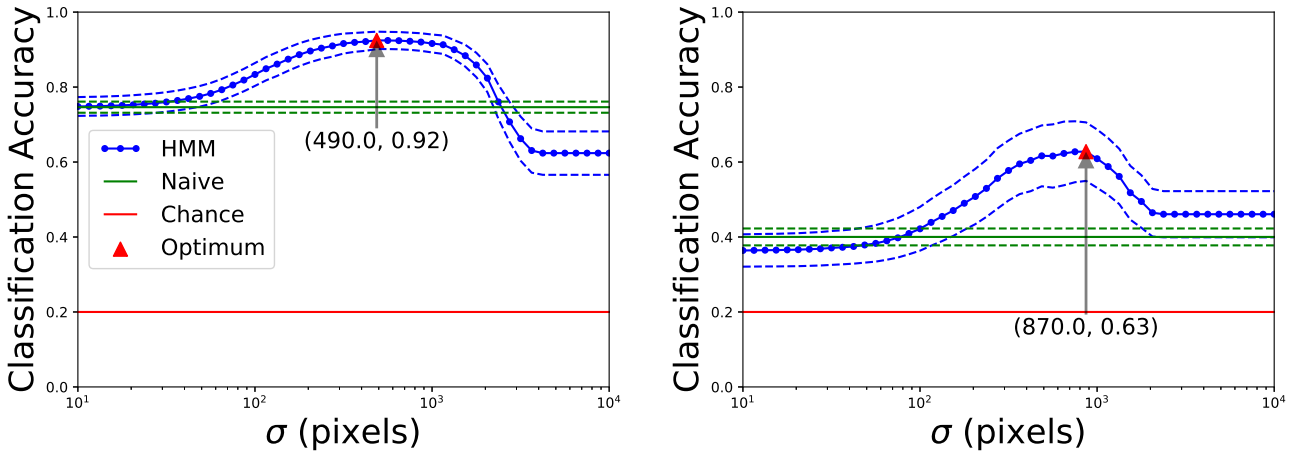
We compared our HMM's performance to that of a 'naive' model that assumed that, at each time point, the subject was looking at the object closest to their gaze. This model is equivalent to a variant of our HMM with uniform transition matrix Π, thus ignoring the underlying Markov model and using only emission probabilities. Figure 3 shows the HMM's accuracy, as a function of σ, along with that of the naive model and 'chance' of 20% (1 out of 5 total objects).

While both models perform better on adult data than on child data, curves are qualitatively similar for both populations. For very small σ (e.g., < 100 ($\approx 2^\circ$)), the cost of selecting an object even slightly further than the closest object outweighs the cost of transitioning states, and so the HMM behaves essentially like the naive model. For very large σ (e.g., > 2000 ($\approx 49^\circ$)), the emission distributions of all objects become similar, and the HMM may fail to ever transition, performing worse than the naive model. As we expected, the optimal σ for children was much larger than that for adults (870 pixels ($\approx 18^\circ$) versus 490 pixels ($\approx 10^\circ$)), reflecting less precise visual tracking of the target object. However, for both adults and children, in the large range of $\sigma \in [10^2, 10^3]$ ($\approx 2^\circ$-21°), the HMM outperforms the naive model.

This analysis shows that superiority of the HMM decoder depends on the value of σ, albeit quite robustly. Hence, to objectively evaluate decoder performance independently of tuning, we next used leave-one-out cross-validation (holding out 1 trial per fold, maximizing accuracy over σ on remaining trials, and measuring accuracy on the held-out trial). Table 1 shows that the HMM provides a large mean improvement ($\geq$ 16.1% in adults, $\geq$ 20.9% in children) over the naive model.

SSA Performance Evaluation (Unsupervised Data)

We next applied our HMM and the naive model to data from the original unsupervised TrackIt experiment, this time with the goal of measuring subject performance (rather than model performance). As shown in Table 2, task performance as scored by the HMM is far higher than that as scored by the naive model, and this difference was statistically significant



(a) Adult

(b) Child

Figure 3: Semi-log plot of HMM, naive, and chance accuracies as functions of HMM parameter σ. Dashed lines indicate bootstrapped 95% confidence bands. The point of optimal HMM performance (our suggested value of σ) is indicated by a triangle. Only accuracies on non-missing frames are shown, but curves computed using all frames were qualitatively similar.

Table 1: Proportion of supervised frames correctly classified.

Population	HMM (95% CI)	Naive (95% CI)
All frames		
Adult	91.4%(2.7%)	75.3%(2.5%)
Child	52.7%(3.9%)	31.8%(2.3%)
Non-missing/interpolated frames only		
Adult	93.5%(1.3%)	76.8%(1.5%)
Child	60.7%(2.2%)	36.8%(2.1%)

Table 2: Proportion unsupervised frames classified on target.

Population	Condition	HMM (95% CI)	Naive (95% CI)
All frames			
Adult	All Same	85.4%(5.4%)	62.9%(6.4%)
Adult	All Diff	90.7%(4.1%)	65.4%(5.4%)
Child	All Same	30.1%(6.0%)	19.3%(3.9%)
Child	All Diff	21.8%(6.3%)	13.6%(3.7%)
Non-missing/interpolated frames only			
Adult	All Same	92.9%(3.4%)	68.4%(4.3%)
Adult	All Diff	97.4%(1.4%)	70.2%(3.0%)
Child	All Same	48.8%(7.6%)	33.3%(4.9%)
Child	All Diff	37.3%(7.0%)	26.5%(3.4%)

($p < 0.05$) in all conditions except in the All Different condition in children. This suggests that the naive model significantly underestimates task performance, losing signal that may be important for downstream data analyses.

While the main contrast (child performance between All Same and All Different conditions) was not statistically significant, the direction of difference is consistent with our hypothesis that 3-year-olds have more limited endogenous control of SSA. This dataset was quite small, and we believe that collecting a larger dataset would show this contrast conclusively. Further work will also explore performance trends (over trial duration) from the output of the HMM model.

Conclusions & Future Directions

This paper proposed a novel analysis using a hidden Markov model for eye-tracking data in the presence of multiple moving objects. We validated and tuned this model in a novel supervised object-tracking task, demonstrating robustness to hyperparameter choices, and we used the model to analyze data from the TrackIt task for measuring SSA in children. The HMM collapses noisy spatiotemporal eye-tracking data into a sequence of a small number of states, simultaneously denoising the data and making it more behaviorally interpretable. The model is quite flexible; input data can be from any visual stimulus with moving objects or areas of interest, and many analyses can be performed on its output. For example, while we only studied the proportion of time spent on target, the HMM can also identify *when* during the trial children are distracted, and, when, if at all, children are able to return attention to the target, allowing us to study the time course of SSA and its possible self-regulatory mechanisms.

The main constraint of the proposed method is that it requires knowing positions of objects of interest. While available for artificially-generated stimuli, these may be difficult to obtain in studies that are not computer-based or use videos of natural scenes. A solution may be to combine an HMM with algorithms for automated object detection in video, which are becoming widely-available (Huang et al., 2017).

Towards a Cognitive Model of Object Tracking Our decoder is based on a generative model of eye-tracking data. This model may be a suggestive first step towards linking eye-tracking data to the cognitive process of visual object tracking, and, perhaps, to the higher-level construct of visual SSA. Using such a model to study subject performance during task (as in this study) requires fixing the HMM with uniform initial and transition probabilities, so that the model does not

intrinsically prefer some states over others (e.g., in the case of TrackIt, the model should treat the target identically to the other objects). Conversely, a realistic cognitive model should have non-uniform probabilities (e.g., preferring to follow the target over distractors, by virtue of SSA). Hence, a major step in developing such a cognitive model would be fitting its parameters to behavioral data. For Gaussian HMMs, this can be done using expectation maximization (Bilmes et al., 1998), which we suggest as a fruitful direction for future work.

Acknowledgments

We thank Melissa Pocsai for help collecting data. We thank children, parents and teachers of CMU Children's School for making this work possible. This work was supported by National Science Foundation grant BCS-1451706 to AVF and EDT and graduate research fellowship DGE-1252522 to SS.

References

Barr, D. J. (2008). Analyzing visual world eyetracking data using multilevel logistic regression. *Journal of memory and language*, *59*(4), 457–474.

Bilmes, J. A., et al. (1998). A gentle tutorial of the EM algorithm and its application to parameter estimation for gaussian mixture and hidden Markov models. *International Computer Science Institute*, *4*(510), 126.

Biondi, J., Fernandez, G., Castro, S., & Agamenonni, O. (2017). Eye-movement behavior identification for ad diagnosis. *arXiv preprint arXiv:1702.00837*.

Cassin, B., Solomon, S., & Rubin, M. L. (1984). *Dictionary of eye terminology*. Triad Pub. Co.

Chuk, T., Chan, A., & Hsiao, J. (2015). Hidden Markov model analysis reveals better eye movement strategies in face recognition. In *Proc. of the Cognitive Science Society.*

Chuk, T., Chan, A. B., & Hsiao, J. H. (2014). Understanding eye movements in face recognition using hidden Markov models. *Journal of vision*, *14*(11), 8–8.

Chuk, T., Chan, A. B., & Hsiao, J. H. (2017). Is having similar eye movement patterns during face learning and recognition beneficial for recognition performance? evidence from hidden Markov modeling. *Vision Research*.

Chuk, T., Chan, A. B., Shimojo, S., & Hsiao, J. (2016). Mind reading: Discovering individual preferences from eye movements using switching hidden Markov models. In *Proceedings of the Cognitive Science Society.*

Chuk, T., Crookes, K., Hayward, W. G., Chan, A. B., & Hsiao, J. H. (2017). Hidden Markov model analysis reveals the advantage of analytic eye movement patterns in face recognition across cultures. *Cognition*, *169*, 102–117.

Coutrot, A., Hsiao, J. H., & Chan, A. B. (2017). Scanpath modeling and classification with hidden Markov models. *Behavior Research Methods*, 1–18.

Dink, J., & Ferguson, B. (2015). *eyetrackingR: An R library for eye-tracking data analysis.*

Doran, M., Hoffman, J., & Scholl, B. (2009). The role of eye fixations in concentration and amplification effects during multiple object tracking. *Visual Cognition*, *17*(4), 574.

Duchowski, A. T. (2017). *Eye tracking methodology: Theory and practice*. Springer.

Fisher, A. V., Thiessen, E., Godwin, K., Kloos, H., & Dickerson, J. (2013). Assessing selective sustained attention in 3-to 5-year-old children: Evidence from a new paradigm. *J of Experimental Child Psychology*, *114*(2), 275–294.

Fisher, A. V., Thiessen, E. D., Dickerson, J. P., & Erickson, L. C. (2013). Development of selective sustained attention: Conceptual and measurement issues. In *Biennial meeting of the cognitive development society (cds).*

Forney, G. D. (1973). The Viterbi algorithm. *Proceedings of the IEEE*, *61*(3), 268–278.

Gegenfurtner, A., Lehtinen, E., & Säljö, R. (2011). Expertise differences in the comprehension of visualizations: A meta-analysis of eye-tracking research in professional domains. *Educational Psychology Review*, *23*(4), 523–552.

Haji-Abolhassani, A., & Clark, J. (2013). A computational model for task inference in visual search. *Journal of vision*.

Haji-Abolhassani, A., & Clark, J. J. (2014). An inverse yarbus process: Predicting observers task from eye movement patterns. *Vision research*, *103*, 127–142.

Holzman, P. S., Proctor, L. R., Levy, D. L., Yasillo, N. J., Meltzer, H. Y., & Hurt, S. W. (1974). Eye-tracking dysfunctions in schizophrenic patients and their relatives. *Archives of general psychiatry*, *31*(2), 143–151.

Huang, J., Rathod, V., Sun, C., Zhu, M., Korattikara, A., Fathi, A., ... others (2017). Speed/accuracy trade-offs for modern convolutional object detectors. In *IEEE CVPR.*

Jacob, R., & Karn, K. S. (2003). Eye tracking in human-computer interaction and usability research: Ready to deliver the promises. *Mind*, *2*(3), 4.

Kärrsgård, I., & Lindholm, A. (2003). *Eye movement tracking using hidden Markov models*. Chalmers tek. högsk.

Kim, J., Vande Velde, A., Thiessen, E. D., & Fisher, A. V. (2017). Variables involved in selective sustained attention development: Advances in measurement. In *Proceedings of the 39th annual conf. of the Cognitive Science Society.*

O'Connor, C., Manly, T., Robertson, I., Hevenor, S., & Levine, B. (2004). An fMRI study of sustained attention with endogenous and exogenous engagement. *Brain and Cognition*, *54*(2), 113–135.

Palinko, O., Kun, A. L., Shyrokov, A., & Heeman, P. (2010). Estimating cognitive load using remote eye tracking in a driving simulator. In *Proceedings of the 2010 symposium on eye-tracking research & applications* (pp. 141–144).

Pyykkönen, P., Hyönä, J., & van Gompel, R. P. (2009). Activating gender stereotypes during online spoken language processing. *Experimental Psychology*.

Rehder, B., & Hoffman, A. (2005). Eyetracking and selective attention in category learning. *Cognitive Psych.*, *51*(1).

Smuc, M., Mayr, E., & Windhager, F. (2010). The game lies in the eye of the beholder: The influence of expertise on watching soccer. In *Proc. of the Cognitive Science Society.*