Bimodal Multi-Object Localisation, Siteswap Inference, and Analysis for Competitive Juggling

James M. Cozens
Signal Processing and Communications Laboratory
University of Cambridge Engineering Department
Cambridge, UK
jmc257@cam.ac.uk

Simon J. Godsill
Signal Processing and Communications Laboratory
University of Cambridge Engineering Department
Cambridge, UK
sjg30@cam.ac.uk

***Abstract*—This paper presents an adaptive approach to real-time multi-object localisation in addition to Siteswap inference, and performance evaluation metrics for juggling routines, employing a proposed bimodal machine learning-enhanced state-space model implementation. Considering the complex multi-modal characteristics exhibited by objects during performances, the paper introduces a bespoke Interacting Multiple Model (IMM) component for increased Siteswap beat detection accuracy and gravitational acceleration inference, and a scheme for causal Siteswap inference derived through machine learning-enhanced IMM mode outputs. The algorithm effectively models the transitory behaviour of the system, enabling rapid and smooth transitions between the two discrete tracking cases (airborne, and caught) and accurate Siteswap inference under a variety of camera and environmental conditions. The employment of beat tracking algorithms that exploit optimal compromises in time domain onset detection functions and Tempograms, enables effective error correction of Siteswap detections, in addition to providing performance analysis and visualisation utilities. Experimentally, the algorithm is capable of object tracking and Siteswap inference with up to 11 objects for a variety of challenging Siteswaps and conditions, serving as a versatile performance analysis, evaluation, and visualisation utility.**

***Index Terms*—Multi-object tracking (MOT), Adaptive Bimodal Kalman Filter, Interacting Multiple Model (IMM), Audio Beat Tracking Algorithms, Time Domain Onset Detection Functions, Tempogram, Machine Learning in Object Tracking, Siteswap Juggling, Sport Performance Analysis.**

I. Introduction

A. Objectives

Juggling is an ancient art, with references to juggling found in writings from ancient China, Ireland, Rome, and Egypt [12]. Over the centuries, juggling has evolved into an immensely diverse and popular art form, with sport juggling featuring as one of the latest notable additions to the art within the past couple of decades. With the potential inclusion of Sport Juggling in the Olympics on the horizon [1], [2], [21], one of the most sought-after technologies within the juggling community is multi-object tracking software, both for performance analysis and as a visualisation utility. Tracking provides an unparalleled perspective on the technique of a performer and offers the capacity to augment performances with elements such as real-time visualisations. In addition, considering the popularity and extensive use of the Siteswap notation system (see II-G1) among jugglers for routines, particularly in competitive scenarios [22] such as in the World Juggling Federation (WJF) Competitions [21], Siteswap inference is of paramount importance both for performance analysis and competitive evaluation. The primary objective is thus to provide the community with a generalised real-time (causal) and non-real-time (non-causal) multi-object tracker capable of accurately inferring Siteswap sequences, catch ('beat') positions, various tracked spatial derivatives, and performance evaluation metrics for a variety of challenging environmental conditions. Currently, the software has been used by various World Record holding jugglers as a performance analysis utility [3]; a playlist of tracked results is available here[1], and the code repository here[2].

B. Overview

Methods exist in literature to extract multiple objects from video, in particular within gravitationally-influenced environments. One such approach employs the Expectation Maximisation algorithm with respect to parabolas extracted from inputted video [7], [10], and almost all employ Kalman filters [19]. The issue currently is that techniques that extract parabolic arcs are optimal for uniformly distributed patterns, such as a cascade or fountain (which involve throwing all the balls at the same height), however, for more complex routines that employ concepts such as Siteswaps, given non-parabolic trajectories from the hand cannot be extracted, objects can be easily lost in the frame during data association. Additionally, juggling typically involves a period of highly predictable motion, whilst airborne, followed by a short period of highly unpredictable motion (whilst being accelerated in the hand). Given the unique challenges facing Siteswap detection, especially in causal contexts, no literature currently exists for the direct inference of Siteswap sequences from video. An unpublished project by A. Alonso [4] explores Siteswap detection for common Siteswaps for 3-6 balls for a non-causal application, employing a conventional mono-modal Kalman filter, with beat positions extracted through

JMC is funded by the EPSRC Doctoral Training Partnership (DTP) (Grant Number: 2738835)

[1]https://youtube.com/playlist?list=PLWtN-KJsWmthw4U7PDDyAyeDx_qVGlbNt&si=OPAJtT5oymRFwO5B

[2]https://github.com/Jamesmcozens/Bimodal-Multi-Object-Localisation-Siteswap-Inference-and-Analysis-for-Competitive-Juggling

peak detection in positional components; 77.27% accuracy is reported, for specific environmental conditions. Given the complex bimodal motion characteristics of juggling routines previously described and the consequent susceptibility of conventional single-mode trackers to multi-modal transitional behaviour [11], this paper alternatively proposes the use of an adaptive Interacting Multiple Model (IMM) filter facilitated by beat tracking algorithms, with outputs augmented through a learning model trained on tracked data to provide Siteswap detection, for both non-causal and real-time applications.

II. Methodology

A. Object Segmentation

Objects (balls) can be detected via a variety of standard techniques, such as through the employment of neural networks [13], hough circle analysis [17], or blob detection, achieved through the usage of image pyramids and difference of gaussians [18]. However, although robust in general, partial occlusion from the hand renders these methods vulnerable, especially considering during occlusion the balls are expected to experience the least predictable motion. This can be mitigated through the employment of image segmentation through colour filtering in addition to background subtraction using Mixture of Gaussian Models (MOGs) [5], given juggling objects are typically contrasting in colour and fast-moving compared to background objects. It is assumed that the camera is aligned vertically, and the performer plane is assumed to be parallel with respect to the camera plane.

B. State Space Model

1) Overview: A key proposed component of this method is that the trajectory behaviour of the objects can be modelled bimodally. The first mode is characterised by airborne motion, with no external forces, except for a gravitational acceleration a in the y axis, which can be assumed to be Gaussian distributed normally prior to observation in pixel coordinates:

$$a \sim \mathcal{N}(a_i,\, \sigma_a^2), \tag{1}$$

where a_i and σ_a^2 are the gravitational acceleration prior mean and variance respectively. The second mode is characterised by highly unpredictable motion, due to the presence of irregular external forces (during a catch), and thus these two distinct cases must be considered carefully. To account for unknown variations in depth relative to the camera, this paper proposes inferring the "perceived" gravitational acceleration in the camera's frame of reference as a function of time, referred to simply as the "gravitational acceleration" throughout. Thus, the following state-space is proposed:

$$\mathbf{x}^{(m)} = \begin{bmatrix} x^{(m)} & y^{(m)} & \dot{x}^{(m)} & \dot{y}^{(m)} & \ddot{x}^{(m)} & \ddot{y}^{(m)} & a^{(m)} \end{bmatrix}^T,$$

where $\mathbf{x}^{(m)}$ is the state vector for object m ($1 \leq m \leq M$). The state space variables correspond to $\{x^{(m)}\}_{m=1}^{M}$, the horizontal pixel coordinates, $\{y^{(m)}\}_{m=1}^{M}$, the vertical components, the gravitational acceleration, $a^{(m)}$, and the two spatial derivatives for each respective component, $\{\dot{x}^{(m)}\}_{m=1}^{M}$, $\{\dot{y}^{(m)}\}_{m=1}^{M}$ and $\{\ddot{x}^{(m)}\}_{m=1}^{M}$, $\{\ddot{y}^{(m)}\}_{m=1}^{M}$. The dynamical model for state $\mathbf{X}$ in model j ($j \in \{1, 2\}$), is assumed to follow a Stochastic Differential Equation (SDE) [15]:

$$\dot{\mathbf{X}}(t) = \mathbf{A}_j \mathbf{X}(t) + \mathbf{B}_j \mathbf{W}_j(t),$$

where $\mathbf{A}_j$ is the system matrix for mode j, $\mathbf{B}_j$ is the noise matrix, and $\mathbf{W}_j$ is the noise vector.

2) Mode 1 (Airborne): The airborne mode is characterised by near-constant velocity in the x component, and vertically-aligned gravitational acceleration in the y component:

$$\mathbf{A}_1 = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \mathbf{B}_1 = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ \sigma_{\ddot{x}} & 0 \\ 0 & \sigma_{\ddot{y}} \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix},$$

$$\mathbf{W}_1 = \begin{bmatrix} w_{\ddot{x}}(t) & w_{\ddot{y}}(t) \end{bmatrix}^T,$$

where $w_{\ddot{x}}(t)$ and $w_{\ddot{y}}(t)$ are independent and assumed to be standard Gaussian white noise processes such that $\frac{\partial^2 x}{\partial t^2} = w_{\ddot{x}}(t)$ and $\frac{\partial^2 y}{\partial t^2} = a + w_{\ddot{y}}(t)$; thus, the velocity components are subjected to the time integral of the noise processes, which are standard Brownian motion. Therefore, we may discretise the system over time interval T to $t = T + \Delta t$ (where Δt is the time step for each frame), by solving the Stochastic Differential Equation (SDE) [15] as:

$$\mathbf{X}(T + \Delta t) = \exp\left(\mathbf{A}_1 \Delta t\right)\mathbf{X}(T) + \int_T^{T+\Delta t} \exp\left(\mathbf{A}_1 \tau\right)\mathbf{B}_1 \mathbf{W}_1 \left(T + \Delta t - \tau\right)\, d\tau.$$

Letting $\mathbf{X}_{n+1} = \mathbf{X}(T + \Delta t)$ and $\mathbf{X}_n = \mathbf{X}(T)$, the discrete time model is obtained as $\mathbf{X}_{n+1} = \mathbf{F}_1 \mathbf{X}_n + \mathbf{w}_1$, where:

$$\mathbf{F}_1 = \exp\left(\mathbf{A}_1 \Delta t\right) = \begin{bmatrix} 1 & 0 & \Delta t & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & \Delta t & 0 & 0 & \frac{\Delta t^2}{2} \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & \Delta t \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\mathbf{w}_1 = \int_T^{T+\Delta t} \exp\left(\mathbf{A}_1 \tau\right)\mathbf{B}_1 \mathbf{W}_1 \left(T + \Delta t - \tau\right)\, d\tau.$$

Thus, $\mathbf{X}_{n+1}$ is Gaussian, $\mathbf{X}_{n+1} \sim \mathcal{N}\left(\mathbf{F}_1 \mathbf{X}_n,\, \mathbf{Q_1}\right)$, where $\mathbf{Q_1}$ is the process noise matrix:

$$\mathbf{Q}_1 = \mathbb{E}\left[\mathbf{w}_1 \mathbf{w}_1^T\right]$$

$$= \mathbb{E}\left[\int_0^{\Delta t}\int_0^{\Delta t} \begin{bmatrix} (\Delta t - u)\sigma_{\ddot{x}} w_{\ddot{x}}(u) \\ (\Delta t - u)\sigma_{\ddot{y}} w_{\ddot{y}}(u) \\ \sigma_{\ddot{x}} w_{\ddot{x}}(u) \\ \sigma_{\ddot{y}} w_{\ddot{y}}(u) \\ 0 \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} (\Delta t - v)\sigma_{\ddot{x}} w_{\ddot{x}}(v) \\ (\Delta t - v)\sigma_{\ddot{y}} w_{\ddot{y}}(v) \\ \sigma_{\ddot{x}} w_{\ddot{x}}(v) \\ \sigma_{\ddot{y}} w_{\ddot{y}}(v) \\ 0 \\ 0 \\ 0 \end{bmatrix}^T du\, dv\right]$$

$$= \begin{bmatrix} \frac{\Delta t^3}{3}\sigma_{\ddot{x}}^2 & 0 & \frac{\Delta t^2}{2}\sigma_{\ddot{x}}^2 & 0 & 0 & 0 & 0 \\ 0 & \frac{\Delta t^3}{3}\sigma_{\ddot{y}}^2 & 0 & \frac{\Delta t^2}{2}\sigma_{\ddot{y}}^2 & 0 & 0 & 0 \\ \frac{\Delta t^2}{2}\sigma_{\ddot{x}}^2 & 0 & \sigma_{\ddot{x}}^2 & 0 & 0 & 0 & 0 \\ 0 & \frac{\Delta t^2}{2}\sigma_{\ddot{y}}^2 & 0 & \sigma_{\ddot{y}}^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

3) Mode 2 (In the hand): To determine suitable state space transition equations for the highly unpredictable mode, motion can modelled via a constant acceleration model, modelling jerk as a white noise process, $\frac{\partial^3 x}{\partial t^3} = w_{\dddot{x}}(t)$, $\frac{\partial^3 y}{\partial t^3} = w_{\dddot{y}}(t)$, where once again two noise processes are the derivative of standard independent Brownian motions. Thus the state transition matrices for this mode are:

$$\mathbf{A}_2 = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \mathbf{B}_2 = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ \sigma_{\dddot{x}} & 0 \\ 0 & \sigma_{\dddot{y}} \\ 0 & 0 \end{bmatrix},$$

$$\mathbf{W}_2 = \begin{bmatrix} w_{\dddot{x}}(t) & w_{\dddot{y}}(t) \end{bmatrix}^T.$$

Then, similarly deriving the solution for $\mathbf{X}_{n+1}$ in this second mode provides $\mathbf{X}_{n+1} = \mathbf{F}_2\mathbf{X}_n + \mathbf{w}_2$, where:

$$\mathbf{F}_2 = \begin{bmatrix} 1 & 0 & \Delta t & 0 & \frac{\Delta t^2}{2} & 0 & 0 \\ 0 & 1 & 0 & \Delta t & 0 & \frac{\Delta t^2}{2} & 0 \\ 0 & 0 & 1 & 0 & \Delta t & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & \Delta t & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\mathbf{w}_2 = \int_T^{T+\Delta t} \exp\left(\mathbf{A}_2\tau\right)\mathbf{B}_2\mathbf{W}_2\left(T + \Delta t - \tau\right)\, d\tau.$$

As before, $\mathbf{X}_{n+1}$ is Gaussian: $\mathbf{X}_{n+1} \sim \mathcal{N}\left(\mathbf{F}_2\mathbf{X}_n,\, \mathbf{Q_2}\right)$, where $\mathbf{Q}_2$ can similarly be derived:

$$\mathbf{Q}_2 = \mathbb{E}\left[\mathbf{w}_2\mathbf{w}_2^T\right]$$

$$= \begin{bmatrix} \frac{\Delta t^5}{20}\sigma_{\ddot{x}}^2 & 0 & \frac{\Delta t^4}{8}\sigma_{\ddot{x}}^2 & 0 & \frac{\Delta t^3}{6}\sigma_{\ddot{x}}^2 & 0 & 0 \\ 0 & \frac{\Delta t^5}{20}\sigma_{\ddot{y}}^2 & 0 & \frac{\Delta t^4}{8}\sigma_{\ddot{y}}^2 & 0 & \frac{\Delta t^3}{6}\sigma_{\ddot{y}}^2 & 0 \\ \frac{\Delta t^4}{8}\sigma_{\ddot{x}}^2 & 0 & \frac{\Delta t^3}{3}\sigma_{\ddot{x}}^2 & 0 & \frac{\Delta t^2}{2}\sigma_{\ddot{x}}^2 & 0 & 0 \\ 0 & \frac{\Delta t^4}{8}\sigma_{\ddot{y}}^2 & 0 & \frac{\Delta t^3}{3}\sigma_{\ddot{y}}^2 & 0 & \frac{\Delta t^2}{2}\sigma_{\ddot{y}}^2 & 0 \\ \frac{\Delta t^3}{6}\sigma_{\ddot{x}}^2 & 0 & \frac{\Delta t^2}{2}\sigma_{\ddot{x}}^2 & 0 & \Delta t\sigma_{\ddot{x}}^2 & 0 & 0 \\ 0 & \frac{\Delta t^3}{6}\sigma_{\ddot{y}}^2 & 0 & \frac{\Delta t^2}{2}\sigma_{\ddot{y}}^2 & 0 & \Delta t\sigma_{\ddot{y}}^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

The observations, $\mathbf{z}_n$, are defined for mode i:

$$\mathbf{z}_n = \mathbf{H}\mathbf{x}_n + \mathbf{v}_i,$$

where $\mathbf{v}_i$ are the observation noise vectors and $\mathbf{H}$ is the observation matrix. Given that only positional observations are made (in addition to gravitational observations, as described in section II-D), $\mathbf{v}_i$ and $\mathbf{H}$ are defined:

$$\mathbf{v}_1 = \begin{bmatrix} \sigma_x v_x & \sigma_y v_y & \sigma_1 v_a \end{bmatrix}^T, \mathbf{v}_2 = \begin{bmatrix} \sigma_x v_x & \sigma_y v_y & \sigma_2 v_h \end{bmatrix}^T,$$

$$\mathbf{H} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix},$$

where v_x, v_y, v_a, and v_h are Gaussian white noise processes. $\sigma_1 \ll \sigma_2$, such that the system has increased confidence in gravitational observations when airborne relative to being in the hand. Thus, the process is Gaussian, $\mathbf{z}_n \sim \mathcal{N}\left(\mathbf{H}\mathbf{x}_n,\, \mathbf{R}_i\right)$, where $\mathbf{R}_i$ are the observation noise covariance matrices:

$$\mathbf{R}_1 = \mathbb{E}\left[\mathbf{v}_1\mathbf{v}_1^T\right] = \begin{bmatrix} \sigma_x^2 & 0 & 0 \\ 0 & \sigma_y^2 & 0 \\ 0 & 0 & \sigma_1^2 \end{bmatrix} \mathbf{R}_2 = \begin{bmatrix} \sigma_x^2 & 0 & 0 \\ 0 & \sigma_y^2 & 0 \\ 0 & 0 & \sigma_2^2 \end{bmatrix}.$$

C. IMM Update Equations

To appropriately account for the bimodality of the system, the scheme proposes an adaptive IMM implementation, comprising M parallel IMM Kalman filters [9], [11]. For each time step, the following steps are performed for each ball m and mode i:

$$\mu_{n-1}^{(j|i,m)} = \frac{\Pi_{ij}\mu_{n-1}^{(j,m)}}{\sum_{l=1}^{2}\Pi_{il}\mu_{n-1}^{(l,m)}}$$

$$\bar{\mathbf{x}}_{n-1|n-1}^{(i,m)} = \sum_{j=1}^{2}\hat{\mathbf{x}}_{n-1|n-1}^{(j,m)}\mu_{n-1}^{(j|i,m)}$$

$$\bar{\mathbf{p}}_{n-1|n-1}^{(i,m)} = \sum_{j=1}^{2}\mu_{n-1}^{(j|i,m)}\mathbf{p}_{n-1|n-1}^{(j,m)} + \left(\mathbf{x}_{n-1|n-1}^{(i,m)} - \hat{\mathbf{x}}_{n-1|n-1}^{(i,m)}\right)\left(\mathbf{x}_{n-1|n-1}^{(j,m)} - \hat{\mathbf{x}}_{n-1|n-1}^{(i,m)}\right)^T,$$

where $\mathbf{\Pi} = \{\Pi_{i,j}\}_{i=1,j=1}^{2,2}$ is the mode transition matrix, corresponding to the probability of mode j at time step n given mode i at $n-1$. $\mu_{n-1}^{(j|i,m)}$ is the conditional mixing probability that mode i accurately described the system at $n-1$ given mode j provides the accurate description at n. $\mu_n^{(i,m)}$ is the posterior mode mixing probability of mode i at n given the observations, and $\mathbf{p}_{n|n}^{(i,m)}$ are the covariance matrices. The state projection step involves:

$$\begin{aligned} \hat{\mathbf{x}}_{n|n-1}^{(i,m)} &= \mathbf{A}_i\bar{\mathbf{x}}_{n-1|n-1}^{(i,m)} \\ \mathbf{p}_{n|n-1}^{(i,m)} &= \mathbf{A}_i\bar{\mathbf{p}}_{n-1|n-1}^{(i,m)}\left(\mathbf{A}_i\right)^T + \mathbf{Q}_i \\ \mathbf{S}_n^{(i,m)} &= \mathbf{R}_i + \mathbf{H}\mathbf{p}_{n|n-1}^{(i,m)}\left(\mathbf{H}\right)^T \\ \mathbf{K}_n^{(i,m)} &= \mathbf{p}_{n|n-1}^{(i,m)}\left(\mathbf{H}\right)^T\left(\mathbf{S}_n^{(i,m)}\right)^{-1} \\ \hat{\mathbf{x}}_{n|n}^{(i,m)} &= \hat{\mathbf{x}}_{n|n-1}^{(i,m)} + \mathbf{K}_n^{(i,m)}\left(\mathbf{z}_n^{(m)} - \mathbf{H}\hat{\mathbf{x}}_{n|n-1}^{(i,m)}\right) \\ \mathbf{p}_{n|n}^{(i,m)} &= \mathbf{p}_{n|n-1}^{(i,m)} - \mathbf{K}_n^{(i,m)}\mathbf{H}\mathbf{p}_{n|n-1}^{(i,m)} \end{aligned}$$

where $\mathbf{S}_n^{(i,m)}$ and $\mathbf{K}_n^{(i,m)}$ are the innovation covariance and Kalman gain matrices respectively. The mode mixing probabilities are subsequently updated as such:

$$\mu_n^{(i,m)} = \frac{\mathcal{N}\left(\mathbf{z}_n^{(m)} - \mathbf{H}\hat{\mathbf{x}}_{n|n-1}^{(i,m)},\, \mathbf{S}_n^{(i,m)}\right)\sum_{l=1}^{2}\Pi_{il}\mu_{n-1}^{(l,m)}}{\sum_{j=1}^{2}\mathcal{N}\left(\mathbf{z}_n^{(m)} - \mathbf{H}\hat{\mathbf{x}}_{n|n-1}^{(j,m)},\, \mathbf{S}_n^{(j,m)}\right)\sum_{l=1}^{2}\Pi_{jl}\mu_{n-1}^{(l,m)}}$$

Final state estimation is then performed on the two modes for the M parallel filters:

$$\hat{\mathbf{x}}_{n|n}^{(m)} = \sum_{i=1}^{2}\mu_n^{(i,m)}\hat{\mathbf{x}}_{n|n}^{(i,m)}$$

$$\mathbf{p}_{n|n}^{(m)} = \sum_{i=1}^{2}\mu_n^{(i,m)}\left[\mathbf{p}_{n|n}^{(i,m)} + \left(\hat{\mathbf{x}}_{n|n}^{(m)} - \hat{\mathbf{x}}_{n|n}^{(i,m)}\right)\left(\hat{\mathbf{x}}_{n|n}^{(m)} - \hat{\mathbf{x}}_{n|n}^{(i,m)}\right)^T\right]$$

D. Accelerational Observations

Given the potential variation in depth during a video resulting from throw corrections and general movement, we propose including dynamic variation in gravitational acceleration, $a^{(m)}$, through its incorporation as an additional state in the model. To provide estimates for $a^{(m)}$, this paper proposes incorporating known y-component acceleration distribution characteristics

into the model through observational inputs. Given that the objects are solely under the influence of gravitational forces for the majority of their trajectories, histogramming the acceleration's y-component reveals a prominent peak associated with the localised gravitational acceleration. Depending on how "dynamic" the system is required to be with respect to changes in depth, for each nth frame, the last K time samples are smoothed and extracted from $\mathbf{x}^{(m)}$:

$$\bar{V}_{n-i,m} = \frac{1}{\Delta t^2} \sum_{j=0}^{J-1} y_{n-i-j}^{(m)} G_j \, \{0 \leq i \leq K-1\},$$

where $\bar{V}_{n-i,m}$ is the 2D array corresponding to the past K smoothed acceleration values (indexed i) for each ball, m, given the nth frame. $y_{n-i-j}^{(m)}$ is the 2nd row of the state space $\mathbf{x}^{(m)}$ evaluated at frame $n-i-j$ and ball m, which is multiplied by a Second Differential Gaussian kernel of standard deviation $\sigma_{v,a}$ (length J), with entries G_j. The acceleration distribution given a specific video frame is modelled as such:

$$p(a \,|\, \bar{V}_{n-K:n,m}) \propto p(a)\, p(\bar{V}_{n-K:n,m} \,|\, a), \tag{2}$$

where $p(a)$ is the prior distribution defined in Eq. (1), and:

$$p(\bar{V}_{n-K:n,m} \,|\, a) = \frac{1}{K\sqrt{2\pi\sigma_k^2}} \sum_{i=1}^{K} \exp\left(-\frac{1}{2\sigma_k^2}\left(a - \bar{V}_{n-i,m}\right)^2\right),$$

which is a Gaussian kernel density estimator [14] of the vertical acceleration likelihood distribution, evaluated over the past K samples given a. Multiplying by the Gaussian prior produces the posterior distribution 2. Thus, the observational value for $a_n^{(m)}$ is taken to be the maximum a posteriori probability (MAP) estimate of distribution 2:

$$a_n^{(m)} = \max_a \{p(a \,|\, \bar{V}_{n-K:n,m})\}.$$

E. Initialisation

The number of objects, M, is inferred by performing an initial pass that extracts the mode of the number of objects detected in each frame from the inputted video. Subsequently, in the main pass, states can be initiated by considering the central detection coordinates extracted from the first frame (initialised as $n = 1$) that returns M detections, $\{(x_{d,k}, y_{d,k})\}_{k=1}^{M}$, where $(x_{d,k}, y_{d,k})$ corresponds to the x and y components of the kth detection. Thus, $\mathbf{x}_0^{(m)} = \begin{bmatrix} x_{d,m} & y_{d,m} & 0 & 0 & 0 & 0 & a_i \end{bmatrix}$, where a_i is the mean of the gravitational acceleration assumed prior, as per Eq. (1). For the case where the start of the pattern is tracked, objects can be initialised in the hand, as observed in figure 6i. The mode transition matrix, $\mathbf{\Pi}$:

$$\mathbf{\Pi} = \begin{bmatrix} 0.98 & 0.05 \\ 0.02 & 0.95 \end{bmatrix}$$

F. Data Association

This paper proposes a Global Nearest Neighbour (GNN) [30] association method through the construction of a probability association matrix, $\mathbf{L}$, such that each (k,m)th entry corresponds to the probability of the mth ball assignment to the kth ($1 \leq k \leq K$) detection. Assuming a uniform prior over associations, $a_n^{(k\to m)}$:

$$\begin{aligned} l_{k,m} &= p\left(a_n^{(k\to m)} \,|\, \mathbf{z}_n^{(k)}, \mathbf{z}_{1:n-1}^{(m)}\right) \\ &\propto p\left(\mathbf{z}_n^{(k)} \,|\, a_n^{(k\to m)}, \mathbf{z}_{1:n-1}^{(m)}\right) p\left(a_n^{(k\to m)}\right) \\ &\propto \sum_{i=1}^{2} p\left(\mathbf{z}_n^{(k)} \,|\, \mathbf{z}_{1:n-1}^{(m)}, i\right) p\left(i \,|\, \mathbf{z}_{1:n-1}^{(m)}\right) \\ &= \sum_{i=1}^{2} \mathcal{N}\left(\mathbf{z}_n^{(k)} - \mathbf{H}\hat{\mathbf{x}}_{n|n-1}^{(i,m)}, \mathbf{S}_n^{(i,m)}\right) \sum_{l=1}^{2} \Pi_{il} \mu_{n-1}^{(l,m)}. \end{aligned}$$

Assignment is subsequently achieved through maximisation with respect to the associations, determined by the Hungarian Sorting Algorithm [8], with padded rows or columns of zeros appended as necessary to the matrix to ensure it is square. If there are fewer detections than objects, resulting objects with no assignments are assigned to their projected values.

G. Siteswap Inference

1) Brief introduction to Siteswap Notation: Siteswap is a popular notation system employed widely by jugglers during performances, developed independently by Paul Klimek and Don Hatchby around 1981, and Mike Day, Adam Chalcraft and Colin Wright, around 1985 [20]. Siteswap sequences are constructed such that each value sequentially specifies a current ball's relative trajectory duration. Taking a simple case, such as a uniform duration Siteswap "3333...", reading each value sequentially provides the number of juggling beats after a release of a ball before the next release should occur. A juggling beat in this context refers to the time between catches (or throws), alternating from left to right, and corresponds to a shift in the current Siteswap index by one when reading the sequence. For this paper, asynchronous Siteswaps only will be inferred, although the analysis for synchronous patterns is similar and implementable from the current framework. Figure 1 presents a graphical interpretation of the duration of the motions for another 3 ball Siteswap, "531", often termed a parabola diagram or profile braid [23].

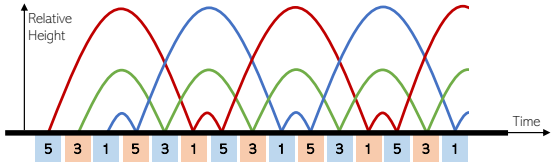


Fig. 1: Graphical representation of the Siteswap sequence "531" (neglecting dwell times) | Blue box = Left Hand, Orange box = Right Hand

Another popular visualisation method is termed a ladder diagram [26], as employed in several of the videos presented in the results section III; it is an extension of the parabola diagram, including both hands now as the sides of the ladder, and the Siteswap transitions as the ladder "rungs". For further literature on Siteswap, the relevant sections in [25], [27], [29] can be consulted.

2) Machine Learning-Enhanced Siteswap Beat Onset Detection: Analogous to time signature and beat tracking tasks performed for musical extracts, this paper proposes the employment of tempo and beat tracking algorithms, given the

inherent beat duration consistency in Siteswap performances. The employment of beat tracking algorithms is to both provide error correction for throws that were not initially detected given small IMM outputs, such as for Siteswap values 1 and 2, in addition to tempo inference, as explored in section II-G3. For beat-tracking algorithms, a beat onset detection function is required. The IMM output probabilities associated with the state-space model proposed in section II-B provide in-hand probabilities, thus, for the outputs to be employed in a detection function, beat onsets must be extracted. For non-causal applications, this is achievable through mode assignment from the IMM in-hand probabilities and subsequent extraction of the throw positions corresponding to the transition points from mode 2 to 1 (in-hand to airborne). To acquire mode assignments, a hard assignment of $\mu_n^{(2,m)} > 0.8$ is sufficient, however, due to temporary dips and noise, morphological operations [28], including erosion and dilation, are advised for increased accuracy. Whilst suitable for non-causal applications, providing real-time estimates for beat onsets through thresholding and morphological operations is challenging, especially given the transition point from mode 2 to 1 is desired as opposed to peaks in the IMM outputs. Given that numerous characteristics of beat transitions observable in trajectory data are not exploited in the proposed IMM that can aid in the provision of real-time direct Siteswap beat onset detection, this paper proposes a convolutional learning model designed to directly determine the probabilities associated with Siteswap beats onsets, $p\left(b_n^{(m)} = 1\right)$, for time step n ($1 \leq n \leq N$) and ball m ($1 \leq m \leq M$) given previously tracked training data. The probability is modelled via a logistic classification model:

$$p\left(b_n^{(m)} = 1 \,|\, \bar{\mathbf{x}}_{1:n}^{(m)}, \mathbf{W}\right) = \frac{1}{1 + e^{-\sum_{i,j} w_{i,j} \bar{x}_{n-i,j}^{(m)}}},$$

where $\bar{\mathbf{x}}_{1:n}^{(m)}$ (with entries $\bar{x}_{n,j}^{(m)}$) is the processed input state space derived from tracked data $\mathbf{x}_{1:n}^{(m)}$ for ball m, frame n and state j, $\bar{\mathbf{x}}_n^{(m)} = \{\bar{x}_{n,j}^{(m)}\}_{j=1}^{7}$. $\mathbf{W}$ (with entries $w_{i,j}$) corresponds to a 7x7 convolutional kernel, determined through training. $\bar{\mathbf{x}}_{1:n}^{(m)}$ is evaluated through taking the output IMM states $\mathbf{x}_{1:N}^{(m)}$, excluding the gravitational acceleration state. A 7th row is also added corresponding to the acceleration magnitude, as such:

$$\bar{\mathbf{x}}_n^{(m)} = \begin{bmatrix} x_n^{(m)} & y_n^{(m)} & \dot{x}_n^{(m)} & \dot{y}_n^{(m)} & \ddot{x}_n^{(m)} & \ddot{y}_n^{(m)} & |\ddot{d}| \end{bmatrix}^T,$$

where $|\ddot{d}| = \left(\left(\ddot{x}_n^{(m)}\right)^2 + \left(\ddot{y}_n^{(m)}\right)^2\right)^{\frac{1}{2}}$. Training is achieved through first evaluating $\bar{\mathbf{x}}_n^{(m)}$ for each ball from successfully tracked runs. A vector $\{t_n^{(m)}\}_{n=1}^{N}$, corresponding to the ground truth probability of a throw given ball m for frames $1:N$, can subsequently be constructed through manually labelling each ball's trajectory, and convolving with a Gaussian kernel to account for uncertainty in labelling. The posterior distribution given the labelled data, $\mathbf{D} = \{t_{1:n}^{(m)}, \bar{\mathbf{x}}_{1:n}^{(m)}\}$ is thus:

$$p\left(\mathbf{W} \,|\, \mathbf{D}\right) \propto p\left(\mathbf{W}\right) \prod_{n=1}^{N} \prod_{m=1}^{M} p\left(b_n^{(m)} = 0 \,|\, \bar{\mathbf{x}}_{1:n}^{(m)}, \mathbf{W}\right)^{1 - t_n^{(m)}} \times p\left(b_n^{(m)} = 1 \,|\, \bar{\mathbf{x}}_{1:n}^{(m)}, \mathbf{W}\right)^{t_n^{(m)}},$$

where $p\left(\mathbf{W}\right)$ refers to the prior, with covariance matrix, $\sigma_w^2 \mathbf{I}$. Thus, the partial derivatives of the negative log of the posterior ($\mathcal{L}\left(\mathbf{W}\right) = -\log L\left(\mathbf{W}\right)$) with respect to $\mathbf{W}$:

$$\frac{\partial}{\partial w_{i,j}}\left(\mathcal{L}\left(\mathbf{W}\right)\right) = \frac{1}{\sigma_w^2} w_{i,j} + \sum_{n=1}^{N} \sum_{m=1}^{M} \bar{x}_{n-i,j}^{(m)} \left[\frac{1}{1 + e^{-\sum_{u,v} w_{u,v} \bar{x}_{n-u,v}^{(m)}}} - t_n^{(m)}\right].$$

For computational efficiency during training, the objective function has been setup specifically such that:

$$z_n^{(m)} = \sum_{u,v} w_{u,v} \bar{x}_{n-u,v}^{(m)}. \tag{3}$$

Thus, for each training step, $\mathbf{Z}^{(m)}$ ($\{z_n^{(m)}\}_{n=1}^{N}$) can be computed through evaluating the vector sum (with respect to j in Eq. (3)) of the processed input space fft convoluted with the convolutional kernel, $\mathbf{W}$. Therefore, the vectorised gradient descent [16] update step becomes:

$$w_{i,j} \Leftarrow w_{i,j} - \alpha \frac{\partial}{\partial w_{i,j}}\left(\mathcal{L}\left(\mathbf{W}\right)\right) = \left(1 - \frac{1}{\sigma_w^2}\right) w_{i,j} - \alpha \sum_{n=1}^{N} \sum_{m=1}^{M} \bar{x}_{n-i,j}^{(m)} \left[\frac{1}{1 + e^{-z_n^{(m)}}} - t_n^{(m)}\right],$$

where α is the learning rate. Thus, given the determined weights, $\mathbf{W}$, the following can employed to provide a real-time Siteswap beat onset probability, $\lambda_n^{(m)}$, for ball m:

$$\lambda_n^{(m)} = p\left(b_n^{(m)} = 1 \,|\, \bar{\mathbf{x}}_{1:n}^{(m)}, \mathbf{W}\right) = \frac{1}{1 + e^{-\sum_{i,j} w_{i,j} \bar{x}_{n-i,j}^{(m)}}},$$

3) Inference: For **non-causal applications**, the Siteswap beat onset detection function can be computed as a series of delta functions corresponding to the transition points from mode 2 to 1 (from in hand to airborne) extracted from the IMM mode probabilities, as per figure 2. The paper subsequently proposes the employment of the Metrogram beat tracking algorithm [24], to perform tempo inference for performance analysis, in addition to beat tracking if necessary to ensure missed beats (such as 0's, 1's, and 2's) are appropriately identified. Another benefit of extracting the tempo profile is to enable the extraction of dwell ratios. The dwell ratio, d_r, defined as the proportion of time an object spends in the hand (the dwell time) relative to twice the local beat length, is a particularly useful measure for performance analysis given its ability to inform the user about certain imbalances and inefficiencies. The dwell ratio as a continuous-time function can be computed: $d_r(t) = d_t(t)T(t)$, where $T(t)$ is the tempo function extracted from the beat tracking algorithm in Hz, and $d_t(t)$ is the interpolated dwell times. The dwell times are the peak widths associated with mode 2 assignments. Siteswap inference can be achieved by formulating a Siteswap Transition Matrix, $S_{i,j}$, where i corresponds to the ball index, and j, to the beat index. For each throw, the entry corresponding to the thrown ball index and the associated beat, is labelled as a 1 in the matrix, else a 0. Each throw can subsequently be assigned the distance to the next throw determined by traversing the matrix horizontally (increasing beat index j); this corresponds

to the Siteswap value. For **causal applications**, a simple thresholding assignment algorithm can be applied on $\lambda_n^{(m)}$, typically $\lambda_n^{(m)} > 0.8$, to register a Siteswap beat onset. To account for potential beat length variability performing different Siteswaps (as observed in figure 5b), this paper proposes for each caught object, backtracking to the previous beat index the object was assigned, and assigning that beat index the number of backtracked beats, which corresponds to the Siteswap value.

III. Results and Discussion

Given the lack of existing datasets for Siteswap detection, a custom Siteswap dataset is proposed, featuring a variety of videos ranging from 3 to 11 balls. All results used here are with permission from the performers. In subsequent work, we plan to further extend and augment the custom dataset with manually annotated ground truth labels, thereby enabling performance comparisons with standardised tracking algorithms. Figures 6a - 6e present tracking and non-causal Siteswap inference results for videos of James juggling several 3 to 7 ball Siteswaps. The results are visualised with software developed by James, including ladder diagrams for real-time Siteswap visualisation and pose recognition, implemented with the MediaPipe library [31]. Object segmentation performance is presented in figure 6i for a 9 ball run by James. For the 5 ball routine (figure 6c), various tracked states are presented in figure 4, and selected IMM in hand-probabilities, mode, and beat assignments presented in figure 2. Additionally, the tracked acceleration values for this example are presented in figure 3a, and the beat Tempogram, tracked tempo, and dwell ratios shown in figure 5, extracted using the beat tracking algorithm. Note to convert the tracked states from pixel to world coordinates, the states are divided by the time-varying tracked acceleration values (and multiplied by 9.81). Figure 6f presents tracking of James performing a 7 ball cascade, and figure 6h displays real-time tracking of Tom Whitfield's[3] 11 ball World Record run (34 catches), with catch recognition and the three most relevant vertical spatial derivatives plotted. Tom uses the software to help recognise irregularities in endurance runs. Furthermore, an example of real-time Siteswap detection is shown in figure 6i. Whilst the algorithm was primarily optimised for ball detection, future research will explore more generalised object detection schemes for joint ball, club, and ring localisation; for instance, figure 6f presents tracking for a 5 club run by Adrian Goldwaser[4].

Overall, the tracking results indicate a high-level accuracy in the various spatial derivatives, indicative of the adaptive IMM Kalman filter implementation. The period of increased spatial uncertainty during the catching and throwing of each object was effectively accounted for by the bimodal parallel IMM filters, even in the presence of occlusion from the hands, as for instance observed frequently in figure 6f. For all the presented examples, the correct Siteswaps were inferred and the beat positions accurately computed for both the causal and non-causal methodologies. From figure 3a, it is interesting to note the local variation in the tracked gravitational acceleration for each object due to variations in depth relative to the camera; distances from the camera (depth) are estimated in figure 3b given the known camera properties. Interestingly, whilst performing the Siteswap 744 at the $4-6$ second mark, the objects are thrown slightly forward, and subsequent error correction results in overcompensation with some balls being thrown slightly backward. Similar localised instability is observed in figure 5c, where the dwell ratio variance increases compared to the previously stable 0.6 value; note the sudden dwell ratio anomaly at $t = 10.8s$ corresponding to a 1 throw. Likewise, the tempo increases whilst performing both the Siteswap 744 $(4-6s)$ and 66661 $(10-11s)$ relative to the steady $\approx 5\text{Hz}$ cascade beat tempo, as observed in figure 5b, which is typical given the greater Siteswap values. As expected, the greatest y-component hand acceleration occurs when releasing the greatest Siteswap-valued throw, at $t = 4.4s$ $(67\text{m}\,\text{s}^{-2})$. As a further performance analysis utility, the minimum distance between any given object across time is plotted in figure 3c. Note that, as expected, the average of the minimum distances decreases whilst performing the Siteswaps 744 and 66661, indicative of the inherent increased trajectory proximity for both patterns, and the temporary instability of the pattern previously observed for these segments. Indeed, there are two points $t = 5.1s, t = 10.9s$ which just exceed the collision threshold (2 times the ball radius); whilst this may technically register as a collision, the collision occurs in the pixel plane, and, as observed in figure 3b, depth variation is not negligible, thus, whilst close, no actual collision truly occurred due to the slight depth variation.

IV. Conclusion

The proposed bimodal adaptive IMM Kalman Filter implementation has been shown to provide effective multi-object localisation for Siteswap-oriented juggling routines. Evaluating the algorithm on numerous videos featuring a variety of background conditions and Siteswap routines, the findings underline the algorithm's proficiency in distinguishing between the two discrete tracking cases, providing accurate state estimation and mode assignment. Likewise, evaluation has demonstrated its capability in deducing accurate Siteswap sequences from video, for both causal and non-causal paradigms, in addition to providing valuable performance utilities. Ultimately, the algorithm serves as a comprehensive tracking, performance analysis, and visualisation utility, enabling Siteswap inference and facilitating real-time competitive evaluation.

[3]https://www.instagram.com/_tom_whitfield_

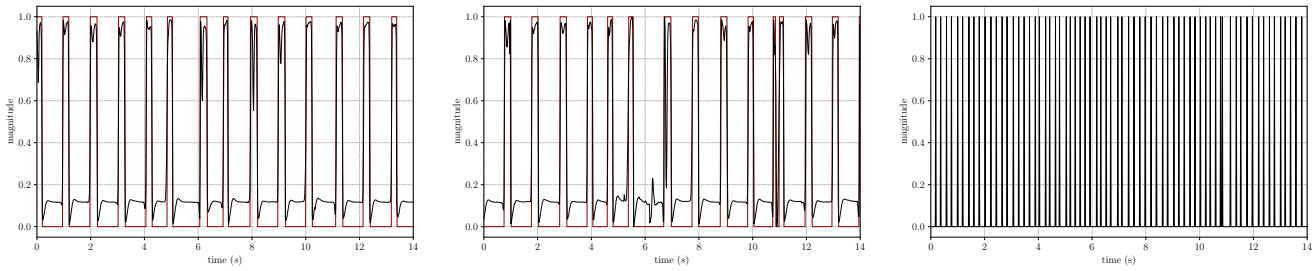
[4]https://www.instagram.com/adriangoldwaser

References

[1] Betuel E. 2021. "A circus act turned extraordinary sport is making its case for the Olympics" Inverse. [Online] Available at: https://www.inverse.com/mind-body/competitive-juggling-is-about-to-get-real

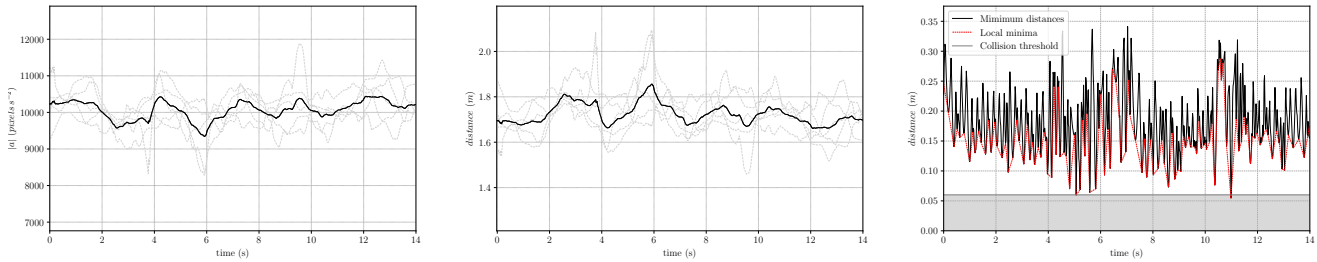
[2] BPG, "Could Juggling Become an Olympic Sport?," Medium, 2021 [Online] Available at: https://authorbpg.medium.com/could-juggling-become-an-olympic-sport-53e811c1897

[3] C. Hester, S. Bevan, "A show of hands: Meet James Cozens – the Guinness World Record-holding juggling engineer," *University of Cambridge Stories*, 2023. [Online] Available at: https://www.cam.ac.uk/stories/Cambridge-guinness-record-juggling-engineer-james-cozens



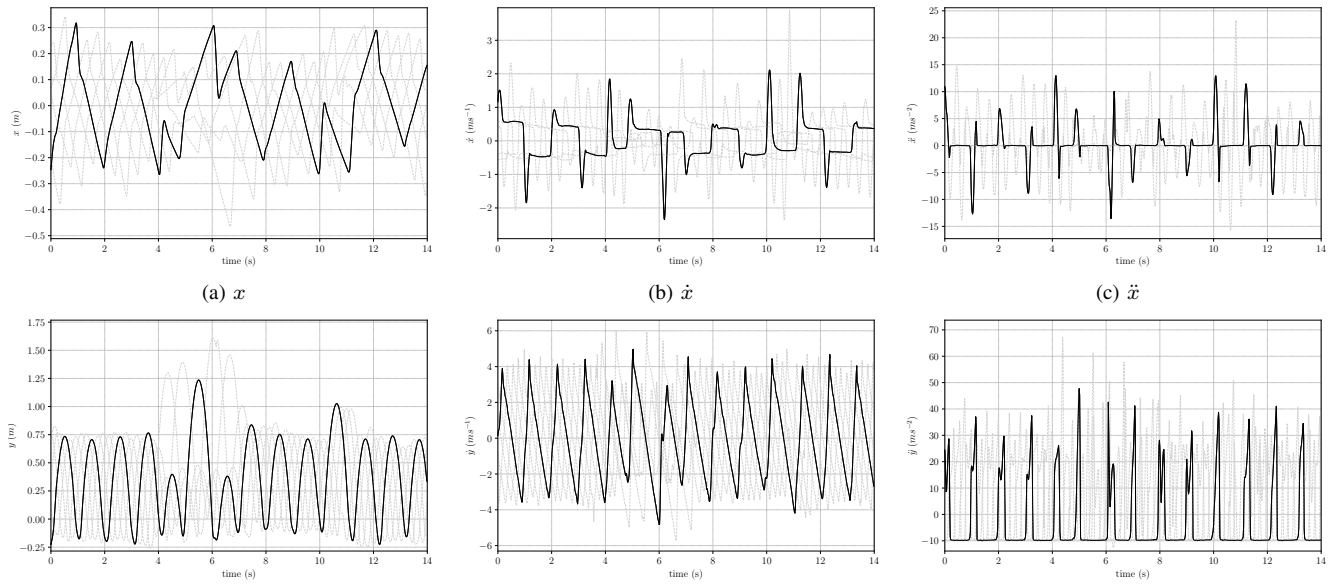
(a) Ball 1 (b) Ball 5 (c) Beat assignments

Fig. 2: In hand probabilities (black), mode (red), and final beat assignments for the 5 ball routine (figure 6c)



(a) $|a|$ (mean emphasised) (b) Depth, z (mean emphasised) (c) Minimum distances

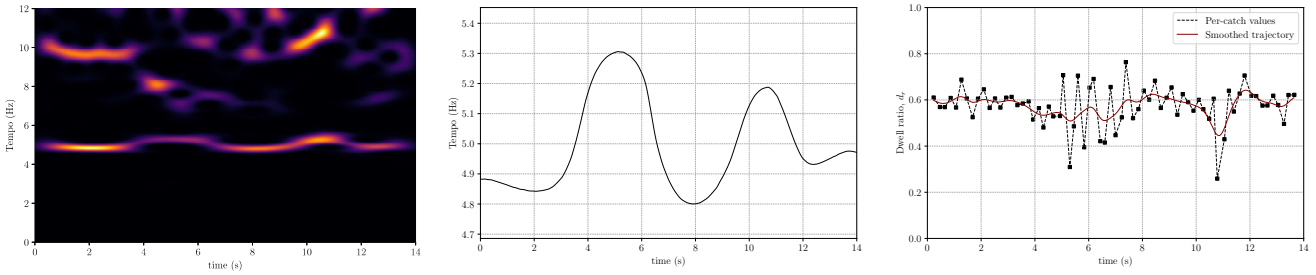
Fig. 3: Tracked acceleration ($a^{(m)}$), estimated distances from the camera, and minimum distances for the 5 ball routine



(a) x (b) $\dot{x}$ (c) $\ddot{x}$

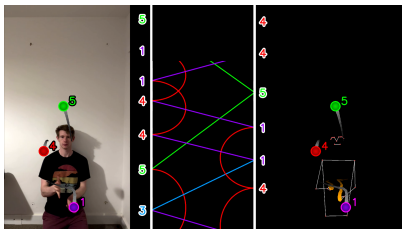
(d) y (e) $\dot{y}$ (f) $\ddot{y}$

Fig. 4: Extracted component states for the 5 ball routine featured in figure 6c, with ball 1 emphasised

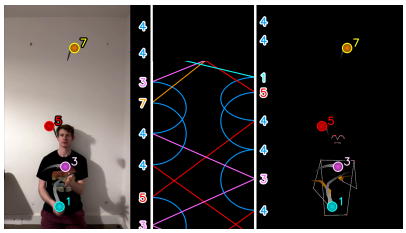


(a) Tempogram (b) Inferred beat tempo trajectory (c) Dwell times, d_r

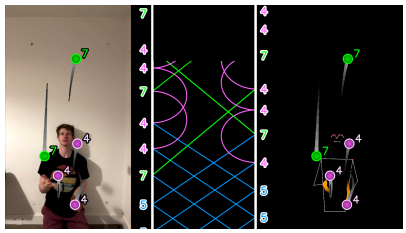
Fig. 5: Beat tempo and dwell ratio inference for the 5 ball routine featured in figure 6c



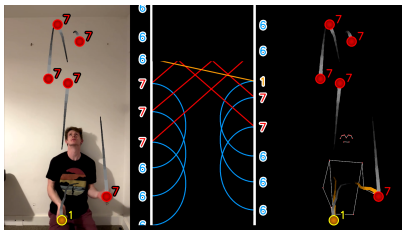
(a) 3 Ball tracking, including the Siteswaps 3, 441, and 51414. Video available below



(b) 4 Ball tracking, including the Siteswaps 4, 534, and 7531. Video available below



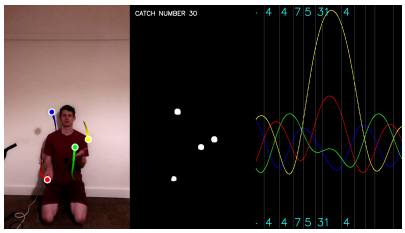
(c) 5 Ball tracking, including the Siteswaps 5, 744, and 66661. Video available below



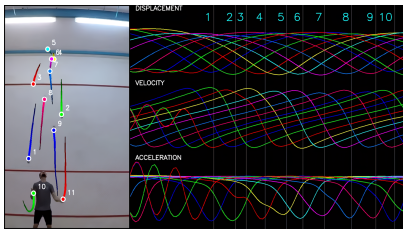
(d) 6 Ball tracking, including the Siteswaps 6, and 777771. Video available below

(e) real-time 7 Ball tracking demonstration with pose recognition. Video available below

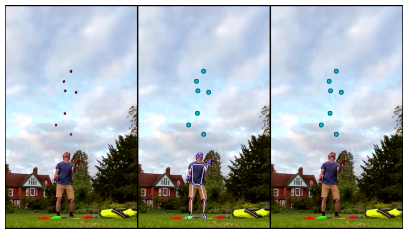
(f) 5 club tracking featuring Adrian Goldwaser. Video available below



(g) Semi-real-time Siteswap tracking for a 4 ball routine, featuring catch detection and visualisation of the vertical displacements for each ball. Video available below



(h) Real-time tracking for Tom Whitfield's 11 Ball World Record run, featuring catch detection and visualisation of the vertical spatial derivatives. Video available below



(i) real-time 9 Ball segmentation demonstration with pose recognition. Video available below

Fig. 6: Dataset tracking results. All tracked videos are available here: https://youtube.com/playlist?list=PLWtN-KJsWmthw4U7PDDyAyeDx_qVGlbNt&feature=shared

[4] A. Alonso, "Obtención de "Siteswap" de malabares mediante análisis de vídeo", Unpublished paper. [Online] Available at: https://github.com/AlejandroAlonsoG/tfg_jugglingTrackingSiteswap

[5] Z. Zivkovic and F. van der Heijden, "Efficient adaptive density estimation per image pixel for the task of background subtraction," *Pattern Recognition Letters*, vol. 27, no. 7, pp. 773–780, 2006

[6] P. C. Mahalanobis, "On the generalised distance in statistics," *Proceedings of the National Institute of Sciences of India*, vol. 2, no. 1, pp. 49–55, 1936.

[7] J. Boyce, "Hawkeye juggling video analysis Github Page," [Online] Available at: https://github.com/jkboyce/hawkeye?fbclid=IwAR2q-f4q9ZYbF9uDskBC2R0fRpB_SeU_vm18S-y2NVMl7g6RkRXXmqoAhZU

[8] H. W. Kuhn, "The Hungarian method for the assignment problem," *Naval Research Logistics Quarterly*, vol. 2, no. 1, 1955

[9] S. Dingler, "State estimation with the Interacting Multiple Model (IMM) method," *arXiv preprint arXiv:2207.04875*, 2022.

[10] E. Ribnick et al., "Detection of Thrown Objects in Indoor and Outdoor Scenes," in *Proceedings of the 2007 IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS 2007*.

[11] A. F. Genovese, "The interacting multiple model algorithm for accurate state estimation of maneuvering targets," *Johns Hopkins APL Technical Digest (Applied Physics Laboratory)*, vol. 22, pp. 614–623, 2001.

[12] M. Truzzi and M. Truzzi, "Notes Toward a History of Juggling," *Bandwagon*, vol. 18, no. 2, 1974. [Online] Available at: http://www.juggling.org/papers/history-2/

[13] C. Szegedy, A. Toshev, and D. Erhan, "Deep Neural Networks for Object Detection," in *Advances in Neural Information Processing Systems 26 (NIPS 2013)*, 2013.

[14] B. Silverman, "Density Estimation for Statistics and Data Analysis," CRC Press, vol. 26, 1986.

[15] B. Øksendal, *Stochastic Differential Equations: An Introduction with Applications*, 6th ed. Berlin, Germany: Springer, 2003.

[16] J. Nocedal and S. J. Wright, *Numerical Optimization*, 2nd ed. New York, NY, USA: Springer, 2006.

[17] M. Rizon, H. Yazid, and P. Saad, "Object Detection using Circular Hough Transform," *American Journal of Applied Sciences*, vol. 2, no. 12, 2005.

[18] T. Lindeberg, "Image Matching Using Generalized Scale-Space Interest Points," *J Math Imaging Vis*, vol. 52, pp. 3–36, 2015.

[19] R. E. Kalman, "A New Approach to Linear Filtering and Prediction Problems," *Transactions of the ASME–Journal of Basic Engineering*, vol. 82, pp. 35-45, 1960.

[20] D. Hayes and T. Shubin, "Mathematical Adventures for Students and Amateurs," *Mathematical Association of America*, p. 99, 2004.

[21] World Juggling Federation Website [Online] Available at: https://www.thewjf.com/

[22] J. Weiner, "The Art of Site Swap Transitioning: How Modern Circus Jugglers Choreograph Performances," TC Public Space, 2016. [Online] Available at: https://tcpublicspace.wordpress.com/2016/04/27/the-art-of-site-swap-transitioning/

[23] M. Macauley, "Braids and Juggling Patterns," Harvey Mudd College Senior Theses, pp. 5 - 6, 2003. [Online] Available at: https://scholarship.claremont.edu/hmc_theses/151/

[24] J. Cozens, S. Godsill, "Dynamic Time Signature Recognition, Tempo Inference, and Beat Tracking Through the Metrogram Transform," in IEEE Open Journal of Signal Processing, vol. 5, pp. 140-149, 2024.

[25] B. Polster, "The Mathematics of Juggling", Springer Verlag, 2003.

[26] S. Devadoss and J. Mugno, "Juggling Braids and Links," *Springer Science and Business Media*, vol. 29, no. 3, pp. 15-17, 2007.

[27] J. Buhler, D. Eisenbud, R. Graham, and C. Wright, "Juggling drops and descents," *American Mathematical Monthly*, vol. 101, no. 6, pp. 507-519, 1994.

[28] L. Vincent, "Morphological Grayscale Reconstruction in Image Analysis: Applications and Efficient Algorithms," *IEEE Transactions on Image Processing*, vol. 2, no. 2, pp. 176-201, April 1993.

[29] E. R. Tou, "Asymptotic counting theorems for primitive juggling patterns," *International Journal of Number Theory*, vol. 15, no. 5, 2019.

[30] D. Reid, "An algorithm for tracking multiple targets," in IEEE Transactions on Automatic Control, vol. 24, no. 6, pp. 843-854, December 1979, doi: 10.1109/TAC.1979.1102177.

[31] V. Bazarevsky, I. Grishchenko, K. Raveendran, T. Zhu, F. Zhang, and M. Grundmann, "BlazePose: On-device Real-time Body Pose tracking," arXiv preprint arXiv:2006.10214, 2020.