Physics-based Activity Modelling in Phase Space

Ricky J. Sethi*
UCLA and UC Riverside
900 University Ave
Riverside, CA 92521
rickys@sethi.org

Amit K. Roy-Chowdhury
UC Riverside
900 University Ave
Riverside, CA 92521
amitrc@ee.ucr.edu

ABSTRACT

In this paper, we employ ideas grounded in physics to examine activities in video. We build the Multi-Resolution Phase Space (MRPS) descriptor, which is a set of feature descriptors that is able to represent complex activities in multiple domains directly from tracks without the need for different heuristics. MRPS is used to do single- and multi-object activity modelling in phase space, which consists of all possible values of the coordinates. The MRPS contains the Sethi Metric (S-Metric), the Hamiltonian Energy Signature (HES), and the Multiple Objects, Pairwise Analysis (MOPA) descriptors: the S-Metric is a distance metric which characterizes the global motion of the object, or the entire scene, with a single, scalar value; the HES is a scalar or multi-dimensional time-series that represents the motion of an object over the course of an activity using either the Hamiltonian or the S-Metric; and the MOPA contains phase space features for paired activities, in which we develop physical models of complex interactions in phase space (specifically, we model paired motion as a damped oscillator in phase space). Finally, we show the S-Metric is a proper distance measure over a metric space and prove its additivity; this allows use of the S-Metric as a distance measure as well as its use in the HES. Experimental validation of the theory is provided on the standard VIVID and UCR Videoweb datasets capturing a variety of problem settings: single agent actions, multi-agent actions, and aerial sequences, including video search.

Keywords

physics-based, activity modeling, phase space

1. INTRODUCTION

Motion underlies all activities; human activities, in fact, are defined by motion. The rigorous study of motion has been the cornerstone of physics and we exploit the physics of motion to understand activities of individual objects as well as the interaction between them.

*Corresponding author

ICVGIP '10, December 12-15, 2010, Chennai, India

1.1 Related Work & Contributions

We build liberally upon theoretical thrusts from several different disciplines, including Analytical Hamiltonian Mechanics and human activity recognition [1], especially for multiple activities [2]. Our approach also draws inspiration from the method employed in [3], which detects global motion patterns by constructing super tracks using flow vectors for tracking high-density crowd flows in low-resolution. Our methodology in this paper, on the other hand, works in both high- and low-resolution and for densely- and sparsely-distributed objects since all it requires is the (x, y, t) tracks for the various objects' motion analysis, as shown in Figure 1.

Using a physics-based methodology, [4] derived the **Hamiltonian Energy Signatures (HES)** for individual objects; the HES is a scalar or multi-dimensional time-series that represents the motion of an object over the course of an activity. The HES is derived from Hamilton's Principle of Least Action and is a time-series built using the Hamiltonian Equations of motion. Hamiltonian Dynamics is an elegant and powerful alternative formulation of classical mechanics that not only gives the equations of motion for a system but, more importantly, provides greater, and often more abstract, insight about the system. Hamilton's equations are primarily of interest in establishing basic theoretical results, rather than determining the motions of particular systems.

In this paper, however, we generalize the HES and phase space analysis and also apply the physics-based methodology to modelling multi-object activities in phase space. The **phase space** of a system consists of all possible values of the coordinates, which is usually represented as the space of position vs momentum (x, p) or position vs velocity (x, v) but can be any set of coordinates like the generalized coordinates and Hamiltonian values discussed in detail below.

In addition, we develop a distance metric, the **Sethi Metric (S-Metric)**, which characterizes the global motion of the object, or the entire scene, with a single, scalar value; we show, in the Appendix, that the S-Metric is a Norm and thus the S-Metric is a proper distance measure over a metric space and can be used as a distance in other distance measures. The S-Metric can also be represented as a series of values in the HES if the total video is broken up into shorter time-segments since we prove the S-Metric is additive in the Appendix, as well. A derivation of the HES and S-Metric from the Principle of Least Action is also given in Section 2.1. The HES and the S-Metric have distinct properties: the

Figure 1: Tracks to Phase Space

S-Metric can be used to characterize the entire scene with a single, scalar, global value; the HES time series, on the other hand, can characterize activities of individual objects.

Finally, we develop physical models of complex interactions in phase space; specifically, we do a **Multiple Objects Pairwise Analysis (MOPA)** in which we model paired motion as a damped oscillator in phase space using relative distances. Others, such as [5], have utilized relative distances within a coupled HMM but our methodology does not require an external stochastic framework and can characterize motion directly from tracks without requiring training or classifiers. In addition, they interpret pedestrian actions only and create prior models of human behaviour by using synthetic agents that encapsulate their assumptions for simple actions that are atomic in nature and only look at single interaction detections. Our approach, on the other hand, looks at complex activities betwen multiple, interacting objects of any variety without the need for synthetic agents or prior models of the objects. Also, unlike the heuristic examination of simple activities in a single domain using relative distances in [6], we create a consistent framework to derive and unify different representations of motion for activity recognition. All these previous approaches also rely on a classifier (coupled HMM in [5] and a simple hypothesis testing framework based on two-class nearest neighbor classification with extensive parameterizations and thresholds in [6]) whereas our approach uses our physics-based models to do the recognition directly.

The HES, S-Metric, and MOPA together form a *Multi-Resolution Phase Space (MRPS)* feature descriptor set that can be applied to any system, from multiple objects to multiple points on a single object to a single point on each object. The advantage of our approach over others is that we build a single, consistent feature descriptor set, the MRPS, to do both single- and multi-object modelling in phase space across different resolutions and domains. We are thus able to model a wide variety of complex motions using a single, consistent framework that operates in phase space. This consistent framework yields a descriptor that is able to represent complex activities directly from tracks without the need for different heuristics. Thus, our main contributions are:

- Building the MRPS feature descriptor set to do single- and multi-object activity modelling in phase space across different resolutions and domains directly from tracks without requiring training or classifiers
- Developing the MOPA, phase space features for paired activities
- Developing the S-Metric and proving it is a proper distance measure over a metric space, as well as its additivity

2. MULTI-RESOLUTION PHASE SPACE FEATURE DESCRIPTOR

The Multi-Resolution Phase Space (MRPS) feature descriptor is composed of the HES, S-Metric, and MOPA and can be used to model both single-object and multi-object activities in phase space, as discussed in detail below.

2.1 S-Metric in Phase Space

Following Hamilton's approach, we define **Hamilton's Action**, S, for motion along a worldline between two fixed physical events (not events in activity recognition) as:

$$S \equiv \int_{t_1}^{t_2} L(q(t), \dot{q}(t), t)dt \tag{1}$$

with q, the generalized coordinates, and L, in this case, the *Lagrangian* which, for a conservative system, is defined as $L = T - U$, where, T is the *Kinetic Energy* and U is the *Potential Energy*. The Action is what we cast as the **Sethi Metric (S-Metric)**. The *Hamiltonian function*, derived from *Hamilton's Variational Principle*, is usually stated most compactly, in generalized coordinates, as [4]:

$$H(q, p, t) = \sum_i p_i \dot{q}_i - L(q, \dot{q}, t) \tag{2}$$

where H is the Hamiltonian, p is the generalized momentum, and $\dot{q}$ is the time derivative of the generalized coordinates, q. This defines the dynamics on the system's phase space, in which the q_i and p_i are regarded as functions of time. The *phase space* of a system consists of all possible values of the generalized coordinate variables q_i and the generalized momenta variables p_i. If the Hamiltonian is time-independent, then phase space is 2-dimensional, (q,p); if the Hamiltonian is time-dependent, then phase space is 3-dimensional, (q,p,t) [7]. We may also look at modified phase plots of (H,t), (H,q,p), etc.

We also prove (in the appendix) the S-Metric is a proper distance measure over a metric space thus allowing its use as a global, scalar signature of the system, possibly consisting of multiple objects. Finally, we prove the additivity of S in the Appendix, which allows us to use the S-Metric within the HES, as explained further in the next section.

2.2 Modelling Single Object Activities in Phase Space

The Hamiltonian in (2) is normally utilized as the Hamiltonian Energy Signature (HES) for various objects (either entire objects or the parts of a single object) involved in an activity, thus representing the motion of each object over the course of the activity as a time series. We can now also use the Action within the HES since we prove the additivity of S in the Appendix. By incorporating the Action or Hamiltonian within it, the HES gives a simple, intuitive expression

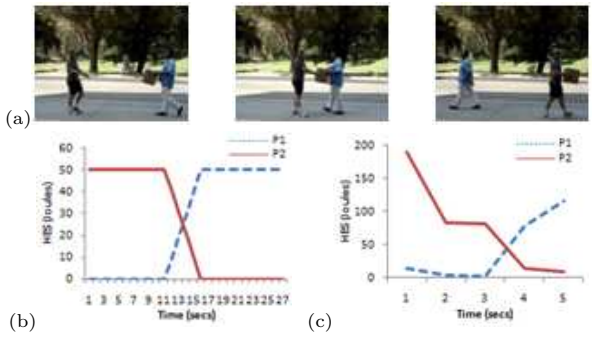

Figure 2: (a) Video of two people exchanging a box and their (b) Ideal vs (c) Actual HES.

for an abstract, compact representation of the system; i.e., the characteristic time-series curves for each object.

In this section, we will deal with individual objects. They can be represented by a single point, like the center of mass, or a collection of points, like points on the contour of the human body or joints of the body. For example, if we track a person in video, we can compute these HES curves for the centroid of the person (considering the person as an entire object, as shown in Figure 2) or consider all the points on the contour of that person's silhouette, thus leading to a multi-dimensional time series (which can, for example, represent the gait of a person, as shown in [8, 9]).

Note that these HES curves can be computed in either the image plane, yielding the Image HES, composed of the pseudo-Hamiltonian or pseudo-Action, as used in this paper, or in the 3D world, giving the Physical HES, composed of the actual Hamiltonian or Action, depending on the application domain and the nature of the tracks extracted. We thus use the motion trajectories to calculate this physically-relevant pseudo-Hamiltonian or pseudo-Action and the more information we have about the objects in the video, the more physically significant they become. In either case, the Hamiltonian framework gives a highly abstract, compact representation for a system and can yield the energy of the system being considered under certain conditions and the more information we have about the objects in the video, the more physically significant it becomes. Therefore, this pseudo-Hamiltonian or pseudo-Action (used to construct the HES) allows us to extract an abstract representation of the motion of the underlying physical systems we consider in video and allows us to represent a video sequence using a physics-driven HES and provides a framework for theoretical extensions.

We compute the Lagrangian and then the pseudo-Hamiltonian or pseudo-Action using $T = \frac{1}{2}mv_o^2$ and $U = mg(y_b - y_a)$, derived directly from the trajectories, (x, y, t). We thus segment the video into systems and sub-systems (e.g., whole body of a person, or parts of the body) and, for each of those, get their tracks, from which we compute T and U, and use that to get the HES curve signature, which can then be evaluated further and the results analyzed accordingly. In the same manner, we compute the S-Metric from the tracks for the relevant time period by first computing the L from the T and U, as shown in Figure 1. Thus, we use the video to gain knowledge of the physics and use the physics to capture the essence of the system being observed via the HES and S-Metric.

Examples: For example, in the general case when U $\neq 0$, the Lagrangian, $T - U$, of a single particle or object acting under a constant force, F (e.g., for a gravitational field, g, F=mg) over a distance, x, is:

$$\begin{array}{l} L(x(t), \dot{x}(t)) = \frac{1}{2}mv^2 - Fx \\ \text{with } x = x_o + v_o t + \frac{1}{2}at^2 \text{ and } a = \frac{F}{m} \end{array} \quad (3)$$

We now use this Lagrangian to calculate Hamilton's Action for the general system:

$$\begin{array}{l} S = \int_{t_a}^{t_b} L dt = \int_{t_a}^{t_b} \left(\frac{1}{2}m(v_o^2 + 2v_o\frac{F}{m}t) - F\left(x_o + v_o t\right)\right) dt \\ = \frac{1}{2}mv_o^2(t_b - t_a) - Fx_o(t_b - t_a) \end{array} \quad (4)$$

Using Hamilton's Variational Principle on (4) for a gravitational force yields (with y being the vertical position, which can be determined from the tracks):

$$H = T + U = \frac{1}{2}mv_o^2 + mgh = \frac{1}{2}mv_o^2 + mg(y_b - y_a) \quad (5)$$

Here, as a first approximation, we treat m as a scale factor and set it to unity; in future, we can estimate mass using the shape of the object or other heuristics, including estimating it as a Bayesian parameter. In addition, mass is not as significant when we consider the same class of objects.

We can thus plot Hamilton's Action vs. time as the HES curve since the partial derivative of the Action is energy [10]. In addition, we can compute differences in the S-Metric between *multiple* objects by just subtracting their S Metrics, since Hamilton's Action is shown to be additive in Appendix. Since the HES is already a time-series, we can compare two characteristic HES curves for two activities using a Dynamic Time Warping (DTW) algorithm.

We can also compare their full-fledged S-Metrics or, when we need a greater granularity of matches, we can segment the video into smaller time-intervals and compute the S-Metric piecewise for each of them, leading to a time-series in this case also. The additivity of S allows this and we can use DTW for matching the S-Metric sequences. In addition, it can be shown that the Image HES allows us to recognize activities in a moderately view-invariant manner while the 3D Physical HES is completely view invariant; the invariance of both comes from the invariance of the HES to affine transformations.

The main advantage of using the Hamiltonian formalism is that it provides a framework for theoretical extensions to more complex physical models. In the next section, we show just such an example of extension to modelling the paired activity of two objects moving together.

2.3 Modelling Multi-Object Activities in Phase Space

In this section, we do a pairwise analysis of multiple objects. In particular, we develop physical models of complex interactions in phase space and do a **Multiple Objects Pairwise Analysis (MOPA)** in which we model paired motion as a damped oscillator in phase space using relative distances. MOPA thus contains phase space features for paired activities, with physical models of complex interactions in phase space.

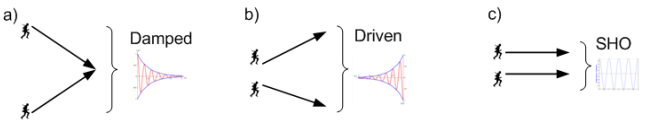

Figure 3: modelling Paired Motion: a) Two people walking towards each other is modeled as a Damped Oscillator; b) Two people walking away from each other is modeled as a Resonant Driven Oscillator; c) Two people walking parallel to each other is modeled as an Un-driven SHO.

We start with the problem of trying to categorize the motion of two objects in video and to see if their motion is correlated (this is taken as a simple example and can be generalized further). We hypothesize that the motion of two objects can be modeled as an oscillation with the envelope of that oscillation, as seen in Figure 4, being the average distance over time between the two. We thus calculate the relative distance with respect to time between these two objects and use that as the envelope for the oscillation.

To further elucidate our method, let us consider two people walking. In order to model the motion of two people walking, we consider the three possibilities: they can walk towards each other, they can walk away from each other, or they can walk parallel to each other. These three situations are shown in Figure 3 where we model all three types of motion as an oscillator. In Figure 3a, two people walking towards each other is modeled as a Damped Oscillator; in Figure 3b, two people walking away from each other is modeled as a Resonant Driven Oscillator; in Figure 3c, two people walking parallel to each other is modeled as an Undriven Simple Harmonic Oscillator (SHO).

A damped oscillator is described by the following second order differential equation of motion, which can represent all three cases and can model mass, damping, and elasticity:

$$m\ddot{x} + c\dot{x} + kx = 0 \tag{6}$$

where c is the *damping constant* and k is the *spring constant*. This leads to:

$$x(t) = e^{-\gamma t}\left(A_1 e^{\sqrt{\gamma^2-\omega_0^2}t} + A_2 e^{-\sqrt{\gamma^2-\omega_0^2}t}\right) \tag{7}$$

where $\gamma = \frac{c}{2m}$ is the *damping factor* with mass m, A_1 and A_2 are the coefficients, and $\omega_0^2 = \frac{k}{m}$. For an under-damped oscillator, this gives $x(t) = Ae^{-\gamma t}cos[\omega_1 t - \phi]$ with amplitude A, phase constant ϕ, and angular frequency $\omega_1 = \sqrt{\omega_0^2 - \gamma^2}$.

Thus, the damping is determined by γ, which is determined by the coefficients, and, as in systems theory, x need not only be the position. This damping is pictured in Figure 4 along with the envelope, which is given by $Ae^{-\gamma t}$. This gives the Hamiltonian for the damped oscillator as [11]:

$$H(x,p) = \frac{p^2}{2m} + \frac{1}{2}m\omega_0^2 x^2 = \frac{1}{2}m\dot{x}^2 + \frac{1}{2}mx^2 \tag{8}$$

The change in energy is then given as $\frac{dH}{dt} = -c\dot{x}^2$. The under-damped oscillator is also shown in the (x,v), (H,t), and $(\frac{dH}{dt},t)$ modified phase space plots in Figure 4.

The *damping ratio* is defined as $\zeta = \frac{c}{2m\omega_0} = \frac{\gamma}{\omega_0}$ and determines whether the damping is critical, under-, or over-damping. *Logarithmic decrement*, δ, is used to find the

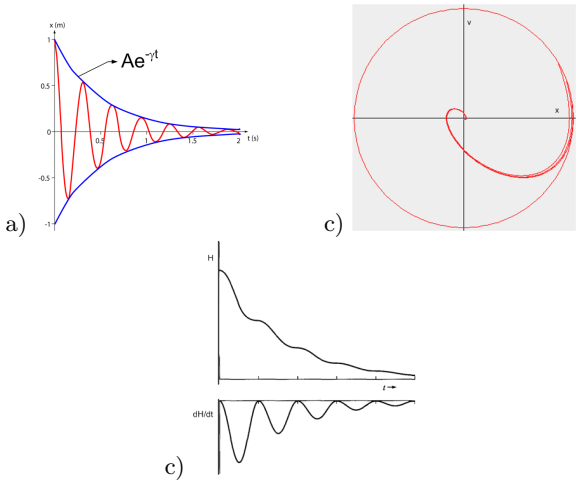

Figure 4: Under-damped oscillator in modified phase space plots: a) Under-damped Oscillation Envelope, b) (x,v) phase space plot, and c) (H,t) and $(\frac{dH}{dt},t)$ phase space plots.

damping ratio of an under-damped system in the time domain. The logarithmic decrement is the natural log of the amplitudes of any two successive peaks:

$$\delta = \frac{1}{n}\ln\frac{x_0}{x_n} \tag{9}$$

where x_0 is the greater of the two amplitudes and x_n is the amplitude of a peak n periods away. The damping ratio is then found from the logarithmic decrement as:

$$\zeta = \frac{1}{\sqrt{1+(\frac{2\pi}{\delta})^2}} \tag{10}$$

When two people are walking as in Figure 3, we can thus estimate the kind of oscillation via the damping factor, γ; in particular, we do an exponential fit to the average distance between the two people with respect to time in order to determine the damping factor. We then conclude SHO if $\gamma = 0$, driven resonant oscillator if $\gamma < 0$, and damping if $\gamma > 0$. To further qualify the damping, we utilize the *quality factor*, Q. We could use the *damping time*, $\tau = \frac{m}{\gamma}$, to define $Q = \omega_0\tau$ provided $\omega_0\tau \gg 2\pi$. Q is also defined as:

$$Q = \frac{1}{2\zeta} \tag{11}$$

We can then use (11) to determine the kind of damping as critical damping when $Q = \frac{1}{2}$; over-damping when $Q < \frac{1}{2}$; and under-damping when $Q > \frac{1}{2}$.

2.3.1 Application to Activity Modelling

For example, we model two people walking towards each other, as in Figure 3a, as an under-damped oscillator. We thus use the logarithmic decrement, δ, to estimate the damping ratio, ζ, by estimating n in (9). We use this damping ratio to compute the quality factor, Q, and determine the specific kind of damping. We can also use the damping ratio to get the angular frequency, ω, and then plot x vs ω or use ω directly. Finally, can also use average distance and

ω to get average velocity since $v = r\omega$. In fact, it is also possible to estimate Q from the (x, v) phase space plot, as in Figure 4a (under-damped oscillations spiral in while SHO is an ellipse, e.g.), or to use an exponential fit on the (H, t) or $(\frac{dH}{dt}, t)$ phase space plot, as in Figure 4b.

The proposed method is generalizable to many other cases of arbitrary motion interactions. For example, it can be applied to deal with intersections since intersections imply a transition from over-damped to under-damped or critically damped. In fact, it can be generalized to more than two objects by considering pairwise combinations and future work can consider more efficient methods than this combinatorial approach.

3. EXPERIMENTAL RESULTS

We experimented with videos consisting of people, vehicles, and buildings, which encompasses a large class of possible activities. We used high-resolution and low-resolution video from standard datasets like the UCR Videoweb (`http://vwdata.ee.ucr.edu/`) and VIVID (`https://www.sdms.afrl.af.mil/request/data_request.php`) datasets. We also assumed tracking and basic object-detection to be available. We utilized these (x,y,t) tracks to compute the Kinetic (T) and Potential (U) energies of the objects (mass can be idealized to unity or computed from shape). The distance and velocity vectors derived from the tracks are thereby used to compute both the HES curves, S-Metric, and MOPA.

3.1 Single-Object Activity Modelling with MRPS

Here we first show an example of the characteristic HES curves and apply them to tracking three cars from the UCR Videoweb dataset, where two cars maintain distance and one starts off together with them and then veers away, as shown in the frame in Figure 5. Since it involves more than two objects, we could then utilize the S-Metric to help characterize the gist of this system. For this experiment, we see the HES vs. Time curves for the two cars which follow all the way are highly correlated while the curve for the third car is not; similarly, the S-Metric (Table 1) calculated for them shows the coupling between Car 1 and Car 2 and the non-coupling between the others.

3.2 Modelling Paired Activities with MRPS

In this section, we show the results of modelling a paired activity from the UCR Videoweb dataset. In Figure 6a, we see three representative samples from a video of a person walking to their car. We model this as a paired activity where the stationary track of the car finally intersects the dynamic track of the person. In Figure 6b, we plot the average distance between the two tracks with respect to time and then do an exponential fit to that curve. The results are analyzed further in Table 2, where we see the analysis of Section 2.3 applied to the video and chart represented in Figure 6. We find a value for γ of 0.007, which indicates Damping; subsequent analysis yields a Q-factor of 37.878,

S(1,2) = 0.05382	S(1,3) = 0.56237	S(2,3) = 0.63720

Table 1: S-Metric Distance between the three cars shows coupling between Car 1 and Car 2 (with a small distance) and the non-coupling between the others (showing larger distances).

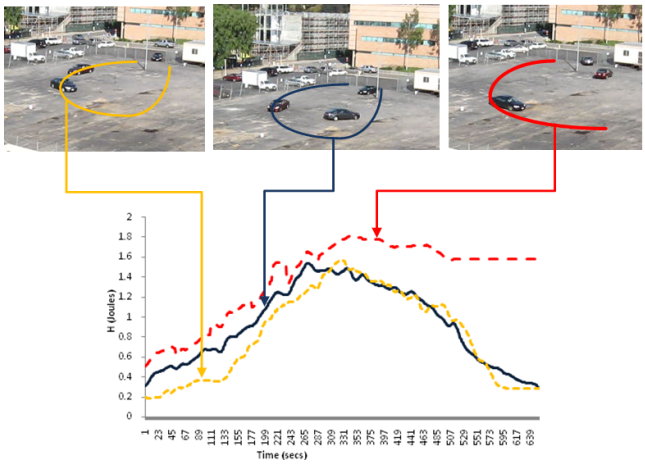

Figure 5: Two cars following; the first car, trajectory in orange, is the lead car and executes a U-turn; the second car, trajectory in blue, follows it and also makes a U-turn; the third car, trajectory in red, follows it for a while and then turns away.

thus indicating Under-Damping and showing that the two tracks eventually converge.

In addition, we show two people running away from each other and towards a car in Figure 7. As can be seen in Figure 7b, we plot the average distance between the two people with respect to time and then do an exponential fit to that curve to see a driven oscillator, while in Figure 7c and Figure 7d, we see the motion of each person running towards the car as being a damped oscillator.

3.2.1 *Advantages*

The advantages of our approach over a simple linear fit are manifold. One of the main is the robustness in tracking: short-term tracking errors would not affect our method since we fit to a model. In addition, the utilization of the Q-factor lets us determine the extent of the motion (if an object is headed for another, this lets us characterize if they head directly there or meander and go back and forth, instead). Finally, the formalism afforded by our method provides a framework that is extensible with more complex models to an even wider variety of situations and domains.

3.2.2 *Application to Activity Modeling*

For example, we model two people walking towards each other, as in Figure 3a, as an under-damped oscillator. We thus use the logarithmic decrement, δ, to estimate the damp-

Factor	Value	Result
γ	0.007	Damping
δ	0.083	
ζ	0.013	
Q-Factor	37.878	Under-Damping

Table 2: Person-Car Paired Activity Values. We find a value for γ of 0.007, which indicates Damping; subsequent analysis yields a Q-factor of 37.878, thus indicating Under-Damping and showing that the two tracks eventually converge.

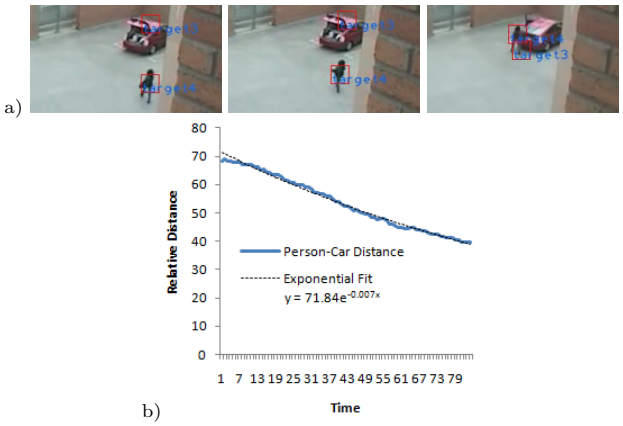

Figure 6: Person-Car Paired Activity modelling. In a) we see three representative samples from a video of a person walking to their car. We model this as a paired activity where the stationary track of the car finally intersects the dynamic track of the person. In b) we plot the average distance between the two tracks with respect to time and then do an exponential fit to that curve. The results are analyzed further in Table 2.

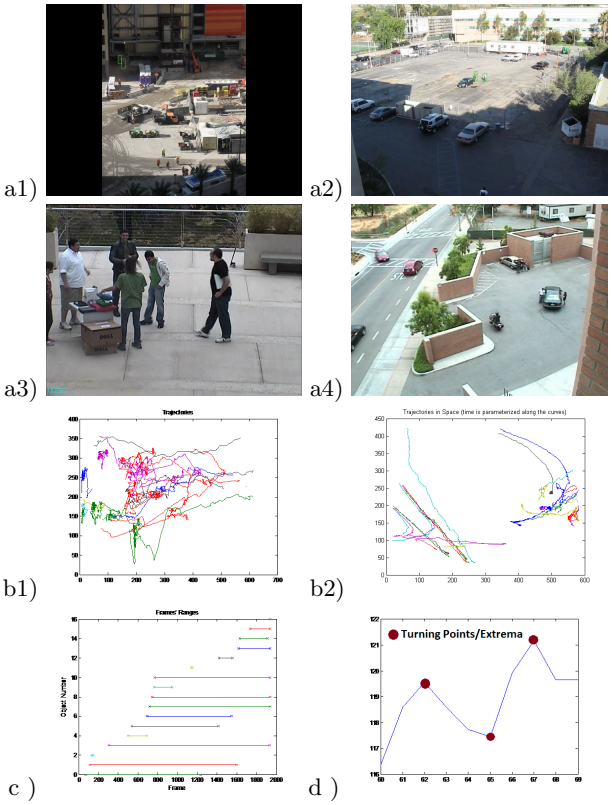

Figure 8: Multi-Object Activity Modeling. a) Sample frames representing complex interactions captured by our approach: construction site, courtyard, and parking lot; (b) Sample trajectories in space (time is parameterized along the curve) for the courtyard and parking lot; (c) Temporal overlap of objects' trajectories; (d) Turning points for a single trajectory.

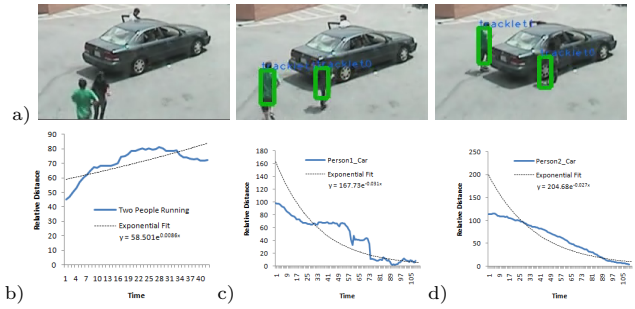

Figure 7: Two People and A Car: in a) we see three representative samples from a video of two people running away from each other and towards a car. In b) we plot the average distance between the two people with respect to time and then do an exponential fit to that curve to see a driven oscillator while in c) and d) we see the motion of each person running towards the car is modeled as a damped oscillator.

ing ratio, ζ, by estimating n in (9). We use this damping ratio to compute the quality factor, Q, and determine the specific kind of damping. We can also use the damping ratio to get the angular frequency, ω, and then plot x vs ω or use ω directly. Finally, we can use average distance and ω to get average velocity since $v = r\omega$. In fact, it is also possible to estimate Q from the (x, v) phase space plot, as in Figure 4a (under-damped oscillations spiral in SHO is an ellipse, e.g.), or to use an exponential fit on the (H, t) or $(\frac{dH}{dt}, t)$ phase space plot, as in Figure 4b.

3.2.3 *Generalization*

The proposed method is also generalizable to many other cases of arbitrary motion interactions. For example, it can be applied to deal with intersections since intersections imply a transition from over-damped to under-damped or critically damped. In fact, it can be generalized to more than two objects by considering pairwise combinations (as seen in Figure 8 and Figure 7) and future work can consider more efficient methods than this combinatorial approach.

This generalization can also be extended to more complex interactions, as shown in Figure 8a. Here, we see interactions that are not atomic, direct interactions; instead, we might observe interactions like people milling together, where they

Activity	Precision	Recall	Total Fetched	True Positive	Ground Truth
Person Entering Building	1	1	9	9	9
Person Exiting Building	1	1	7	7	7
Person Entering Vehicle	0.9	0.9	11	10	10
Person Exiting Vehicle	1	1	6	6	6
People Walking Together	1	0.71	5	5	7
People Coming Together	0.86	0.86	7	6	6
People Going Apart	0.8	1	5	4	5

Table 3: Precision/Recall Values for database query and retrieval using the combined VIVID and UCR Videoweb database.

alternately approach and recede from each other. Similarly, they might interact with a static object, like a car or a building, by meandering around it, rather than approaching it directly and then becoming static in its vicinity. In video of activities in the "wild", as in the UCR Videoweb dataset, these are exactly the kinds of activities observed, as shown in Figure 8b, where we see how convoluted the trajectories of individual objects can seem. When two objects interact, there's a temporal overlap to their trajectories, as shown in Figure 8c, where we see the overlap of frames between all the objects in a scene.

Thus, in these complex situations, people don't generally exhibit the simple, direct motions examined in [5, 6]. Instead, there are multiple *turning points* of their motion as they might alternate between approaching each other or moving away from each other; these turning points are characterized as the extrema of the (r, t) plot between two objects, as shown in Figure 8d. Thus, in order to model these behaviours, we do a pathwise MOPA analysis between all the turning points by using the extrema of the path to indicate the turning points for each segment.

3.3 Activity Recognition in Video Search using the MRPS

We also apply our modelling methodology to activity recognition by testing it within a query-based retrieval framework. We use a combined database from the UCR Videoweb and VIVID datasets and the results are shown in Table 3, where we see the precision/recall values for this experiment. In addition to tracking, we also assumed basic object detection. As shown in the table, the detection rate for the activity "People Walking Together" is lower because, as can be seen in the video, the participants tend to walk towards or away from each other instead of continuing perfectly parallel to each other.

4. CONCLUSION

We propose a set of descriptors motivated by Hamiltonian mechanics that forms a physics-based framework and provides a structured approach for activity recognition that only requires tracks for the motion. The descriptors provide, loosely speaking, a measure of global energy over a sequence (S-metric), a time series representation of motion or energy of an object (HES), and representation of relative position of pairs of interacting objects, characterized as damped oscillatory motion (MOPA). These descriptors form the Multi-Resolution Phase Space (MRPS) descriptor set. All measures can be computed directly form (x, y, t) trajectories of each interacting object.

Acknowledgement: Both authors were partially supported by NSF grant IIS-0712253 and the DARPA VIRAT program. The first author was also partially supported by NSF Award 1019343/Sub Award CIF-B-17.

5. REFERENCES

[1] P. Turaga, R. Chellappa, V. S. Subrahmanian, and O. Udrea. Machine recognition of human activities: A survey. *CSVT*, 2008.

[2] M.S. Ryoo and J.K. Aggarwal. Spatio-temporal relationship match: Video structure comparison for recognition of complex human activities. *ICCV*, 2009.

[3] M. Hu, S. Ali, and M. Shah. Detecting global motion patterns in complex videos. In *ICPR*, 2008.

[4] R.J. Sethi, A.K. Roy-Chowdhury, and S. Ali. Activity recognition by integrating the physics of motion with a neuromorphic model of perception. WMVC, 2009.

[5] N. Oliver, B. Rosario, and A. Pentland. A bayesian computer vision system for modeling human interactions. *ICVS*, 1999.

[6] U. Gaur, B. Song, and A. Roy-Chowdhury. Query-based retrieval of complex activities using "strings of motion-words". *WMVC*, 2009.

[7] J. R. Taylor. *Classical Mechanics*. University Science Books, 2005.

[8] R.J. Sethi and A.K. Roy-Chowdhury. The human action image and its application to motion recognition. *ICVGIP*, 2010.

[9] R.J. Sethi, A.K. Roy-Chowdhury, and A. Veeraraghavan. *Multibiometrics for Human Identification*, chapter Gait Recognition Using Motion Physics in a Neuromorphic Computing Framework. Cambridge University Press, 2010.

[10] L. Landau and E. Lifshitz. *Course of Theoretical Physics: Mechanics*. 3rd edition, 1976.

[11] G.R. Fowles and G.L. Cassiday. *Analytical Mechanics*. Brooks Cole, 6th edition, 2004.

APPENDIX

A. PROOF ACTION IS A NORM

Given the Action for a free particle ($U = 0$):

$$\begin{aligned} S_f &= \int_{t_a}^{t_b} L\left(q, \dot{q}, t\right) dt = \tfrac{1}{2} m \tfrac{\left(x_b - x_a\right)^2}{t_b - t_a} \\ &= \tfrac{1}{2} m v^2 \left(t_b - t_a\right) = \tfrac{1}{2} m v^2 \Delta t \end{aligned} \tag{12}$$

We want to prove that $\sqrt{S_f(m, v, \Delta t)} = \|v\|_{m,\Delta t}$ is a

norm on the vector space $\mathbb{R}^3$of the velocities. We have to prove that the following holds:

1. $\|v\|_{m,\Delta t} \geq 0, \|v\|_{m,\Delta t} = 0$ if and only if $v = 0$
2. $\|\lambda v\|_{m,\Delta t} = |\lambda| \, \|v\|_{m,\Delta t}$
3. Given two free particles, (m_1, v_1) and (m_2, v_2), and the system made up by the two, (m_{TM}, v_{CM}), where m_{TM} is their Total Mass and v_{CM} is the velocity of their Center of Mass:

$$\|v_{CM}\|_{m_{TM},\Delta t} \leq \|v_1\|_{m_1,\Delta t} + \|v_2\|_{m_2,\Delta t} \tag{13}$$

Properties 1 and 2 are trivially true. In order to prove Property 3, the **triangle inequality**, let (m_1, v_1) and (m_2, v_2) be two free particles. The system of the two particles is characterized by its total mass (TM), its center of mass (CM), and the velocity of its center of mass:

$$\begin{array}{l} m_{TM} = m_1 + m_2, \\ xCM = \frac{m_1 \vec{x}_1 + m_2 \vec{x}_2}{m_1 + m_2}, \\ vCM = \frac{m_1 \vec{v}_1 + m_2 \vec{v}_2}{m_1 + m_2} \end{array} \tag{14}$$

and the Action of the two particles, from (12), considered as one gives the norm:

$$\|v_{CM}\|_{m_{TM},\Delta t} = \sqrt{\frac{m_{TM} \vec{v}_{CM}^2}{2} \Delta t} \tag{15}$$

In order to prove the triangle inequality, we need:

$$\begin{array}{rcl} \|v_{CM}\|^2_{m_{TM},\Delta t} & = & \frac{m_{TM} \vec{v}_{CM}^2}{2} \Delta t \\ & = & \frac{1}{2}(m_1+m_2)\left(\frac{m_1 \vec{v}_1 + m_2 \vec{v}_2}{m_1+m_2}\right)^2 \Delta t \\ & = & \frac{1}{2} \frac{m_1^2 \vec{v}_1^2 + 2 m_1 m_2 \vec{v}_1 \cdot \vec{v}_2 + m_2^2 \vec{v}_2^2}{m_1+m_2} \Delta t \end{array} \tag{16}$$

and

$$\begin{array}{l} \left(\|v_1\|_{m_1,\Delta t} + \|v_2\|_{m_2,\Delta t}\right)^2 \\ = \frac{m_1 \vec{v}_1^2}{2} \Delta t + \sqrt{m_1 m_2}\, v_1 v_2 \Delta t + \frac{m_2 \vec{v}_2^2}{2} \Delta t \end{array} \tag{17}$$

Their difference gives:

$$\begin{array}{l} \left(\|v_1\|_{m_1,\Delta t} + \|v_2\|_{m_2,\Delta t}\right)^2 - \|v_{CM}\|^2_{m_G,\Delta t} \\ = \frac{m_1 \vec{v}_1^2}{2} \Delta t + \sqrt{m_1 m_2} v_1 v_2 \Delta t + \frac{m_2 \vec{v}_2^2}{2} \Delta t \\ \quad - \frac{1}{2} \frac{m_1^2 \vec{v}_1^2 + 2 m_1 m_2 \vec{v}_1 \cdot \vec{v}_2 + m_2^2 \vec{v}_2^2}{m_1+m_2} \Delta t \\ = \frac{m_1 m_2 \vec{v}_1^2 + m_1 m_2 \vec{v}_2^2 + 2(m_1+m_2)\sqrt{m_1 m_2} v_1 v_2 - 2 m_1 m_2 \vec{v}_1 \cdot \vec{v}_2}{2(m_1+m_2)} \Delta t \end{array} \tag{18}$$

The difference is positive because:

$$m_1 m_2 \vec{v}_1^2, \; m_1 m_2 \vec{v}_2^2 > 0 \tag{19}$$

and

$$2(m_1 + m_2)\sqrt{m_1 m_2} v_1 v_2 - 2 m_1 m_2 \vec{v}_1 \cdot \vec{v}_2 > 0 \tag{20}$$

with strict inequality, because of the *Cauchy-Schwarz inequality*:

$$v_1 v_2 \geq \vec{v}_1 \cdot \vec{v}_2 \tag{21}$$

and, because the *arithmetic mean* is greater than the *geometric mean*, we have: $m_1 + m_2 \geq 2\sqrt{m_1 m_2} > \sqrt{m_1 m_2}$. Substituting these in and expanding, gives:

$$\begin{array}{l} 2(m_1 + m_2)\sqrt{m_1 m_2} v_1 v_2 - 2 m_1 m_2 \vec{v}_1 \cdot \vec{v}_2 \\ = 2\sqrt{m_1 m_2}((m_1 + m_2) v_1 v_2 - \sqrt{m_1 m_2} \vec{v}_1 \cdot \vec{v}_2) \\ > 2\sqrt{m_1 m_2}(\sqrt{m_1 m_2} v_1 v_2 - \sqrt{m_1 m_2} \vec{v}_1 \cdot \vec{v}_2) \\ = 2 m_1 m_1 (v_1 v_2 - \vec{v}_1 \cdot \vec{v}_2) \geq 0 \end{array} \tag{22}$$

Putting it all together, this finally yields:

$$\left(\|v_1\|_{m_1,\Delta t} + \|v_2\|_{m_2,\Delta t}\right)^2 - \|v_{CM}\|^2_{m_{TM},\Delta t} > 0 \tag{23}$$

And so, we have the triangle inequality:

$$\|v_{CM}\|_{m_{TM},\Delta t} < \|v_1\|_{m_1,\Delta t} + \|v_2\|_{m_2,\Delta t} \tag{24}$$

Thus, this proves that

$$\|v_{CM}\|_{m_{TM},\Delta t} < \|v_1\|_{m_1,\Delta t} + \|v_2\|_{m_2,\Delta t} \tag{25}$$

is a norm and therefore it induces a distance in the vector space of the velocities. Although it can be shown to apply to situations where the potential is not zero, the exact nature of this distance needs to be verified and explored further.

A.1 Additivity of Actions

To prove the additivity of Actions, we start off by computing the S-Metric for two objects by first constructing the combined Action for the two objects, S_{12}. Again under the assumption of $U = 0$, we start off by using the S for one object, as shown in (1). From this, we compute the Action for both objects by first constructing their Lagrangian:

$$L_{12} = \frac{1}{2} m_1 v_1^2 + \frac{1}{2} m_2 v_2^2 \tag{26}$$

This leads to a combined Action for the two objects:

$$\begin{array}{l} S_{12} = \int_{t_a}^{t_b} L(q, \dot{q}, t) dt \\ = \int_{t_a}^{t_b} \frac{1}{2} m_1 \left(\frac{x_{1,b} - x_{1,a}}{t_b - t_a}\right)^2 + \frac{1}{2} m_2 \left(\frac{x_{2,b} - x_{2,a}}{t_b - t_a}\right)^2 dt \\ = \frac{1}{2} m_1 \frac{(x_{1,b} - x_{1,a})^2}{t_b - t_a} + \frac{1}{2} m_2 \frac{(x_{2,b} - x_{2,a})^2}{t_b - t_a} = S_1 + S_2 \end{array} \tag{27}$$

Thus showing the combined Action is just the sum of the individual Actions:

$$S_{12} = S_1 + S_2 \tag{28}$$

where S_{12} is used as the S-Metric for composite systems.