The Human Action Image and its Application to Motion Recognition

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

Recognizing a person's motion is intuitive for humans but represents a challenging problem in machine vision. In this paper, we present a multi-disciplinary framework for recognizing human actions. We develop a novel descriptor, the **Human Action Image** (HAI), a physically-significant, compact representation for the motion of a person, which we derive from Hamilton's Action.[1] We prove the additivity of Hamilton's Action in order to formulate the HAI and then embed the HAI as the Motion Energy Pathway of the Neurobiological model of motion recognition. The Form Pathway is modelled using existing low-level feature descriptors based on shape and appearance. Finally, we propose a **Weighted Integration** (WI) methodology to combine the two pathways via statistical Hypothesis Testing using the bootstrap to do the final recognition. Experimental validation of the theory is provided on the well-known Weizmann and USF Gait datasets.

Keywords

Hamilton's Action, motion recognition, integration

1. INTRODUCTION

Interpreting how people walk is intuitive for humans. From birth, we observe physical motion in the world around us and create perceptual models to make sense of it. Neurobiologically, we invent a framework within which we understand and interpret human activities like walking [2]. Analogously, in this paper, we develop a computational model that seeks to understand human motion from its neural basis to its physical essence. Human motion, including gait, the study of the motion of the walking style of humans, has been examined via motion methods in computer vision [3, 4, 5, 6].

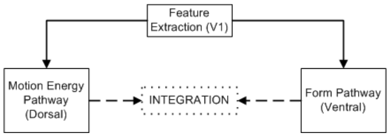

Figure 1: Feature extraction in V1 and then division along Motion Energy Pathway (Dorsal) and Form/Shape Pathway (Ventral)

In this paper, we develop the Human Action Image (HAI), a physically-significant, compact representation for the motion of a person, which we derive from first principles in physics. We embed the HAI within a neurobiologically-inspired model for the final motion recognition, thus providing a unifying framework for the analysis of human motion.

2. RELATED WORK & CONTRIBUTIONS

Recent work in Neurobiology [7, 8] suggests the brain examines both the form aspects of motion (e.g., shape, colour, orientation, etc.) as well as the motion energy (the kinematics and dynamics) when it attempts motion recognition. Visual processing in the brain, as shown in Figure 1, bifurcates into two streams at V1: a Dorsal Motion Energy Pathway and a Ventral Form/Shape Pathway [8, 9]. This neural basis for motion recognition has garnered much attention of late and [10] used a pseudo-Hamiltonian as the equivalent of the Motion Pathway and a Multiple Hypothesis Testing model as the final integration for the two pathways.

However, that Multiple Hypothesis Testing approach is a preliminary framework that is not able to model the variabilities in the inference process of each pathway; our methodology in this paper, on the other hand, incorporates a sampling-based approach using the bootstrap. In this work, we focus solely on the problem of analyzing the motion of an individual and propose a computational model for integration as well as a physics-based signature which has a strong physical significance and is a generalization of much previous work in human motion and gait analysis [3, 4, 5], as discussed below.

In the Appendix, we prove the additivity of Hamilton's Action and we use the additivity of the Action to develop the **Human Action Image (HAI)**, a spatio-temporal gait representation in which we create an average silhouette image that assigns an intensity value to each point on a person's contour. In addition, it can be shown that the Action is invariant to affine transforms; this allows for moderate view and scale invariance of HAI. HAI thus generalizes the

[*] Corresponding author

[1] Action can refer to both the usual Computer Vision meaning (as primitives of activities humans perform), as well as its use in the Physics community as Hamilton's Action [1].

ICVGIP '10, December 12-15, 2010, Chennai, India
Copyright 2010 ACM 978-1-4503-0060-5 ...$10.00.

motion analysis approaches of Motion Energy Image (MEI) [3], Motion History Image (MHI) [4], and Gait Energy Image (GEI) [5], which are widely used in gait recognition and represent an integration of image intensities over an image sequence, to a physics-based, compact representation, the HAI.

Our HAI can be used to recognize individuals on the basis of their gait as well as human actions, in general; therefore, it is a descriptor for human motion as well as gait. HAI unifies and extends ideas from MEI, MHI, and GEI and encapsulates the dynamic motion element of gait; thus, we use our physics-based HAI to represent the Motion Energy Pathway of the Neurobiological model of motion recognition, which provides a unifying framework for gait recognition.

We also propose a computational model for Integration, **Weighted Integration** (WI), which does statistical Hypothesis Testing using the *bootstrap* [11]. The bootstrap is used to ensure reasonable limits and allows WI to make the final Integration/gait recognition decision. WI also ensures that the Integration does no worse than either of the two pathways individually. The bootstrap is used to find the variance of a statistic on a sample; the statistic, in our case, is the quantiles. Our WI approach also builds upon recent work in the neurobiological community, which shows the dorsal and ventral processes could be integrated through a process of feature integration [12] or biased competition [13, 14] as originally outlined by [15]. Also, in the Appendix, we prove the additivity of Hamilton's Action which allows us to develop the HAI from the Action and to define distance measures on the HAI, our representation for the Motion Energy Pathway.

In addition to motion analysis, the Neurobiological model of motion recognition requires a Form Pathway. Most modern approaches in activity recognition also involve some kind of image analysis of form. In fact, image analysis of the form (based on shape, colour, orientation, etc.) is a well-known area in activity recognition [16, 17]. Our methodology provides flexibility in the Form Pathway since new approaches in low-level feature extraction can be employed easily within our framework. Indeed, Form can be represented as not just shape but any method like Bag of Video Words, Spatio-temporal Interest Points, etc., since we examine all of human motion with HAI.

For the present work, we use the methodology from [18] for the Form Pathway. It presents an approach for comparing two sequences of deforming shapes using both parametric models and nonparametric methods, where we use the latter. They apply this algorithm for gait-based human recognition on a subset of the USF dataset by exploiting the shape deformations of a person's silhouette as a discriminating feature and they also provide results for motion recognition. Significant effort has been devoted to the study of human gait, driven by its potential use as a biometric for person identification, as the comprehensive review on gait recognition found in [19] shows and any of these pre-existing approaches can also be used in the Form Pathway.

Our main contributions are thus:

- Development of a novel spatio-temporal human motion descriptor, HAI
- Development of a general framework to recognize human actions that utilizes WI integration via the bootstrap for sensitivity analysis in the integration of motion and form information
- Proof of Additivity of Hamilton's Action

3. HUMAN ACTION IMAGE

In this section, we introduce a special extension of the Hamiltonian framework as applied to the problem of gait recognition. Compact, image-based representations of gait have been an active area of research, where MHI, MEI, and GEI are three popular descriptors [3, 4, 6]. Extending current approaches that use MHI, MEI, and GEI, as well as the analysis of the dense optical flow by [20], we develop a spatio-temporal gait representation, the *Human Action Image* (*HAI*), which builds upon all three of these but is also based upon the fundamental Hamilton's Principle of Least Action. Hamilton's Principle of Least Action is built upon the idea of the Action of a system observed in video and is usually denoted as:

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

with q, the generalized coordinates [2], and L, in this case, the *Lagrangian* which, for a conservative system, is defined as:

$$L = T - U \tag{2}$$

with T (the Kinetic Energy) and U (the Potential Energy) derived directly from the trajectories, (x, y, t), as described in Section 4.1. Please see [21] for more on Hamilton's Action as well as the Sethi Metric (S-Metric). We use this to demonstrate the additivity of Hamilton's Action (in the Appendix), which we employ in the development of the HAI and its distance measure. However, before describing the HAI, it would be useful to briefly review the basic concepts behind MEI, MHI, and GEI.

MHI and MEI were proposed by [4] as formulations for human movement recognition. Both MEI and MHI are vector-valued images where the vector value at each pixel is a function of the motion properties at that particular location in an image sequence. MEI is a binary image which represents *where* motion has occurred in an image sequence:

$$MEI_\tau(x, y, t) = \bigcup_{i=0}^{\tau-1} D(x, y, t-i) \tag{3}$$

where $D(x, y, t-i)$ is a binary sequence indicating regions of motion, τ is the length of time, t is a particular moment in time, and (x, y) are the values of the 2D image coordinates. In similar fashion, MHI is a grey-level image which represents *how* a motion region in the image is moving:

$$MHI_\tau(x, y, t) = \begin{cases} \tau, \text{ if } D(x, y, t) = 1; \\ max\{0, MHI_\tau(x, y, t-1) - 1\}, otherwise. \end{cases} \tag{4}$$

Similarly, GEI [6] is a robust, widely used spatio-temporal gait descriptor for gait recognition. GEI builds upon the

[2] Generalized coordinates are the configurational parameters of a system; the natural, minimal, complete set of parameters by which you can completely specify the configuration of the system.

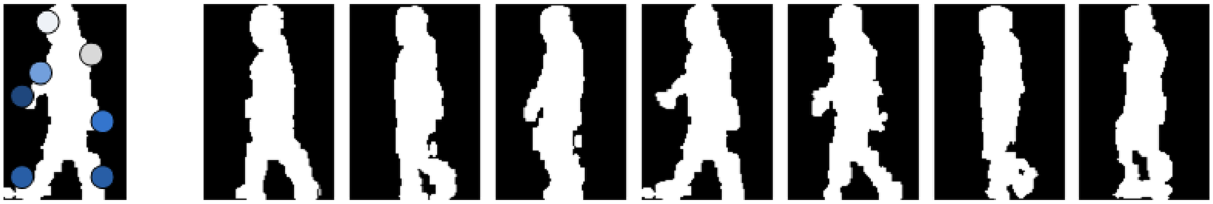

Figure 2: Examples of the Human Action Image (HAI) formed by averaging the row of silhouettes, with darker blues representing higher Action values and lighter blues representing lower Action values for points on the contour

approach of [4], who proposed MEI and MHI formulations for general human movement recognition. Both MEI and MHI assign a value to each pixel as a function of the motion properties at that location in an image sequence. GEI also creates an average silhouette image that assigns an intensity value to each pixel; it does so by starting with a size-normalized and horizontally-aligned binary silhouette, $B(x, y, t)$, and defines a grey-level GEI, $GEI(x, y)$, as:

$$GEI(x,y) = \frac{1}{N}\sum_{t=1}^{N} B(x,y,t) \tag{5}$$

where N is the number of frames in a complete cycle of the sequence, t is the frame number of the sequence, and (x, y) are the 2D image coordinates. Although, in general, MEI and MHI are different motion representations than GEI, a correspondence between the binary version of GEI and a modified MEI can be shown [5].

In this work, we use the ideas behind GEI, MEI, and MHI as motivation to extend our physics-based approach and generalize them to a physically-significant **Human Action Image (HAI)**. The GEI is an averaged silhouette summed over the temporal sequence; Hamilton's Action is a similarly integrated quantity over a specific time interval, as shown in (1). We combine these ideas by computing a physically-relevant pseudo-Action for each point on the human silhouette contour or body parts in a given cycle (described in Section 4.1) as:

$$HAI(x,y) = HAI(q) = \frac{1}{N}\int_{t=1}^{N} L(q(t),\dot{q}(t),t)dt \tag{6}$$

where N is again the number of frames in a complete cycle and q and $\dot{q}$ are the generalized coordinate and generalized velocity, respectively (L is again the Lagrangian). Following the example of [5, 22], we measure the similarity between the gallery (training) and probe (test) templates of two gait sequences, HAI_g and HAI_p respectively, by calculating their distance as the normalized matching error:

$$\begin{aligned} D(HAI_g, HAI_p) =& \frac{\sum_{x,y}|HAI_g(x,y) - HAI_p(x,y)|}{\sqrt{\sum_{x,y} HAI_g(x,y)\sum_{x,y} HAI_p(x,y)}} \\ =& \frac{\sum_q |HAI_g(q) - HAI_p(q)|}{\sqrt{\sum_q HAI_g(q)\sum_q HAI_p(q)}} \end{aligned} \tag{7}$$

where $\sum_{x,y}|HAI_g(x,y) - HAI_p(x,y)|$ is the matching error between two HAIs (sum of the magnitudes of the difference between two HAIs) and $\sum_{x,y} HAI(x,y)$ is the total energy/action in a HAI.

Because HAI mirrors the GEI, MEI, and MHI formulations and representations so closely, all the extensions and proposed algorithms for them should be immediately extensible to HAI, as well. In addition, we can use distance or similarity measures computed using HAI directly in our Integration framework by combining that similarity distribution with one of the standard shape/form methodologies, as described in Section 4. We show an example of the HAI in Figure 2 and show experimental results of using the HAI as the motion pathway and shape sequence for the form pathway in the Integration in Section 5.

4. PROPOSED FRAMEWORK FOR HUMAN MOTION ANALYSIS

We cast our physics-driven compact representation of the gait of a person, the HAI, within the neurobiologically-inspired infrastructure for gait recognition, as shown in Figure 3, by integrating the motion signature of the person's gait (the HAI) for the Dorsal Pathway and shape features for the Ventral Pathway via the Integration module. As can be seen there, the task we have is to take a probe/test query, containing the motion of a subset of the people from the gallery/database, and match the motion of each person in the probe to the gallery/database. We start off by computing the HAI for the Motion Energy Pathway for each person in the probe. Simultaneously, we compute the shape information for the Form Pathway for each person in the probe. These are then compared, individually, with each person in the gallery. These normalized similarity measures are then sent to the Integration module which does Weighted Integration using the bootstrap.

4.1 The Motion Pathway: Human Action Image

Our approach is to get the tracks for each point on the contour of each person, from which we compute T (Kinetic Energy, KE) and U (Potential Energy, PE), and use that to get the HAI, as shown in Figure 4. Thus, we use the video to gain knowledge of the physics and use the physics to capture the Motion Energy of the person being observed via the HAI. In order to compute the HAI, we use the tracks from the video to compute the kinematic quantities that drop out of the Lagrangian formalism, thus giving a theoretical basis for their examination from (x, y, t).

We compute the Lagrangian and then the pseudo-Action as in [10] by using $T = \frac{1}{2}mv_o^2$ and $U = mg(y_b - y_a)$, derived directly from the trajectories, (x, y, t). Note that the Lagrangian can be computed in either the image plane, yielding the Image Lagrangian (which composes the pseudo-Action),

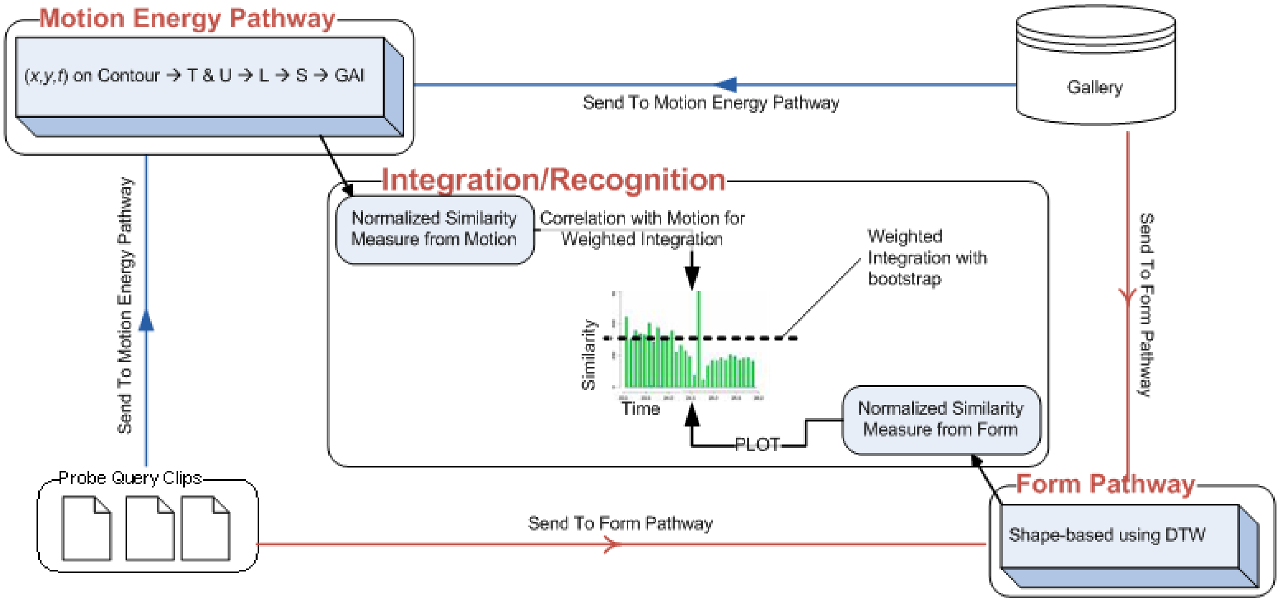
Motion Energy Pathway
(x,y,f) on Contour → T & U → L → S → GAI
Send To Motion Energy Pathway
Gallery
Integration/Recognition
Normalized Similarity Measure from Motion
Correlation with Motion for Weighted Integration
Weighted Integration with bootstrap
Similarity
Time
PLOT
Normalized Similarity Measure from Form
Send To Form Pathway
Send To Motion Energy Pathway
Probe Query Clips
Send To Form Pathway
Form Pathway
Shape-based using DTW

Figure 3: Proposed Framework for motion recognition by searching a database for a query: final recognition decision is made in the Integration module

or in the 3D world, giving the Physical Lagrangian (which composes the actual 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-Action and the more information we have about the objects in the video, the more physically significant this pseudo-Action becomes. Regardless, though, this pseudo-Action allows us to extract an abstract representation of the motion of the underlying physical systems we consider in video and allows us to build a physics-driven pseudo-Action to represent a video sequence.

The HAI does require tracking of contour points; however, adjacent points approximate the same trend and we are trying to capture the uniqueness in the trend in the pseudo-Actions we compute. This is exactly the same as for MHI, MEI, and GEI as our work generalizes that methodology using the Hamiltonian formulation; thus, the same assumptions and extensions applicable to MHI, MEI, and GEI apply to HAI, as well. On the other extreme, if more detailed information is available, we can compute more complex interactions between the points on a person's contour/body joints or the kinematics of the different body parts when we calculate the pseudo-Action, which can approximate the actual Action in the case of full knowledge. While the HAI will, in general, not be sufficient for complete activity recognition, it formalizes a first level of discrimination using only the motion information and provides a framework for theoretical extensions.

4.2 The Form Pathway: Shape-based Features

Since the Form Pathway is posited to have orientation detectors and also recognizes body shapes and colour [23], we use well-established methods in machine vision to calculate exactly these features in order to develop its computational representation. For the present work, we use shape features with Dynamic Time Warping (DTW). Also, since we are analyzing video, we consider the shape over a sequence of frames.

For modelling the sequence of shapes for an activity, we use shape features with DTW; in particular, we use the approach in [18], which presents a method for comparing two sequences of deforming shapes using both parametric models and nonparametric methods. In their approach, Kendall's definition of shape is used for feature extraction. Since the shape feature rests on a non-Euclidean manifold, they propose parametric models like the autoregressive model and autoregressive moving average model on the tangent space. The nonparametric model is based on DTW but they employ a modification of the DTW algorithm to include the nature of the non-Euclidean space in which the shape deformations take place. They apply this algorithm for gait-based human recognition on a subset of the USF dataset by exploiting the shape deformations of a person's silhouette as a discriminating feature and then providing recognition results using the nonparametric model for gait-based person authentication.

4.3 Integration

Given a set of distance scores of a probe sequence against all elements of the gallery, Hypothesis Testing lets us choose between the motion and form features or come up with a combination of them, with the bootstrap being used to find the variance of the quantiles on the sample. Building upon recent work in the neurobiological community, which shows the dorsal and ventral processes could be integrated through a process of feature integration or biased competition, we propose a computational model for the Integration of the two pathways by implementing this Integration in a statistical Hypothesis Testing framework, creating a *Weighted Integration (WI)* model (Ventral values are weighted by Dorsal values and does no worse than Ventral values), using the *bootstrap*, which is a better method than simple hypothesis testing since it allows resampling and is thus able to model the distributions. Although the exact mechanism of the In-

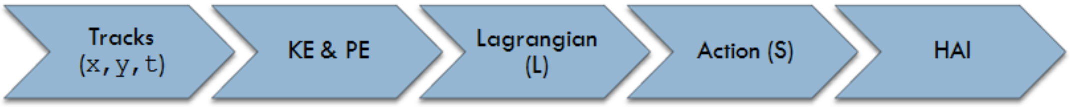

Figure 4: Tracks to Hamiltonian to HAI

tegration of these pathways is an open question in Neurobiology [7], we are motivated by the neurobiological models in the development of the WI, which provides a natural framework for the integration of image and motion components (however, in this paper, we do not claim that our method provides a model for the neurobiogical integration mechanism).

The bootstrap is used to find the variance of a statistic on a sample; the statistic, in our case, is the quantiles. After a sample is collected from an experiment, we can calculate a statistic on it (like the mean or quantiles, for example), and then use the bootstrap to figure out the variance in that statistic, e.g., via a Confidence Interval (CI). A CI is a range of values that tries to quantify the uncertainty in the sample and can be two-sided or one-sided, as shown in Figure 5; e.g., the 95% 2-sided confidence interval shows the bounds within which you find 95% of the population (similarly for the 1-sided upper and lower confidence bounds). Confidence intervals are also equivalent to encapsulating the results of many hypothesis tests [24].

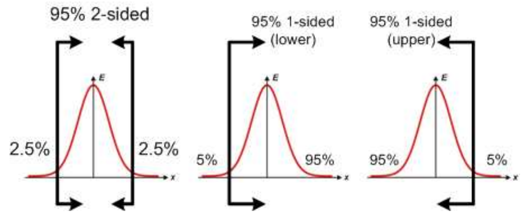

Figure 5: 2-sided and 1-sided Confidence Intervals (CI): the first diagram shows a 2-sided CI showing the confidence interval in the middle and the critical regions to the left and right; the second diagram shows a 1-sided lower bound with the critical region to the left; the final diagram shows a 1-sided upper bound with the critical region to the right; the E just indicates the mean expectation value.

The bootstrap itself works by re-sampling with replacement, as described in Figure 6. One way to estimate confidence intervals from bootstrap samples is to take the α and $1-\alpha$ quantiles of the estimated values, called bootstrap percentile intervals, where α is the standard significance level. For example, for the upper quantile, this confidence interval would then be given as $CI = (q^u_{lower}, q^u_{upper})$, with $lower = \lfloor N\alpha/2 \rfloor$ and $upper = N - lower + 1$, where N is the number of bootstrap samples and $(q^u_{lower}, q^u_{upper})$ are the lower and upper critical values of the bootstrap confidence interval bounds.

Weighted Integration uses a two-sided upper bound CI, where the Form values are weighted based on the Motion values; if the observed distance value of the Form and the Motion is lower than the lower distance quantile obtained from the bootstrap quantile analysis for both, then the value is set to 0; if either is higher than the upper quantile analysis, it is set to the max value; all other values are set to the unaltered Form value. In this way, WI ensures that it always does no worse than the Form. This can be inverted trivially to ensure that the Integration does no worse than either pathway individually.

5. EXPERIMENTAL RESULTS

In the experiments that follow, we show that the Integration mechanism helps reduce the search space (using the Weizmann dataset) and also helps with overall recognition when either the motion or the form (or both) pathways fail or under-perform. We show how, in the USF Gait dataset, although the form model performs well, when we integrate that with the motion energy computational model, it improves the overall performance; although both do reasonably well on their own, the integrated version does better than either alone.

We conduct experiments on the Weizmann dataset to demonstrate how the integration afforded by WI helps reduce the search space, as well as the hierarchical scheme for recognition, as shown in Figure 7. Also, the flexibility of the WI approach allows our framework to accomodate any method to compute the Form features. Please note, in these experiments, we have just used the basic form methodology since specialized feature sets is not the focus of this work. We can utilize any form methodology that allows us to generate similarity scores between a probe query and a video database of clips.

Also, these results should not be compared to absolute recognition scores but rather as the gain over the image-based approach and the pruning of the search space in our hierarchical approach. Since gait recognition and activity search in video is becoming a very important problem, we expect this work to be an important contribution in this direction.

The entire code for the project will be made available to the research community once the paper is accepted.

5.1 Experimental Background

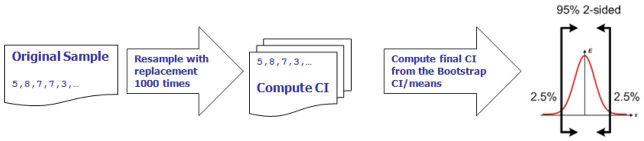

Figure 6: Overview of Bootstrap: the original sample is re-sampled (with replacement), say, 1000 times. In each re-sampling, a Confidence Interval is computed based on that sample. Eventually, the final Confidence Interval is estimated from either the Bootstrap Confidence Interval (on the CI computed on each re-sample) or the means (again, of the CI computed on each re-sample)

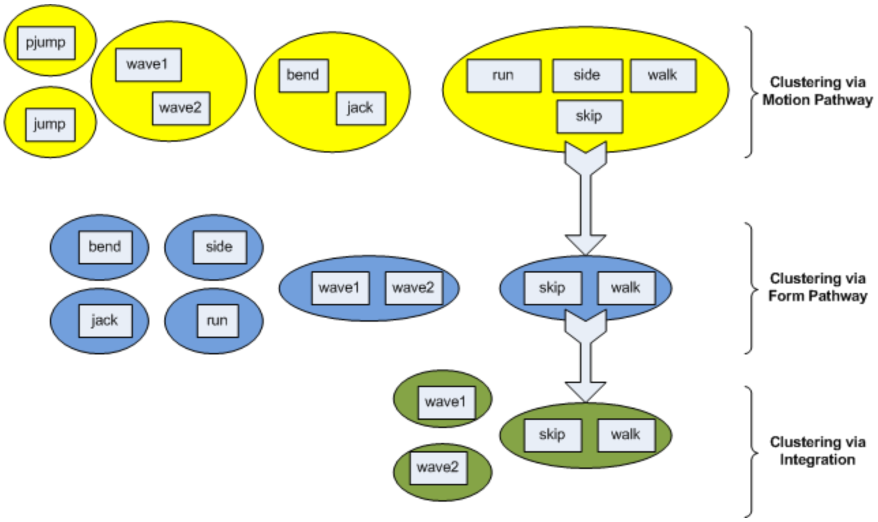

Figure 7: Data Clustering via Motion Pathway, then Form Pathway, and finally Integration. As seen here, the Motion correctly isolates pjump and jump; Form further clarifies bend, jack, side, and run; finally, Integration discerns wave1 and wave2, with skip and walk remaining grouped. Please also refer to Figure 9

For all of these experiments, tracking and basic object-detection was already available [25] and we utilized these (x,y,t) tracks to compute the Kinetic (T) and Potential (U) energies of each point on the contour of each person as detailed in Section 4.1 (mass can be idealized to unity or computed from shape and, when we assume gait is characterized only by the horizontal motion of the body, U is set to zero). The distance and velocity vectors derived from the tracks are thereby used to compute the HAI, which is then used as the Dorsal Pathway component of the framework.

We utilized shape (as defined in [18]) for the Form component. We then utilized Weighted Integration to bias the Ventral Pathway component with the Dorsal Pathway component and used the bootstrap to set the threshold for peaks in the distributions that might compete for selection/matching. We biased these peaks by doing pointwise multiplication with the Dorsal Pathway values computed earlier to make our final selections/matches. The results are then plotted as both heatmaps of the distance matrices as well as Cumulative Match Score (CMS) graphs, which plot probability vs. rank.

5.2 HAI for Activity Recognition

We now show the applicability of HAI and WI for human action analysis. We will show that it reduces the search space leading to a hierarchical search mechanism, which is a huge benefit when searching through a large database. In this case, we demonstrate on the Weizmann dataset, as in [18]. The Weizmann dataset (`http://www.wisdom.weizmann.ac.il/~vision/SpaceTimeActions.html`) consists of a database of 90 low-resolution (180 x 144, deinterlaced 50 fps) video sequences showing nine different people, each performing 10 natural actions. We analyze these using both shape methods [18] (as discussed in Section 4.2), as well as via the HAI. Using both procedures, we see the resulting similarity matrices in Figure 8 (a) and (b), respectively. Finally, in Figure 8 (c), we see the result of integrating via WI. In each of the distance matrices, both axes consist of the people grouped by the activity: bend, jack, jump, pjump, run, side, skip, walk, wave1, wave2. So the first nine rows are each person bending, the next nine rows are each person doing a jumping jack, etc. This clustering by the different methods is shown explicitly in Figure 7, where we see the Motion pathway correctly isolates pjump and jump; the Form pathway further clarifies bend, jack, side, and run; finally, Integration discerns wave1 and wave2, with skip and walk remaining grouped.

In addition, one of our main contributions is in combining shape and motion with a sensitivity analysis, which we show results in improvements over previous results as in Figures 8, 9, and 10. The experiments are meant to demonstrate that the hierarchical search our novel model affords prunes search results (c.f. Figure 7). This kind of analysis requires larger databases to fully show its efficacy and such a database is not available for the complex activities we consider. Thus, the improvement will be significantly better on a larger database as opposed to a small database like Weizmann. Also, our approach is not fine-tuned for any specific database and can be widely and generically applied. Finally, we can also increase the “burn-in” period for larger datasets, which should yield better results.

As can be seen in the matrices in Figure 8 and the diagrams in Figure 7, HAI alone, in Figure 8 (a), groups together bending and jumping jacks; partially groups the jumping sideways; fully groups jumping in place; confuses running, galloping sideways, skipping, and walking; and confuses waving 1 and waving 2. Form alone, in Figure 8 (b), groups bending and jumping jacks correctly; partially groups the jumping sideways; fully groups jumping in place; partially groups running; partially groups galloping sideways; confuses skipping and walking; and partially confuses waving 1 and waving 2.

The Integration, however, in Figure 8 (c), does better than both in most cases and no worse than the better method, form, in all cases. As can be seen, it groups bending and jumping jacks correctly; partially groups the jumping sideways; fully groups jumping in place; partially groups running; fully groups galloping sideways; confuses skipping and

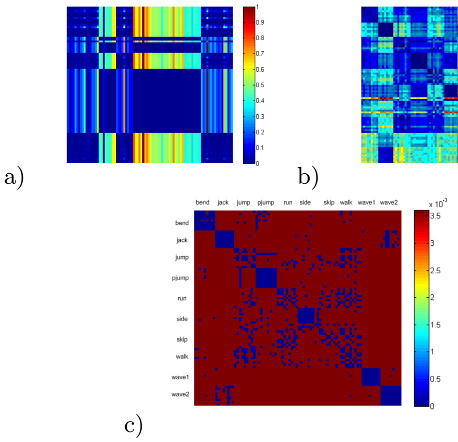

Figure 8: Similarity matrices on the Weizmann dataset for a) HAI only, b) Shape Methods only, and c) Integration using WI. Both axes consist of the people grouped by the activity: bend, jack, jump, pjump, run, side, skip, walk, wave1, wave2. So the first nine rows are each person bending, the next nine rows are each person doing a jumping jack, etc. In (c), we see the result of integrating via WI. As seen in the matrices, WI combines both pathways in such a way as to do no worse than either pathway by itself

walking; and fully groups waving 1 and waving 2. We thus show that the integrated combination works better than using only one source of information.

5.3 HAI for Gait Recognition

We utilize the same subset of the USF Gait dataset used in [18] in order to compare our results to previously published results; we then show how, although the form model performs well, when we integrate that with the motion energy computational model, it improves the overall performance. We experimented with videos from this subset of the standard USF gait dataset consisting of 67 people walking on different surfaces (grass and concrete) with different shoe types and different camera views. The Gallery contains video of four cycles of walking by all 67 people under standard conditions. There are also videos of different combinations of the 67 people (between 40 and 67) in the seven different probes, labelled Probe A to Probe G. The goal is then to compare the gait of each person in a probe with the gait of all the people in the Gallery to determine which one(s) it matches most closely. This is done for each of the probes in turn.

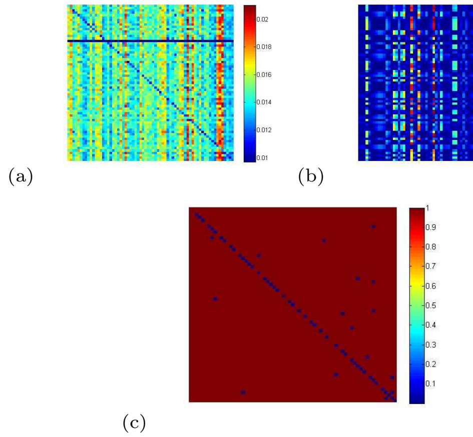

Figure 9: Similarity Matrices for USF Gait dataset examined using (a) Form Pathway, (b) Motion Pathway, and (c) the WI Integrated Framework on Probe A for all seven probes in the USF Gait. Although the form model performs well, when we integrate that with the motion energy computational model, it improves the overall performance as seen by the matching in (c). The overall CMS matching is shown in (d) and explained in Figure 7

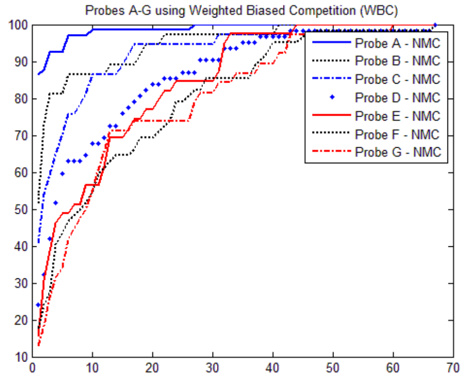

Figure 10: CMS Curves for the USF Gait dataset

The Motion Pathway is represented by the HAI for each person, as shown in Figure 2. The form component was calculated using the shape of the silhouettes and computing similarity using DTW in the shape space. We utilized WI to bias the Form component with the Motion component and then used the bootstrap to set the threshold for peaks in the distributions that might compete for selection/matching. The results are plotted as both distance matrices as well as Cumulative Match Score (CMS) graphs, which plot probability vs. rank; results are in Figures 9 and Figure 10. We also see the Integration approach consistently outperforms the Form Pathway approach alone, as seen in Figure 10. The singular exception is Probe B in rank 1; this is because WI favours the Form method more heavily than the Motion Energy Pathway method and, in this case, the Form method misses the real match and guesses matches that are far removed from the real match, as seen in the similarity matrix in Figure 9. Please note that although these results are specific to our Form approach, it is expected that similar improvements would be realized using other approaches.

6. CONCLUSION

We propose a novel spatio-temporal human motion descriptor, the Human Action Image (HAI), which is a natural extension of existing MEI, MHI, and GEI approaches. The HAI is derived from Hamilton's Principle of Least Action in classical physics and, in a way, rationalizes these prior approaches. The HAI is combined with shaped based features for human motion analysis. We then cast this HAI within the neurobiological model of motion recognition and propose a novel Integration mechanism in order to create a framework for gait recognition which we apply to real world datasets. This principled and coherent framework approach is evaluated on gesture (human activity within the Weizmann dataset) and gait databases (on a subset of the USF Gait dataset). The infrastructure we present in this work provides a structured approach to gait analysis using motion analysis neurobiological models within a single, unifying framework which mimics the processing in the dorsal and ventral pathways of the human brain. Our new framework is a general approach which uses Weighted Integration to give sensitivity analysis concurrent with extraction of a description of motion. We further prove the additivity of Hamilton's Action.

We see much room for future research, especially developing a more complex physical model for the HAI, deriving new distance measures for the HAI, and exploring alternative Integration strategies. We also intend to address robustness of our high-level approach to low-level errors in the tracks; techniques for addressing this include potentially creating a Stochastic HAI. Finally, we are considering other integration mechanisms, including using MCMC/DDMCMC approaches adapted for the Hamiltonian dynamics model.

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.

7. REFERENCES

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

[2] E. Kandel, J. Schwartz, and T. Jessell. *Principles of Neural Science.* McGraw-Hill Medical, USA, 4th edition, 2000.

[3] A. Bobick and J. Davis. An appearance-based representation of action. *ICPR*, 1996.

[4] A. Bobick and J. Davis. The recognition of human movement using temporal templates. *PAMI*, 2001.

[5] J. Han and B. Bhanu. Individual recognition using gait energy image. *Workshop on MMUA*, pages 181–188, December 2003.

[6] J. Han and B. Bhanu. Individual recognition using gait energy image. *PAMI*, 28:316–322, 2006.

[7] M.A. Giese and T. Poggio. Neural mechanisms for the recognition of biological movements and action. *Nature Reviews Neuroscience*, 4:179–192, 2003.

[8] R. Sigala, T. Serre, T. Poggio, and M. Giese. Learning features of intermediate complexity for the recognition of biological motion. In *ICANN*. Springer Berlin, 2005.

[9] H. Jhuang, T. Serre, L. Wolf, and T. Poggio. A biologically inspired system for action recognition. ICCV, 2007.

[10] 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.

[11] A.M. Zoubir and B. Boashash. The bootstrap and its application in signal processing. *IEEE Signal Processing Magazine*, 15:56–76, 1998.

[12] A. M. Treisman and G. Gelade. A feature-integration theory of attention. *Cogn. Psychol.*, 12:97–136, 1980.

[13] G. Deco and E. Rolls. Neurodynamics of biased competition and cooperation for attention: A model with spiking neurons. *J Neurophysiol*, pages 295–313, 2005.

[14] D. M. Beck and S. Kastner. Top-down and bottom-up mechanisms in biasing competition in the human brain. *Vision Research*, 2008.

[15] R. Desimone and J. Duncan. Neural mechanisms of selective visual attention. *Annu. Rev. Neurosci.*, 18:193–222, 1995.

[16] J.K. Aggarwal and Q. Cai. Human motion analysis: A review. *CVIU*, 1999.

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

[18] A. Veeraraghavan, A.K. Roy-Chowdhury, and R. Chellappa. Matching shape sequences in video with applications in human motion analysis. *PAMI*, 2005.

[19] M. Nixon, T. Tan, and R. Chellappa. *Human Identification Based on Gait.* Springer, 2005.

[20] J.J. Little and Boyd.J.E. Recognizing people by their gait : The shape of motion. *JCVR*, 1998.

[21] R.J. Sethi and A.K. Roy-Chowdhury. Physics-based activity modelling in phase space. *ICVGIP*, 2010.

[22] Z. Liu and S. Sarkar. Simplest representation yet for gait recognition: Averaged silhoutte. In *ICPR*, 2004.

[23] M. A. Giese. Neural model for biological movement recognition. In *Optic Flow and Beyond*, pages 443–470. Kluwer Academic Publihers, 2004.

[24] G. Kochanski. Confidence intervals and hypothesis testing, 02 2005.

[25] A. Kale, A.N Rajagopalan, Sundaresan.A., N. Cuntoor, A. Roy-Chowdhury, V. Krueger, and R. Chellappa. Identification of humans using gait. Sept. 2004.

APPENDIX

A. ADDITIVITY OF ACTIONS

To prove the additivity of Actions, we start off by computing the Sethi Metric (S-Metric) [21] 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:

$$L_{12} = \frac{1}{2}m_1 v_1^2 + \frac{1}{2}m_2 v_2^2 \tag{8}$$

This leads to a combined Action for the two objects:

$$\begin{aligned} &S_{12} = \int_{t_a}^{t_b} L(q,\dot{q},t)dt \\ &= \int_{t_a}^{t_b} \tfrac{1}{2}m_1 \left(\tfrac{x_{1,b}-x_{1,a}}{t_b-t_a}\right)^2 + \tfrac{1}{2}m_2 \left(\tfrac{x_{2,b}-x_{2,a}}{t_b-t_a}\right)^2 dt \\ &= \tfrac{1}{2}m_1 \tfrac{(x_{1,b}-x_{1,a})^2}{t_b-t_a} + \tfrac{1}{2}m_2 \tfrac{(x_{2,b}-x_{2,a})^2}{t_b-t_a} = S_1 + S_2 \end{aligned} \tag{9}$$

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

$$S_{12} = S_1 + S_2 \tag{10}$$

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