The Human Action Image

Ricky J. Sethi
UC Riverside
rickys@sethi.org

Amit K. Roy-Chowdhury
UC Riverside
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 first principles in physics using Hamilton's Action.[1] We embed the HAI as the Motion Energy Pathway of the latest Neurobiological model of motion recognition. The Form Pathway is modelled using existing low-level feature descriptors based on shape and appearance. Experimental validation of the theory is provided on the well-known Weizmann and USF Gait datasets.

1 Introduction

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 studied via motion methods in computer vision [1, 2, 3]. We provide a rigorous analysis for the motion of a person via Physics and Neurobiology and thus provide a unifying framework for the analysis of human motion. We develop the physics-based Human Action Image (HAI), a physically-significant, compact representation for the motion of a person, which we derive from first principles in physics. We then embed the HAI within a neurobiologically-inspired framework for the final motion recognition.

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

2 Related Work & Contributions

In this work, we 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. An example of the HAI is seen in Figure 1, which shows the HAI is 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. In addition, the invariance of Hamilton's Action to affine transforms [4, 5] allows for moderate view and scale invariance of HAI. HAI thus generalizes the motion analysis approaches of Motion Energy Image (MEI) [1], Motion History Image (MHI) [2], and Gait Energy Image (GEI) [3], 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.

The 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 also 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 [4, 5], thereby providing a unifying framework for gait recognition.

In addition to motion analysis, the Neurobiological model of motion recognition requires a Form Pathway. Image analysis of the Form (based on shape, colour, orientation, etc.) is a well-known area in activity recognition [6, 7] and, as the construction in [4, 5], Form can be represented as not just shape but any method like Bag

Figure 1. The Human Action Image (HAI).

of Video Words, Spatio-temporal Interest Points, etc., since we examine all of human motion with HAI.

3 Human Action Image

Compact, image-based representations of gait have been an active area of research, where MHI, MEI, and GEI are three popular descriptors [1, 2, 8]. Extending current approaches that use MHI, MEI, and GEI, as well as the analysis of the dense optical flow by [9], we develop a spatio-temporal gait representation, the *Human Action Image* (*HAI*), which builds upon all three of these but is also based upon Hamilton's Principle of Least Action from fundamental physics. Hamilton's Principle of Least Action, as detailed in [4, 5], is built upon the idea of the *Hamilton's Action* of a system, which 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:

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

where, T is the *Kinetic Energy* and U is the *Potential Energy*. 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 [2] 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 [8] is a robust, widely used spatio-temporal gait descriptor for gait recognition. GEI builds upon the approach of [2], 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 [3].

In this work, we use the ideas behind GEI, MEI, and MHI as motivation to extend our physics-based approach and generalizing them to a physically-significant HAI. The GEI is an averaged silhouetted 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 Hamilton's Action for each point on the human silhouette contour or body parts in a given cycle 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 [3, 10], 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:

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

[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.

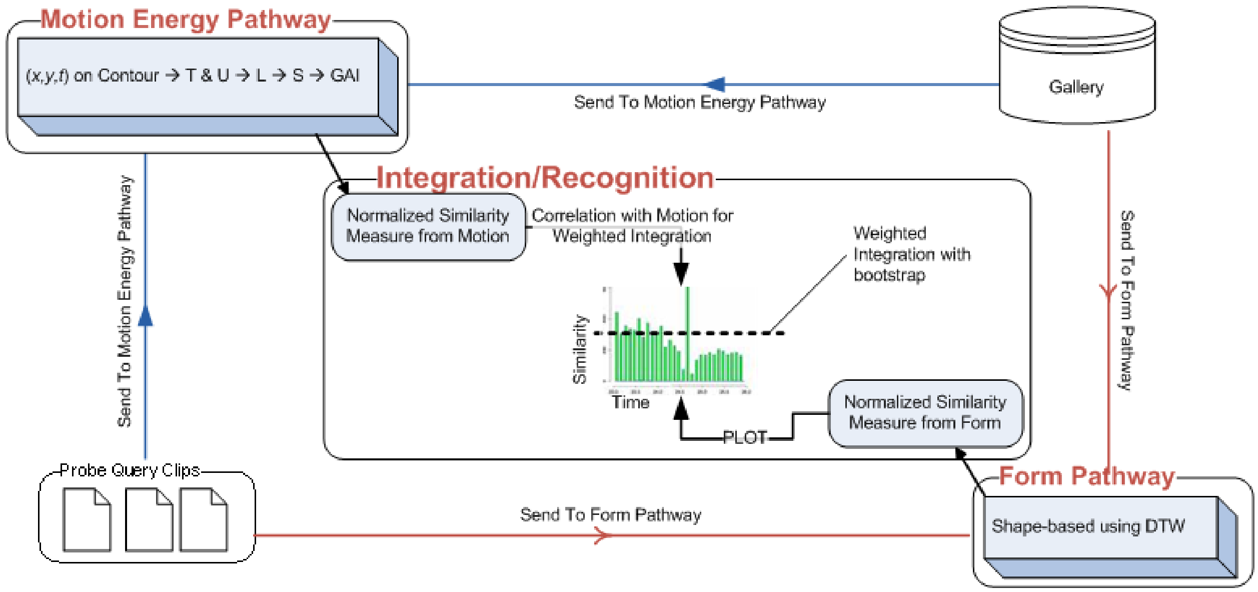

Figure 2. Motion recognition framework.

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. We next show an example of the HAI in Figure 1 and show experimental results of using the HAI as the motion pathway and shape sequence for the form pathway in the Integration.

4 Human Motion Analysis Framework

We cast our physics-driven compact representation for the gait of a person, the HAI, within the neurobiologically-inspired infrastructure for gait recognition, as shown in Figure 2. 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.

The Motion Pathway: Our approach is to get the tracks for each point on the contour of each person, from which we compute T and U, and use that to get the HAI, as shown in Figure 3. 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 examination of their energy from (*x,y,t*). On the other extreme, we can compute more complex interactions between the points on a person's contour/body joints or the kinematics of the different body parts. While it 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.

The Form Pathway: For the present work, we use the shape features with DTW we developed in [11], where we presented an approach for comparing two

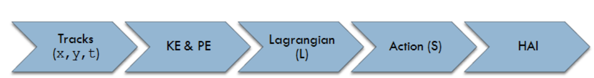

Figure 3. Tracks to Hamiltonian to HAI

sequences of deforming shapes using both parametric models and nonparametric methods and we applied this algorithm for gait-based human recognition on the USF dataset by exploiting the shape deformations of a person's silhouette as a discriminating feature. 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 [12] shows and any of these preexisting approaches can also be used in the Form Pathway.

5 Experimental Results

For all experiments, tracking and basic object-detection was already available [13] and we utilized these (x,y,t) tracks to compute the HAI, which is then used as the Dorsal/Motion Pathway component of the framework. For the Ventral/Form Pathway, we utilized shape as defined in [11], where we calculated the shape of the silhouettes and computed similarity using DTW in the shape space. We then utilized Weighted Integration [5] to bias the Form Pathway component with the Motion 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.

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. We experimented with videos from 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.

Results are in Figures 4 and Figure 5. In Figure 4, we show similarity matrices for the USF Gait dataset examined using (a) Form Pathway, (b) Motion Pathway, and (c) the Integrated Framework on Probe A for all seven probes in the USF Gait. We see the Integration of Motion approach consistently outperforms the Form

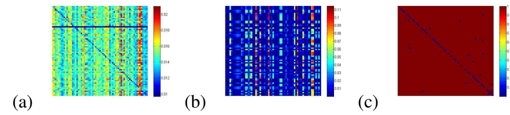

Figure 4. Similarity Matrices for USF Gait.

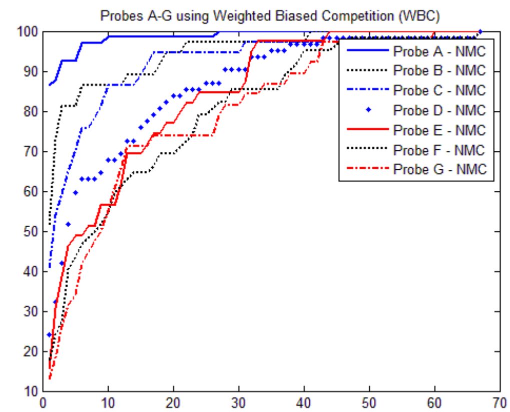

Figure 5. CMS Curves for USF Gait.

Pathway approach alone, as seen in Figure 5. As seen in Table 1, 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 4.

For example, in Probe A, the Ventral Pathway method ranked 81.8% of the people correctly in rank 1 whereas the NMC approach ranked 86.4% in rank 1; in Probe D, the Ventral Pathway ranked 54.8% in rank 5 while the NMC ranked 58.1% there; in Probe F, the Ventral Pathway got 54.8% in rank 10 whereas NMC ranked 60%. 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 employ a physics-based development to create a new spatio-temporal gait representation, called the Human Action Image. We then cast this HAI within the neurobiological model of motion recognition to create a framework for gait recognition which we apply to real world datasets. We see much room for future research,

Probe	Rank 1		Rank 5	
	Ventral	NMC	Ventral	NMC
A	81.8	86.4	92.4	92.4
B	59.5	51.4	81.1	83.8
C	40.5	40.5	70.3	70.3
D	21.0	24.2	54.8	58.1
E	15.4	15.4	46.2	46.2
F	16.1	17.7	41.9	43.6
G	13.2	13.2	34.2	34.2

Table 1. Ventral Pathway vs. Integration.

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.

Acknowledgement: The authors were partially supported by NSF grant IIS-0712253 and the DARPA VIRAT program.

References

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

[2] ——, "The recognition of human movement using temporal templates," *PAMI*, 2001.

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

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

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

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

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

[8] J. Han and B. Bhanu, "Individual recognition using gait energy image," *PAMI*, vol. 28, pp. 316–322, 2006.

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

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

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

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

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