A Perceptually-Inspired Stochastic Framework for Video Search and Analysis

Ricky J. Sethi
Fitchburg State University
160 Pearl St, Fitchburg MA 01420
rickys@sethi.org

ABSTRACT

The neurobiological model for motion recognition and visual processing posits that visual stimuli in the brain bifurcate into a Motion Energy Pathway and a Form Pathway, both of which are finally Integrated in the resulting motion recognition. In this article, we propose a perceptually-inspired computational framework for video search and analysis of human activity recognition. For high-resolution video, our computational framework relies upon the physics-based Human Action Image for the Motion Energy Pathway, shapes of trajectories and Dynamic Time Warping for the Form Pathway, and a variant of the bootstrap for the Integration. For low-resolution video, the computational equivalent uses simple, physics-based kinematic features for the Motion Energy Pathway, Space-time Interest Points and Nearest Neighbour classification for the Form Pathway, and a variant of the Markov chain Monte Carlo method for the Integration. We demonstrate the efficacy of our system on real-life video sequences from the well-known USF, Weizmann, UCR Videoweb, and YouTube/Hollywood Human Actions datasets.

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Miscellaneous

General Terms

Theory

Keywords

neurobiologically-inspired; motion recognition; human activity recognition

1. INTRODUCTION

Video search and analysis of activity recognition is difficult because the dynamic nature of video makes it necessary to combine both the image and motion elements represented in videos in order to recognize activities. Understanding activities, however, 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 visual processing framework within which we understand and interpret human activities [1].

Neurobiological and Neuromorphic Computing models of visual processing: Recent work in neurobiology [2, 3, 4] 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 bifurcates into two streams at V1: a Dorsal Motion Energy Pathway and a Ventral Form/Shape Pathway [4, 5], although the exact mechanism of the Integration is an open question in neurobiology [2, 3, 6, 7, 8]. Some Neuromorphic Computing models [9, 10, 11] relate the ventral and dorsal processing to the ideas of gist and saliency. However, we use the standard notation for these two pathways (Dorsal and Ventral) as outlined in the original neurobiological models of motion recognition for visual processing [2, 3, 12]. In this article, we develop a perceptually-inspired computational model of motion recognition and apply it to the problem of video search and analysis of human activity recognition.

Eventually, or perhaps concurrently [13], the results of these two pathways are integrated; although existing neurobiological models for motion recognition do posit the existence of a coupling or integration of these two pathways, they leave any specific mechanism for combination of the two pathways as an open question [2, 3].

Activity Recognition Background: In terms of activity recognition [14, 15], some of the newer research uses the fusion of multiple features (e.g., [16]). Their approach to features fusion comes closest to the idea of combining features that express both the form and motion components. Our approach also draws inspiration from the method employed in [17], which detects global motion patterns by constructing super tracks using flow vectors for tracking high-density crowd flows in low-resolution. Our methodology, on the other hand, works in both high- and low-resolution and for densely- and sparsely-distributed objects.

Overview of Approach: We propose both the features that model the biological motion energy pathway as well as stochastic methods for the integration of the two pathways. We start with the requisite background and then address the computational framework for both high- and low-resolution videos. In the case of high-resolution videos, we use the physics-based Human Action Image for the Motion Energy Pathway [18, 19], shapes of trajectories and Dynamic Time

PIVP'14, November 7, 2014, Orlando, FL, USA.

http://dx.doi.org/10.1145/2662996.2663011.

Warping for the Form Pathway [20, 21], and a variant of the bootstrap for the Integration. For the low-resolution case, we use simple, physics-based kinematic features for the Motion Energy Pathway [22, 23], Space-time Interest Points and Nearest Neighbour classification for the Form Pathway [24], and a simple variant of the Markov chain Monte Carlo [25, 26], similar to the data-driven Markov chain Monte Carlo [27, 28], in order to sample both the form and motion energy spaces simultaneously for the Integration[25, 26]. Finally, we show the experimental results of our framework in pruning video search and analysis using real video from the USF, Weizmann, UCR Videoweb, and YouTube/Hollywood Human Actions datasets.

2. HIGH-RESOLUTION CASE

For high-resolution videos, where we can distinguish the contours of humans, we utilize shape features with Dynamic Time-Warping for the Form Pathway, the Human Action Image for the Motion Energy Pathway, and a variant of the bootstrap for the Integration.

2.1 The Form Pathway

Our construction provides flexibility since new approaches in low-level feature extraction can be employed easily within our framework. For the present work, we use shape features with Dynamic Time-Warping. In general, the neurobiology literature suggests that orientation, shape, and color might serve as the best form-based components [1]. In particular, we used the approach in [21]. The integration is directly on the similarity matrices, so the exact same method can be used for other low-level features.

2.2 Motion Energy Pathway

The Human Action Image (HAI) [18, 19] captures a unique, physics-based energy signature of the motion of humans; as such, we utilize it to characterize the Motion Energy Pathway in our neurobiologically-inspired framework for high-resolution videos. In particular, the HAI is constructed by getting 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 in [18, 19], where $T = \frac{1}{2}mv_o^2$ and $U = mg_a(y_b - y_a)$, derived directly from the trajectories, (x, y, t) and with mass m, gravitational acceleration g_a, and velocity v, with the mass and gravitation acceleration simply being scale factors to first approximation. HAI is then defined as $HAI = \frac{1}{N}\int_1^N L dt$, with $L = T - U$ and N number of frames in a complete cycle. Comparison of two Human Action Images, HAI_g of the video clip from the database gallery and HAI_p of the probe/query video clip, is done in the usual way utilizing the distance between them:

$$D(HAI_g, HAI_p) = \frac{\sum_{x,y} |HAI_g(x,y) - HAI_p(x,y)|}{\sqrt{\sum_{x,y} HAI_g(x,y) \sum_{x,y} HAI_p(x,y)}} \tag{1}$$

In our case, we have a test video sequence, the Probe HAI_p, which we compare against all the video clips in the video database Gallery, where we compute HAI for each gallery sequence clip,HAI_g.

2.3 Integration

The Integration is accomplished by first creating a distance matrix for each person in a given Probe with each person in the Gallery using the method for the Form Pathway which, in our case, was using shape features with Dynamic Time Warping. We also created a distance matrix for each person in that same Probe with each person in the Gallery using our method for the Motion Energy Pathway, the HAI. Finally, we apply the Integration model to the two matrices in order to get the result matrix, which contains our final matches.

Our Integration mechanism is based on hypothesis testing with a 2-sided Confidence Interval where we use the bootstrap to determine the variance. Following the work in [29, 30, 31], we use the bootstrap to find the variance of the desired quantile threshold of the 2-sided Confidence Interval used within the Integration. Bootstrap is a nonparametric method which lets us compute some statistics when distributional assumptions and asymptotic results are not available. In statistics, it is more appealing to compute the two sided α significance threshold (confidence interval) via bootstrapping because of its accuracy and lack of assumptions.

The bootstrap works by re-sampling with replacement to find the variance of a statistic on a sample. We use this algorithm twice: once for the upper quantile and once for the lower quantile. 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. 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.

So, in our case, we use the hypothesis testing framework to establish the critical region quantiles for the Confidence Interval associated with our significance level, α, for each probe in the Distance Matrix. In order to find the variance of the desired quantiles (both lower and upper), we use the bootstrap method from Figure 1. We use the same significance level, α, as is standard in Hypothesis Testing ($\alpha = 0.05$) and derive the bootstrap critical region, $CI = (q^u_{lower}, q^u_{upper})$, for the upper quantile and $CI = (q^l_{lower}, q^l_{upper})$ for the lower quantile. We also use the alternate method (using just the mean of the quantile threshold) from the bootstrap for comparison.

Our Integration method then uses the Motion Energy Pathway values to weight the Form Pathway values; if the observed distance value of the Form and the Motion Energy Pathway 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 Pathway value.

3. LOW-RESOLUTION CASE

For low-resolution videos where we cannot distinguish the contours of humans, we utilize Space-time Interest Points and Nearest Neighbour classification for the Form Pathway, simple physics-based kinematic features for the Motion Energy Pathway, and a variant of the Markov chain Monte Carlo method for the Integration.

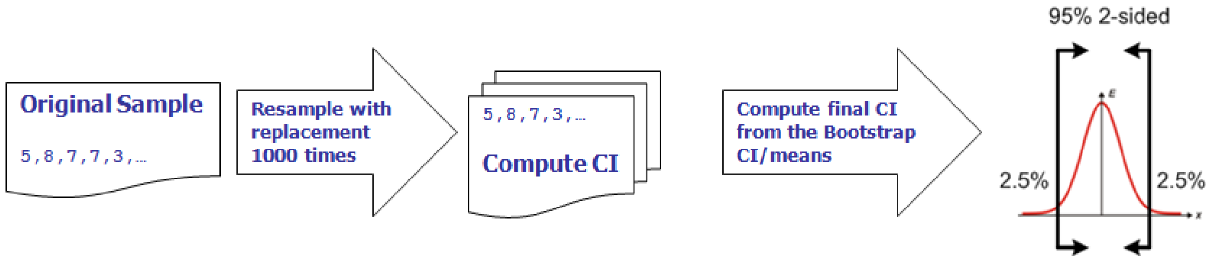


Figure 1: Overview of Bootstrap. This figure shows how the original sample is re-sampled (with replacement), say, 1000 times. In each re-sampling, a Confidence Interval is computed based on that sample. Eventually, the final Confidence Interval is estimated from either the Bootstrap Confidence Interval (on the CI computed on each re-sample) or the means (again, of the CI computed on each re-sample).

Algorithm 1 Markov chain Monte Carlo based Integration.

Integration based on Markov chain Monte Carlo

Initialize chain with τ_o
for $i = 1$ to $nsamples$ **do**
  // 1. Initial Metropolis-Hastings
  flag = true
  while (flag) **do**
    draw $\tau_i^{'} \sim e^{-K(D_{Motion}(\tau_i, \tau_q))}$
    draw $\alpha \sim \mathcal{U}[0,1]$
    if $\alpha > min\left(1, \frac{D_{Motion}\left(\tau_i^{'}, \tau_q\right)}{D_{Motion}\left(\tau_{i-1}, \tau_q\right)}\right)$ **then**
      flag = false
    end if
  end while
  // Prepare for Leapfrog
  $q_o = D_{Form}(\tau_i^{'}, \tau_q)$
  $p_o = D_{Motion}(\tau_i^{'}, \tau_q)$
  $q^o = D_{Form}(\tau_i^{'}, \tau_q)$
  draw $p^o \sim \mathcal{N}(0,1)$
  // 2. Perturbation using Leapfrog
  for j=1 to l **do**
    $p^{j-\frac{1}{2}} = p^{j-1} - \frac{\Delta t}{2} \bullet \nabla U(q^{j-1})$
    $q^j = q^{j-1} - \Delta t \bullet p^{j-\frac{1}{2}}$
    $p^j = p^{j-\frac{1}{2}} - \frac{\Delta t}{2} \bullet \nabla U(q^j)$
  end for
  $(q^{'}, p^{'}) = (q^l, p^l)$
  // 3. Final Metropolis-Hastings
  draw $\alpha \sim \mathcal{U}[0,1]$
  $\delta H = \left| H(q_o, p_o) - H^{'}(q^{'}, p^{'}) \right|$
  if $\alpha < min\left(1, e^{-\delta H}\right)$ **then**
    $(q_i, p_i) = (q^{'}, p^{'})$
  else
    $(q_i, p_i) = (q_{i-1}, p_{i-1})$
  end if
end for
return $\{\tau_i(q_i)\}_{i=0}^{nsamples}$

3.1 The Form Pathway

For the Form Pathway in the case of low-resolution videos, we use both mean shape [21], for the Weizmann dataset as above, and Space-time Interest Points and Nearest Neighbour classification for the Form Pathway [24], for the YouTube Hollywood Human Actions datasets, following the example of [32].

3.2 Motion Energy Pathway

Since we cannot track details of the objects, we instead use simple physics-based kinematic features calculated on the center of mass of the objects. In particular, we again use Kinetic (T) and Potential (U) Energy, where $T = \frac{1}{2}mv_o^2$ and $U = mg(y_b - y_a)$, derived directly from the trajectories, (x, y, t) of the center of mass of each human object, analogous to the HAI where the same kinematic quantities are computed for each point on the contour rather than on the center of mass.

3.3 Integration

For the Integration of the low-resolution case, we use the Data-Driven Hamiltonian Monte Carlo [25, 26], a simple variant of the well-known Markov chain Monte Carlo method. In particular, we utilize a standard Metropolis-Hastings component to select a proposal from the Motion Energy Pathway, we combine these values with corresponding Form Pathway value, do a standard Leapfrog perturbation , and, finally, use another Metropolis-Hastings component to accept or reject the final value.

Starting with tracks for an object, we calculate the Motion Energy and Form features for each object. The integration allows us to simultaneously search over both the motion energy space and image space in a concerted manner. We can then classify activities based on the likelihood of a particular track under a KDE or Gibbs estimator. We thus use form-based methods to compute similarities between the shapes obtained from a query track and all the database test tracks. We then convert the similarities to probability density functions by casting them as a KDE or Gibbs estimator.

This results in a potentially complex and difficult-to-sample joint distribution: $\pi(\tau, f) = \pi(\tau|f)\pi(f) = \pi(f|\tau)\pi(\tau)$, where $\pi(f)$ is the probability density function for the Form pathway and $\pi(\tau)$ is the probability density function for the Motion Energy pathway. Our goal is to sample this joint space, $\pi(\tau, f)$, and we employ our approach to do exactly this since Markov chain Monte Carlo methods have proved so successful in analyzing high-dimensional spaces in phase space. We expect the peaks to be highest in our joint distribution where both individual distributions exhibit higher values and we use the data-driven proposals to narrow in on those areas specifically. We are thus exploring the joint space by drawing proposals from one dimension/distribution and then searching in that vicinity in the other dimension or distribution. Because we expect the peaks in the joint distribution to correspond to areas where peaks of the motion and form distributions maximally overlap, we can use our approach to sample from just the $\pi(\tau)$ or the $\pi(f)$ instead of the $\pi(\tau, f)$, as well. In this article, we sample from the distribution of form similarities, $\pi(f)$, and the final algorithm is shown below in Algorithm 1.

4. EXPERIMENTS

We provide validation of our approach by showing the results of Integration via both the variant of the bootstrap and the variant of Markov chain Monte Carlo method. The experiments are meant to demonstrate both the efficacy and the hierarchical search our model affords as it prunes the search results.

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 and we are forced to combine datasets like the YouTube Action and Hollywood Human Actions datasets; we can also increase the "burn-in" period for larger datasets, which should yield better results. Thus, the improvement will be significantly better on a larger database as opposed to a small database.

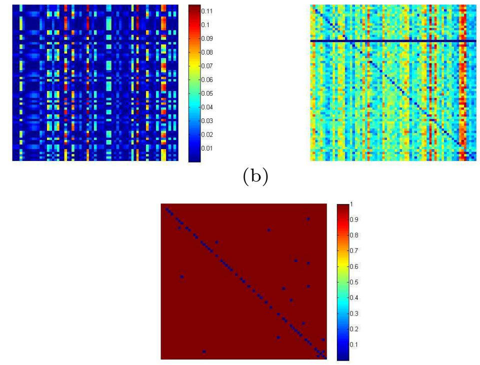
However, the amalgamated database shows the advantage of our approach, which is not fine-tuned for any specific database and can be widely and generically applied.

4.1 Motion Energy Experiments

We first show an example of the mechanical energy $(T+U)$ computed from video for simple objects. We apply this 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. For this experiment, we see the $(T+U)$ vs. Time curves for the two cars which follow all the way are highly correlated while the curve for the third car is not.

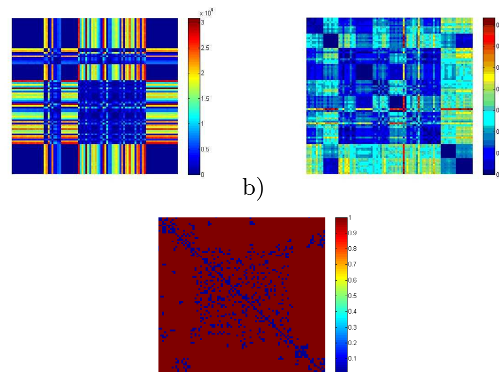
4.2 Integration via bootstrap

For the high-resolution case, we use the USF Gait dataset to demonstrate the efficacy of our method. For all experiments, tracking and basic object-detection was already available [33] and we utilized these (x, y, t) tracks to compute the HAI, which is then used as the Motion Energy Pathway component of the framework. For the Form Pathway, we utilized shape as defined in [21], where we calculated the shape of the silhouettes and computed similarity using Dynamic Time Warping in the shape space. We then utilized Integration with bootstrap 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 Motion Energy 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.



(a) (b) (c)

Figure 2: Similarity Matrices for USF Gait, where each probe is compared to each clip in the gallery for a) Motion only, b) Form (Shape) only, c) Integration via the Markov chain Monte Carlo variant (single diagonal line is best).



a) b) c)

Figure 3: Similarity matrices using the Weizmann dataset for a) Motion only, b) Form (Shape) only, c) Integration via the Markov chain Monte Carlo variant (single square blocks on diagonal indicate best matching of the block of similar/same activities).

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 2 and Figure 4. In Figure 2, 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 Pathway approach alone, as

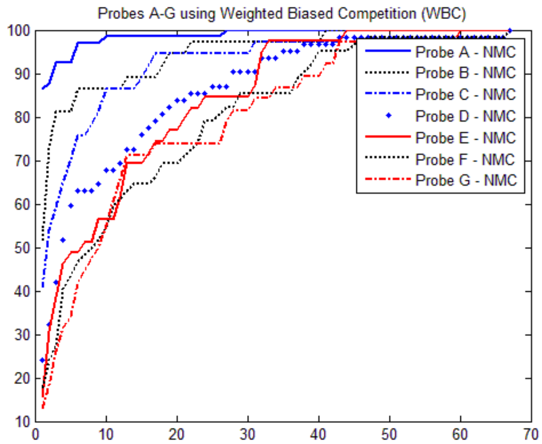


Figure 4: CMS Curves for USF Gait.

Table 1: Ventral Pathway vs. Integration.

Probe	Rank 1		Rank 5	
	Ventral	Integ	Ventral	Integ
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

seen in Figure 4. As seen in Table 1, the singular exception is Probe B in rank 1; this is because Integration with the bootstrap 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 2.

For example, in Probe A, the Ventral Pathway method ranked 81.8% of the people correctly in rank 1 whereas the Integration approach ranked 86.4% in rank 1; in Probe D, the Ventral Pathway ranked 54.8% in rank 5 while the Integration ranked 58.1% there; in Probe F, the Ventral Pathway got 54.8% in rank 10 whereas Integration 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.

4.3 Integration via Markov chain Monte Carlo

For the low-resolution case, we combined the YouTube Action (`http://cs.ucf.edu/~liujg/YouTube_Action_dataset.html`) and Hollywood Human Actions (`http://www.irisa.fr/vista/actions/`) datasets. The YouTube Action dataset contains 11 action categories and is very challenging due to large variations in camera motion, object appearance and pose, object scale, viewpoint, etc. The Hollywood Human Actions dataset contains video clips of short sequences from 32 movies and each sample is labeled according to one or more of 8 action classes. We combine these into a single dataset and use a query clip to find similar clips in the combined dataset that belong to the same class. We see the algorithm is able to identify the main activity without confusing it with any other activity and that the image method narrows in on the general class while the motion method achieves the final categorization of the hierarchical search. As seen in Figure 5, we see the multi-tiered search outperforms the image method alone; e.g., [32] showed an average precision of 16.4% (non-annotated) but our method of using Integration after the same image method returned 27.3% average precision. Since activity search in video is becoming a very important problem, we expect this algorithm to be an important contribution in this direction.

Figure 5: Grouping of GetOutCar Activity. This shows how our approach improves upon the performance of the Image method alone on the combined YouTube Action and Hollywood Human Actions dataset. Similar results are seen for the other examples

Further experiments on the well-known Weizmann dataset demonstrate how the Integration helps reduce the search space, as well as the hierarchical scheme for recognition. For all of these experiments, tracking and basic object-detection was already available and we utilized these (x, y, t) tracks to compute the kinetic and potential energy of the Motion Energy Pathway.

The Weizmann (`http://www.wisdom.weizmann.ac.il/~vision/SpaceTimeActions.html`) dataset 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 shape methods for the Form Pathway, using mean shape [21], as well as via the mechanical energy for the Motion Energy Pathway. In Figure 3, the rows and columns represent 10 activities by people and are organized according to activity. The plots show the clarification of matches using the different methods: in (a), Motion tends to confuse most activities and does a gross classification; in (b), Image/Form tends to do a little finer granularity of classification; in (c) the Integration based on the Markov chain Monte Carlo variant shows the finest granularity and distinction of matches and classification.

5. CONCLUSION

We presented the computational equivalent for a neurobiologically inspired model for motion recognition in video search analysis. The framework we developed is applicable to both high- and low-resolution videos. In the high-resolution case, we used the Human Action Image for the Motion Energy Pathway, shapes of trajectories and Dynamic Time Warping for the Form Pathway, and a variant of the bootstrap for the Integration. For low-resolution videos, we used simple physics-based features (the Kinetic and Potential Energy of the center of mass of the object) for the Motion Energy Pathway, Space-time Interest Points and Nearest Neighbour classification for the Form Pathway, and a variant of the Markov chain Monte Carlo method for the Integration. We demonstrated the efficacy of this framework on real videos from the USF, Weizmann, UCR Videoweb, and YouTube/Hollywood Human Actions datasets.

ACKNOWLEDGMENTS

This work is based upon work partially supported by the National Science Foundation under Grant #1019343 to the Computing Research Association for the CIFellows Project.

6. REFERENCES

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

[2] M. Giese and T. Poggio, "Neural mechanisms for the recognition of biological movements and action," *Nature Reviews Neuroscience*, vol. 4, pp. 179–192, 2003.

[3] M. A. Giese, "Neural model for biological movement recognition," in *Optic Flow and Beyond*, pp. 443–470, Kluwer Academic Publihers, 2004.

[4] 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.

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

[6] R. J. Sethi, A. K. Roy-Chowdhury, and S. Ali, "Activity Recognition by Integrating the Physics of Motion with a Neuromorphic Model of Perception," in *IEEE Workshop on Motion and Video Computing (WMVC)/IEEE Workshop on Applications of Computer Vision (WACV)*, WMVC, 2009.

[7] R. J. Sethi, A. K. Roy-Chowdhury, and A. Veeraraghavan, "Gait Recognition Using Motion Physics in a Neuromorphic Computing Framework," in *Multibiometrics for Human Identification* (B. Bhanu and V. Govindaraju, eds.), Cambridge University Press, 2010.

[8] R. J. Sethi and A. K. Roy-Chowdhury, "A Neurobiologically Motivated Stochastic Method for Analysis of Human Activities in Video," *International Conference on Pattern Recognition (ICPR)*, 2010.

[9] C. Siagian and L. Itti, "Biologically-inspired robotics vision monte-carlo localization in the outdoor environment," IROS, 2007.

[10] R. Peters and L. Itti, "Beyond bottom-up: Incorporating task-dependent influences into a computational model of spatial attention," CVPR, 2007.

[11] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," pp. 1254 – 1259, PAMI, 1998.

[12] M. Giese, "Neural model for the recognition of biological motion," in *Dynamische Perzeption 2*, 2000.

[13] L. Yang and M. Jabri, "Sparse visual models for biologically inspired sensorimotor control," pp. 131–138, Proceedings Third International Workshop on Epigenetic Robotics, 2003.

[14] W. Hu, T. Tan, L. Wang, and S. Maybank, "A survey on visual surveillance of object motion and behaviors," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 34, pp. 334–352, 2004.

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

[16] J. Liu, S. Ali, and M. Shah, "Recognizing human actions using multiple features," CVPR, 2008.

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

[18] R. J. Sethi and A. K. Roy-Chowdhury, "The Human Action Image," *International Conference on Pattern Recognition (ICPR)*, 2010.

[19] R. J. Sethi, A. T. Kamal, and A. K. Roy-Chowdhury, "Activity Recognition Results on UCR Videoweb Dataset," tech. rep., UCR, 2010.

[20] A. Veeraraghavan, A. Roy-Chowdhury, and R. Chellappa, "Role of shape and kinematics in human movement analysis," *CVPR*, 2004.

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

[22] R. J. Sethi and A. K. Roy-Chowdhury, "Modeling and Recognition of Complex Multi-Person Interactions in Video," in *MPVA at ACM International Conference on Multimedia (ACM MM)*, pp. 0–3, 2010.

[23] H. Jo, K. Chug, and R. J. Sethi, "A Review of Physics-based Methods for Group and Crowd Analysis in Computer Vision," *Journal of Postdoctoral Research*, vol. 1, no. 1, pp. 4–7, 2013.

[24] C. Schuldt, I. Laptev, and B. Caputo, "Recognizing human actions: a local svm approach," *ICPR*, 2004.

[25] R. J. Sethi, H. Jo, and A. K. Roy-Chowdhury, "A Generalized Data-Driven Hamiltonian Monte Carlo for Hierarchical Activity Search," in *IEEE International Conference on Image Processing (ICIP)*, 2013.

[26] N. M. Nayak, R. J. Sethi, B. Song, and A. K. Roy-Chowdhury, "Modeling and Recognition of Complex Human Activities," in *Visual Analysis of Humans* (T. B. Moeslund, L. Sigal, V. Kruger, and A. Hilton, eds.), pp. 289–309, Springer-Verlag, 2011.

[27] Z. Tu, S. Zhu, and H. Shum, "Image segmentation by data driven markov chain monte carlo," ICCV, 2001.

[28] S. Zhu, R. Zhang, and Z. Tu, "Integrating top-down/bottom-up for object recognition by ddmcmc," *CVPR*, 2000.

[29] B. Efron and R. J. Tibshiriani, *An Introduction to the Bootstrap, Monographs on Statistics and Applied Probability 57*. Chapman & Hall, 1993.

[30] E. Zio and F. D. Maio, "Bootstrap and order statistics for quantifying thermal-hydraulic code uncertainties in the estimation of safety margins," *Science and Technology of Nuclear Installations*, 2008.

[31] A. Zoubir and B. Boashash, "The bootstrap and its application in signal processing," *IEEE Signal Processing Magazine*, vol. 15, pp. 56–76, 1998.

[32] I. Laptev and P. Perez, "Retrieving actions in movies," *ICCV*, 2007.

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