2010 International Conference on Pattern Recognition

A Neurobiologically Motivated Stochastic Method for Analysis of Human Activities in Video

Ricky J. Sethi
UC Riverside
rickys@sethi.org

Amit K. Roy-Chowdhury
UC Riverside
amitrc@ee.ucr.edu

Abstract

In this paper, we develop a neurobiologically-motivated statistical method for video analysis that simultaneously searches the combined motion and form space in a concerted and efficient manner using well-known Markov chain Monte Carlo (MCMC) techniques. Specifically, we leverage upon an MCMC variant called the Hamiltonian Monte Carlo (HMC), which we extend to utilize data-based proposals rather than the blind proposals in a traditional HMC, thus creating the Data-Driven HMC (DDHMC). We demonstrate the efficacy of our system on real-life video sequences.

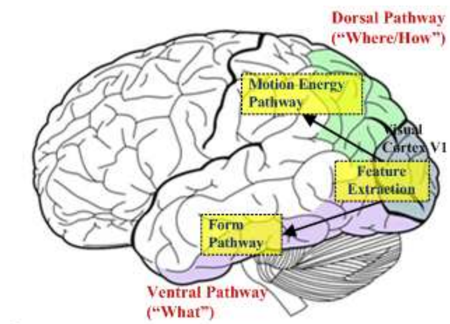

Figure 1. Feature extraction in V1.

1. Introduction

Recent work in Neurobiology [1, 2] 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 [2, 3]. The Form Pathway gives the visual context while the Motion Energy Pathway corresponds to visual saliency [4].

Although the exact mechanism of the Integration of these pathways is an open question in Neurobiology [1], we create a computational equivalent for the neurobiological model of Integration, with a focus on applications in complex video analysis. The need for integration is especially apparent in activity recognition which involves both form and motion information.

2. Related Work

Some researchers [5, 6] suggest the two pathways' integration is similar to object recognition; while biologically-inspired approaches like those used by [3, 7, 8] for image-based recognition have validated this approach, the success of non-biologically-motivated systems for extending object recognition descriptors to actions [9] also lend support to such a thrust. Building upon this, we developed [10, 11] integration strategies based upon hypothesis testing and the bootstrap.

However, searching the joint space of form and motion, or even the form or motion space individually, is difficult because the distributions describing these spaces are often very complex; thus, integrating these two disparate forms of analysis is a significant problem which we address in more detail in this work. Because statistical sampling techniques have proven so effective at analyzing such complex spaces, we turn to Markov chain Monte Carlo (MCMC) techniques to sample this joint space and develop a computational equivalent for the neural integration.

MCMC is a way to generate a random sequence of values for parameter x from a target probability distribution function (pdf), $\pi(x)$; standard MCMC techniques assume that you start off with some target distribution that you can evaluate but cannot sample from. Monte Carlo is a sampling method for iteratively evaluating a deterministic model using sets of random numbers as inputs where the samples are drawn from a probability distribution while a Markov chain is a stochastic

1051-4651/10 $26.00 © 2010 IEEE
DOI 10.1109/ICPR.2010.78

process that consists of a finite number of states with probabilities for transitions from each state to the next and having the property that future states depend only on the present state. Hamiltonian Monte Carlo (HMC), also referred to as the Hybrid Monte Carlo, is an alternative MCMC technique in which an auxiliary (fictitious) momentum variable is introduced for each parameter of the target pdf. HMC tries to avoid the random walk behavior of regular MCMC and allows proposals to move across the sample space in larger steps, thus allowing the proposals to be less correlated and converge to the target distribution more rapidly.

In general, the HMC is faster than classical stochastic sampling-based (Gibbs sampling, Metropolis-Hastings, etc.) MCMC optimization algorithms. By following the dynamical path in phase-space, we can propose candidate moves that are far away from the current state but that still have a substantial chance of being accepted. This gives us a way to efficiently explore large regions of phase-space by simulating Hamiltonian dynamics in fictitious time in the traditional HMC. The benefit of following the trajectory of the system in phase-space is that it eliminates the random walk aspect of the chain while also improving mixing and producing more accurate estimates and allowing us to explore quickly regions that are far away from the current state.

Besides the HMC, another recent innovation in the development of MCMC was the Data-Driven MCMC (DDMCMC), which uses data-driven proposals to make the Markov chain efficient. DDMCMC has mainly been applied to image segmentation and object recognition [12]; similarly, although HMC has been applied to particle filters and tracking [13, 14, 15], these techniques have never been applied to activity recognition to the best of our knowledge. The ability of HMC to incorporate the underlying system dynamics in the MCMC process makes it an ideal candidate for problems in activity search and recognition.

Relying upon a data-driven component to make more informed proposals than the blind proposals generated within a traditional HMC, we form the logical next step in HMC development by introducing the Data-Driven HMC (DDHMC)*, which use these data-driven proposals to make the HMC search more efficient*. In addition, we apply the HMC framework, via the DDHMC, to activity recognition for the first time. Almost all activity recognition methods use some variant of context (form; e.g., appearance) and saliency (motion) analysis but utilize different heuristics to conduct that analysis. Thus, the integration afforded by DDHMC provides a stochastic search framework that is especially suited for activity recognition.

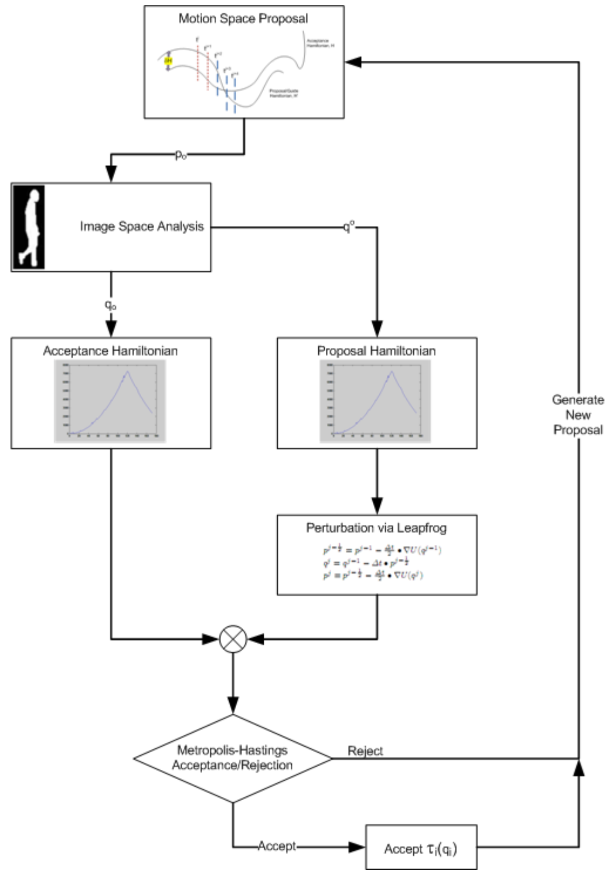

Figure 2. Motion-Proposal Generation.

3. Integration via DDHMC

Starting with tracks for an object, we calculate the motion and form features for each object and use them to compute similarities between the video 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 kernel density or Gibbs estimator.

However, 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 pdf for the Form pathway and $\pi(\tau)$ is the pdf for the Motion pathway. Our goal is to sample this joint space, $\pi(\tau, f)$, and we employ our DDHMC to do exactly this since the HMC has 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/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 the DDHMC to sample from just the $\pi(\tau)$ or the $\pi(f)$ instead of the $\pi(\tau, f)$, as well. In the motion-based DDHMC, we sample from the distribution of motion similarities, $\pi(\tau)$.

Our integration affords a hierarchical classification scheme in which the data-driven proposal does an initial, gross classification. Thus, we use the Hamiltonian analysis we developed in [10, 11] to generate the motion-based proposals whereas the mean shape method we developed in [16] is used for the form pathway and confirms the acceptance, as seen in Algorithm 1. In Algorithm 1, τ_o is the initial trajectory from the gallery/database, τ_q is the trajectory of the query, *nsamples* is the number of samples in the gallery/database, and $D_{Motion}(\tau_o, \tau_q)$ is the motion energy distance measure for the trajectory τ_o from τ_q and $D_{Form}(\tau_o, \tau_q)$ is its shape/form-based distance. $H(\tau_q)$ is the Acceptance Hamiltonian and $H^{'}(\tau_i)$ is the Proposal/Guide Hamiltonian and $H(\tau; t) = H(q(t), p(t))$ for the trajectory $\tau(q, p)$.

We form a pseudo-Hamiltonian, where the image-based and motion-based similarities are cast into a Hamiltonian function and we use the distance measures for the form and motion as the generalized coordinates and momenta, respectively, in lines 12-16, in order to create the Proposal Hamiltonian and the Acceptance Hamiltonian. In particular, we form the Acceptance Hamiltonian, $H(q_o, p_o)$, by using the motion-based distance measure to determine the p_o and the image-based distance measure to determine the q_o. We also form the Proposal Hamiltonian, $H^{'}(q^o, p^o)$, by using the image-based distance measure to determine the q^o and by sampling a normal distribution to determine the p^o.

This Proposal Hamiltonian is then subjected to a perturbation via Dynamic Transitions using Leapfrog in phase space because we assume the Shape/Form method is not perfect and the perturbation, just like the Dynamic Transitions in the Traditional HMC, accounts for such errors.

In Step 3, a normal HMC Metropolis-Hastings is used on the difference between the Acceptance and Proposal/Guide Hamiltonians. We finally accept the proposed trajectory if $\delta H \leq 0$ because it penetrates the Acceptance Hamiltonian's trajectory in phase space then (and so, we conclude the Guide Hamiltonian's trajectory is the same); but if $\delta H > 0$, we only accept with probability α. The penetration of one Hamiltonian trajectory by the other means they intersect which implies they are the same, as per [17].

An overview of the proposal generation is shown in Figure 2: here we see that the motion-based proposal suggests a fictitious momentum, p, and the image-based method is used within the HMC framework to get the fictitious position coordinate, q; finally, both the q and the p are used to create the Hamiltonian, $H(q, p)$, which is then analyzed via the HMC framework to make the final acceptance decision. Please note that our Hamiltonians are based on real data, unlike the traditional HMC. In Figure 2, where we see that the motion-based proposal suggests an artificial momentum, p, and the form-based method is used within the HMC framework to get the artificial position coordinate, q; finally, both the q and the p are used to create the Proposal Hamiltonian, $H^{'}(q^o, p^o)$ and the Acceptance Hamiltonian, $H(q_o, p_o)$, which are then analyzed via the HMC framework to make the final acceptance decision.

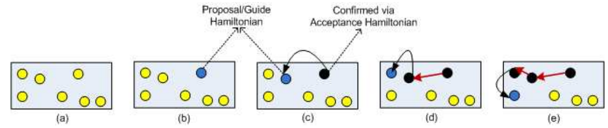

Figure 3. Proposals in Trajectory Space.

A diagrammatic representation of the overall evolution and eventual matching approach of the algorithm in trajectory space is shown in Figure 3. Figure 3(a) shows seven trajectories in trajectory space (represented as yellow circles). The DDHMC Algorithm starts off in Figure 3(b), where we enter Step 1 of the algorithm and find a proposal trajectory. This Proposal Hamiltonian is then compared with the Acceptance Hamiltonian and, if it is accepted, the algorithm continues with the loop by finding a new Proposal Hamiltonian in (c). In this way, the algorithm maneuvers through trajectory-space, only picking out those trajectories whose Proposal Hamiltonians are confirmed by the Acceptance Hamiltonian of the query clip in (c)-(e).

4. Experiments

Experiments on the well-known Weizmann dataset demonstrate how the Integration afforded by the DDHMC helps reduce the search space using the data-driven portion, as well as the hierarchical scheme for recognition, as shown in Figure 5. For all of these experiments, tracking and basic object-detection was already available and we utilized these (*x,y,t*) tracks to compute the Hamiltonian, as shown in [10, 11].

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 shape methods for the image component, using mean shape [16], as

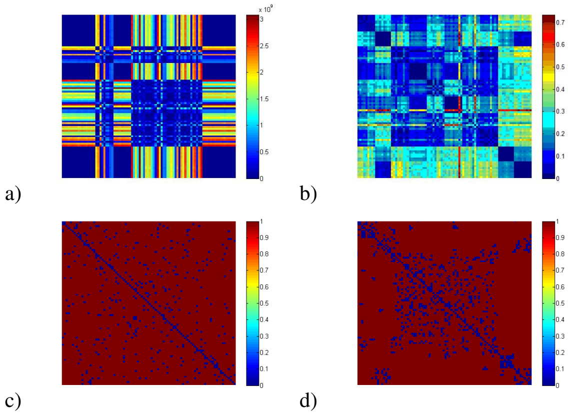

Figure 4. Weizmann Similarity Matrices.

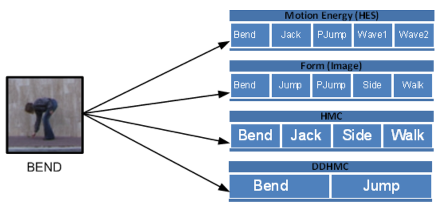

Figure 5. Grouping of Bend Activity.

well as via the Hamiltonian Energy Signature for motion [10, 11], by computing the tracks of the points on the contour and constructing the Hamiltonian.

We show the clustering of activities by the various algorithms where at least three people doing the same activity with an acceptance ratio of at least 0.75 are grouped together. Using this analysis, we see in Figure 5 how the Bend activity is confused with four others by just the Motion Energy examination and Image examination; the HMC narrows it down to just three others, but the DDHMC does the best.

Similar results are seen for the other activities. This dataset shows the potential to significantly reduce the search space in video database search problems. Since activity search in video is becoming a very important problem, we expect the DDHMC to be an important contribution in this direction.

In Figure 4, we see similarity matrices using using the Weizmann dataset for a) Motion only, b) Form only, c) HMC Integration, and d) Integration using DDHMC. 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 confuses most activities and does a gross classification; in (b), Image/Form tends to do a little finer granularity of classification; in (c) traditional HMC tends to have slightly better matching; but (d) DDHMC shows the finest granularity and distinction of matches and classification.

5. Conclusion

We present a DDHMC framework in which we first analyze motion-based information and then integrate in the form information via the HMC framework. The DDHMC thus uses motion energy-based data-driven proposals to make more informed proposals than the blind proposals generated within a Traditional HMC. These informed proposals are then used as the data-driven portion of the HMC to do an initial classification of the activities. Our proposed approach, using motion plus form information, thereby provides a natural framework for the integration of image and motion information, and brings the robustness of statistical methods to activity recognition.

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

References

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

[2] 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.

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

[4] T. Kadir and M. Brady, "Scale, saliency and image description," *IJCV*, vol. 45, pp. 83–105, 2001.

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

[6] N. H. Goddard, "The perception of articulated motion: Recognizing moving light displays," Ph.D. dissertation, University of Rochester, 1992.

[7] T. Serre, L. Wolf, and T. Poggio, "Object recognition with features inspired by visual cortex." CVPR, 2005.

[8] M. Ranzato, F. Huang, Y. Boureau, and Y. LeCun, "Unsupervised learning of invariant feature hierarchies with application to object recognition." CVPR, 2007.

Algorithm 1 Data Driven HMC Algorithm.

DDHMC (motion-based proposals)

Initialize chain with τ_o
for $i = 1$ to *nsamples* **do**
  // 1. Data-Driven: Get Proposal/Guide trajectory $\tau_i^{'}$ from data-based Gibbs distribution
  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 $\alpha > min\left(1, \frac{D_{Motion}\left(\tau_i^{'},\tau_q\right)}{D_{Motion}(\tau_{i-1},\tau_q)}\right)$ **then**
      flag = false
    end if
  end while
  // Initialize the Acceptance, $H(q_o, p_o)$, and the Proposal, $H^{'}(q^o, p^o)$ Hamiltonians
  $q_o = D_{Form}(\tau_i^{'}, \tau_q)$
  $p_o = D_{Motion}(\tau_i^{'}, \tau_q)$
  $q^o = D_{Form}(\tau_i^{'}, \tau_q)$
  draw $p^o \sim \mathcal{N}(0,1)$
  // 2. Perturbation on $H^{'}$ 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 $\{\tau_i(q_i)\}_{i=0}^{nsamples}$

[9] P. Dollar, V. Rabaud, G. Cottrell, and S. Belongie, “Behavior recognition via sparse spatio-temporal feature.” VS-PETS, 2005.

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

[11] 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.

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

[13] K. Choo and D. Fleet, “Tracking people using hybrid monte carlo,” *ICCV*, pp. 321–328, 2001.

[14] E. Poon and D. Fleet, “Hybrid monte carlo filtering: Edge-based people tracking,” in *WMVC*, 2002.

[15] C. Sminchisescu and B. Triggs, “Hyperdynamics importance sampling,” *ECCV*, pp. 769–783, 2002.

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

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