Online Sparse Gaussian Process (GP) Regression Model For Human Motion Tracking

Sri Lavanya Sajja¹, Dr.R.V.Krishnaiah²
¹Department of CSE, DRK Institute of Science & Technology, Ranga Reddy, Andhra Pradesh, India
²Department of CSE, DRK Institute of Science & Technology, Ranga Reddy, Andhra Pradesh, India

Abstract: Human motion tracking is an important aspect in digital image processing domain as it can be used in many real world applications including surveillance. However, it is very challenging to achieve this and dealing with multimodality. The existing solutions need huge training dataset in the learning phase and also they are not effective in dealing with multimodality. In this paper, we use a new technique known as online sparse Gaussian Process (GP) regression model. This model is a combination of temporal and spatial local GP experts model for efficient estimation of human pose. Thus it is a hybrid system that initializes itself and does robust human motion tracking. We developed a prototype application using this novel approach. Experiments are made on two real world datasets known as PEAR and HumanEva. The empirical results reveal that the proposed model is effective and has improved results when compared with existing solutions.

Index Terms – Human motion tracking, local experts model, temporal-spatial model, pose estimation, Gaussian process regression

I. Introduction

Human motion detection has many real world applications. By detecting human motion expert systems can take decisions. The real time applications [1] of human motion detection include human computer interaction, diagnostics, rehabilitation, monitoring of certain areas for human movements, and video surveillance. This kind of application is very useful both in civilian and military applications where human being’s motion can be tracked and appropriate decisions can be taken automatically. There are many approaches for human motion detection. However, an approach named “discriminative approach” [2] is widely used in real world applications. Many researches [3], [4] were made on the concept of discriminative framework. They proposed visual observations mapped to configurations of human pose. The methods used involve nearest neighbor retrieval [5], regression [6], manifold learning [7] and probabilistic mixture of predictions approach [3]. Discriminative approaches have a difficulty due to visual-to-pose ambiguity with respect to modeling multimodality. In order to handle multimodality some approaches came into existence. One of them is the category of model mixtures. The conditional BME is also effective in case of multimodality. This is achieved by BME with the help of input sensitive gate function.

With high-dimensional data, many parametric models are not robust in motion tracking. Even the BME model degrades when training data is small as it heavily depends on data distribution regions. Gaussian Process [8] is widely used discriminative approach applied for applications that involve human motion tracking. The GP model is very powerful approach as it defines probability distribution prior. Moreover, it works with infinite function spaces that makes it flexible and work along kernelized covariance function. On small scale databases [9], [10], GP has more advantages. Human motion tracking is done within the discriminative framework using GP regression. However, GP has certain limitations. They include computational cost is more; and inadequate to manage multimodality [11]. The GP with sparse approximation [8] is capable of overcoming the first limitation of GP without loosing its key characteristics. Mixture of GP experts [12], [13] is an effective alternative for minimizing the limitations of original GP.

In this paper we propose a model that combines local GP experts. This makes use of both temporal and spatial information required by human motion tracking. Only temporal information or spatial information is not sufficient for effective motion tracking. However, the existing discriminative methods do not use both. However, [14] and [15] tried to use both in different means. In this paper we use both of them to make it more robust in human motion tracking. Our model is in a discriminative and regression framework. Moreover, it is local model and GPDM is global. Our work is inspired by sparse GP regression [16]; it is non-parametric and integrated with both temporal and spatial information. The main contributions of this paper include definition of local GP experts containing samples localized in terms of inputs and outputs that supports multimodality;

www.iosrjen.org
integration of temporal and spatial experts to form a hybrid model; the experiments are made on the database i.e. PEAR (Pose Estimation and Action Recognition); the proposed system is evaluated using real time databases such as PEAR and Human Eva [17].

II. Sparse Strategy Of Gp Regression

The sparse GP regression in the local input output space leads to the proposed GP experts model. The GP model and algorithms are described in the following sub sections.

**GP Model**

Gaussian Process is the general form of Gaussian distribution model. This model over infinite index sets is described in [12]. Gaussian distribution with mean and covariance are given by

\[
\mu(x^*) = k^{*}_\zeta^k K^{-1}_{\zeta \zeta} Y_\zeta,
\]

\[
\sigma(x^*) = k^{*}_\zeta^k - k^{*}_\zeta^k K^{-1}_{\zeta \zeta} k^{*}_\zeta^k.
\]

From another viewpoint the mean prediction is a weighted voting from N training outputs.

\[
\mu(x^*) = \sum_{n=1}^{T} w_n y_n
\]

Motivated by the above facts local mixture of GP experts can be made similar to the model described in [13]. According to that the mean prediction for a given test input is:

\[
\mu(x^*) = \sum_{i=1}^{T} \prod_{j} \kappa^{*}_{i,j} K^{-1}_{\zeta_j \zeta_j} Y_{\zeta_j} \sum_{i=1}^{T} \sum_{j=1}^{S} \prod_{j} \omega_{i,j} y_{i,j}
\]

Number of local experts is represented by \( T \) while the \( S \) is the size of each expert. Fig. 1 shows the algorithm for training of the local experts.

---

**Algorithm 1 Local mixture of GP experts: learning and inference**

1: **OFFLINE**: Training of the Local Experts
2: \( R \): number of local GP experts \( (C_R, D_R) = \text{known}(D, R) \)
3: for \( i = 1 \ldots R \) do
4: \( \{ \bar{\theta} \} = \text{min} \left( -\ln p \left( \bar{Y}_{R}, |X_{R}, \bar{\theta} \right) \right) \)
5: end for
6: **ONLINE**: Inference of test point \( x_\eta \)
7: \( T \): number of experts, \( S \): size of each expert
8: \( \eta = \text{findNN}(X, x_\eta, T) \)
9: for \( j = 1 \ldots T \) do
10: \( \zeta = \text{findNN}(D_\eta, d_{\eta j}, S) \)
11: \( t = \text{findNN}(C_{R}, d_{\eta j}, 1) \)
12: \( \bar{\theta} = \bar{\theta} \)
13: \( \mu_j = k_{\zeta \zeta} K^{-1}_{\zeta \zeta} Y_{\zeta} \)
14: end for
15: \( p(y_{\zeta} | X, Y) \approx \sum_{\zeta=1}^{T} \alpha_{\zeta} \mathcal{N} (\mu_\zeta, \sigma_\zeta^2) \)

Fig. 1 – Algorithm for learning and inference

As see in fig. 1 a full GP with a covariance function is approximated by the local GP experts. For a given test point the local GP experts are centered at the neighbors. However, the capability of this to deal with multimodality depends on the training data distribution in the context of multimodality. Theoretically it is not sufficient to deal with multimodality with only using spatial information.

III. Integrated Local Gp Experts

With the knowledge of spatial experts, we introduce temporal experts as an extension in order to deal effectively with multimodality. It is nothing but combined GP Experts model. In this model the local experts learn the relationship between the output space and input space. In the same fashion, the temporal local experts find the importance of underlying context of the outer space.

The proposed process is described by:

\[
P(y_i | y_{i-1}, x_i) = \int p(y_i | y_{i-1}) p(y_{i-1}) dy_i.
\]

Fig. 2 shows algorithm that actually provides instructions for online inference with temporal and spatial local GP experts. Provided a dataset, a set of hyper parameters are learnt for local spatial GP experts. Then the local temporal model is built. Finally local experts model is substituted by appropriate values.
Online sparse Gaussian Process (GP) regression model for Human Motion Tracking

Algorithm 2 Online inference with temporal-spatial local GP experts

1: \( p(\mathbf{y}_t|\mathbf{Y}_1, \mathbf{Y}_2, \mathbf{y}_{t-1}) \approx \sum_{j=1}^{M} p_j(\mu_j, \sigma_j^2) \) (see Algorithm 1)
2: COMBINATION of two classes of local experts
3: \( T_1 \): number of spatial experts
4: \( T_2 \): number of temporal experts
5: \( \eta \) : size of each expert
6: \( \eta(\cdot) = \text{findNN}(\mathbf{x}_t, \mathbf{x}^*, T_1) \)
7: \( \eta^t(\cdot) = \text{findNN}(\mathbf{Y}, \mathbf{y}_t, T_2) \)
8: \( \eta = \eta^{s} \cup \eta^{t} \)
9: ONLINE Inference
10: \( T = T_1 + T_2 \): number of all experts
11: for \( j = 1 \ldots T \) do
12: \( \zeta = \text{findNN}(\mathbf{d}_i, \mathbf{d}_{ij}, S) \)
13: \( \tau = \text{findNN}(\mathbf{C}_k, \mathbf{d}_{ij}, 1) \)
14: \( k_{ij} = k_{\tau} K_{\tau}^{-1} \mathbf{y}_t \)
15: \( \mathbb{p}(\mathbf{y}_t^*|\mathbf{X}, \mathbf{Y}) \approx \sum_{j=1}^{M} p_j(\mu_j, \sigma_j^2) \)

Fig. 2 – Algorithm for online inference with temporal – spatial local GP experts

IV. Experiments

A prototype is built with the given algorithms and two real-time datasets are used for making experiments. The first data set is HumanEva-I used for evaluation of human pose estimation. It is collected at Brown University [17]. The second dataset known as PEAR is collected from Shanghai Jiao Tong University. The description of HumanEva database as used in the experiments is shown in Table 1.

<table>
<thead>
<tr>
<th>Set Partition</th>
<th>Action</th>
<th>S1</th>
<th>S2</th>
<th>S3</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training set</td>
<td>Walking</td>
<td>613</td>
<td>438</td>
<td>393</td>
<td>1444</td>
</tr>
<tr>
<td></td>
<td>Box</td>
<td>97</td>
<td>81</td>
<td>507</td>
<td>685</td>
</tr>
<tr>
<td></td>
<td>Jog</td>
<td>228</td>
<td>397</td>
<td>348</td>
<td>973</td>
</tr>
<tr>
<td>Test set</td>
<td>Walking</td>
<td>386</td>
<td>433</td>
<td>267</td>
<td>1086</td>
</tr>
<tr>
<td></td>
<td>Box</td>
<td>126</td>
<td>110</td>
<td>271</td>
<td>507</td>
</tr>
<tr>
<td></td>
<td>Jog</td>
<td>85</td>
<td>393</td>
<td>396</td>
<td>874</td>
</tr>
</tbody>
</table>

Table 1 – HumanEva Dataset with Frame numbers

Table 1 shows the details of the training set, test set actions and number of frames used in the experiments. Only consistent frames are used in experiments. Walking motion 2530, job motion 1847 and box motion 1192 frames are used. PEAR dataset details are shown in Table 2.

<table>
<thead>
<tr>
<th>Set Partition</th>
<th>Action</th>
<th>S1</th>
<th>S2</th>
<th>S3</th>
<th>S4</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training Set</td>
<td>Walk</td>
<td>250</td>
<td>218</td>
<td>350</td>
<td>367</td>
<td>1185</td>
</tr>
<tr>
<td></td>
<td>Jog</td>
<td>250</td>
<td>251</td>
<td>320</td>
<td>372</td>
<td>1193</td>
</tr>
<tr>
<td></td>
<td>Jump</td>
<td>150</td>
<td>97</td>
<td>240</td>
<td>249</td>
<td>736</td>
</tr>
<tr>
<td></td>
<td>Skip</td>
<td>150</td>
<td>124</td>
<td>238</td>
<td>242</td>
<td>754</td>
</tr>
<tr>
<td></td>
<td>Wave</td>
<td>150</td>
<td>119</td>
<td>240</td>
<td>246</td>
<td>755</td>
</tr>
<tr>
<td></td>
<td>Stretch</td>
<td>150</td>
<td>126</td>
<td>240</td>
<td>239</td>
<td>755</td>
</tr>
<tr>
<td>Validate Set</td>
<td>Walk</td>
<td>233</td>
<td>211</td>
<td>240</td>
<td>335</td>
<td>1019</td>
</tr>
<tr>
<td></td>
<td>Jog</td>
<td>250</td>
<td>208</td>
<td>320</td>
<td>369</td>
<td>1147</td>
</tr>
<tr>
<td></td>
<td>Jump</td>
<td>150</td>
<td>120</td>
<td>215</td>
<td>241</td>
<td>726</td>
</tr>
<tr>
<td></td>
<td>Skip</td>
<td>150</td>
<td>124</td>
<td>240</td>
<td>246</td>
<td>760</td>
</tr>
</tbody>
</table>
Online sparse Gaussian Process (GP) regression model for Human Motion Tracking

<table>
<thead>
<tr>
<th>Wave</th>
<th>150</th>
<th>85</th>
<th>240</th>
<th>250</th>
<th>725</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stretch</td>
<td>150</td>
<td>127</td>
<td>240</td>
<td>231</td>
<td>748</td>
</tr>
</tbody>
</table>

Table 2 – Configuration of PEAR database with multiple subsets and frame numbers

**Impacts of Number of Experts**

Fig. 3 – Impact of number of experts and the size of each local expert

As seen in fig. 3, number of local GP experts are taken in X axis while the Y axis represents average 3D error. The relative small values of both number of GP experts (T) and size of each expert (S) can provide satisfactory results.

![Graph showing impacts of number of experts](image)

**Curve Comparison of Joint Angles**

Fig. 5 – Left Shoulder of subject S2 in walking action

As can be seen in fig. 5, the left shoulder of subject S2 is taken in X axis while the Y axis represents joint angle in degrees. With respect to walking action of subject S2, the graph shows results of FULL-GP, TSL-GP and also ground truth.

![Graph showing joint angles](image)
Fig. 6 – Right hip of subject S3 in job action
As can be seen in fig. 6, the right hip of subject S3 is taken in X axis while the Y axis represents joint angle in degrees. With respect to jog action of subject S3, the graph shows results of FULL-GP, TSL-GP and also ground truth.

Fig. 7 – Sample 3-D pose estimation results
As can be seen in fig. 7, the first column shows box actions of subject S3. The second column represents the jog action of subject S1 and the third column shows the walking action of subject S2.

Fig. 8 – Sample images in the PEAR database
As seen in fig. 8, the PEAR database has images with various actions such as subjects involved in jogging, jumping, waving, skipping, stretching, and walking respectively from left to right.

Fig. 9 – Sample 3-D pose estimation results
As can be seen in fig. 9, the first column shows walk actions of subject S2. The second column represents the jog action of subject S1 and the third column shows the jump action of subject S3.

V. Conclusion

In this paper, we combine spatial and temporal local GP experts for good estimation of 3-D human pose. The proposed model is a combination of GP experts containing information with respect of temporal and spatial nature. Thus the proposed system is a hybrid system that can effectively track human poses and also deals with multimodality. We built an application to demonstrate this work. Real world databases such as PEAR and Human Eva are used for the experiments. The experimental results reveal that the proposed system is able to track human motion accurately and effectively. The future work is to enhance the proposed system so as to adapt to the videos containing disjoint human motions or temporal jumps in the motion.

References

[1] Xu Zhao, Member, IEEE, Yun Fu, Senior Member, IEEE, and Yuncai Liu, Member, IEEE. Human Motion Tracking by Temporal-Spatial Local Gaussian Process Experts. IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 20, NO. 4, APRIL 2011.


Authors

Sri Lavanya Sajja is a student of DRK Institute of science and Technology, Ranga Reddy, Andhra Pradesh, India. She has received M.SC degree in Computer Science and is pursuing her M.Tech in CSE. Her main research interest includes Image Processing and Networking.

Dr. R. V. Krishnaiah has received Ph.D from JNTU Ananthapur, M.Tech in CSE from JNTU Hyderabad, M.Tech in EIE from NIT (former REC) Warangal and B.Tech in ECE from Bapatla Engineering College. His main research interest includes Data Mining and Image Processing.