IOSR Journal of Engineering (IOSRJEN)
e-ISSN: 2250-3021, p-ISSN: 2278-8719, www.iosrjen.org
Volume 2, Issue 10 (October 2012), PP 17-22

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

Sri Lavanya Sajja[1], Dr.R.V.Krishnaiah[2]
[1]*Department of CSE, DRK Institute of Science & Technology, Ranga Reddy, Andhra Pradesh, India*
Principal
[2]*Department 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;

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

$\mu(x_*) = k_{*,\zeta}\, K^{-1}_{\zeta,\,\zeta}\, Y_\zeta$

$\sigma(x_*) = k_{*,*} - k_{*,\zeta}\, k^{-1}_{\zeta,\,\zeta}\, k_{\zeta,*}$

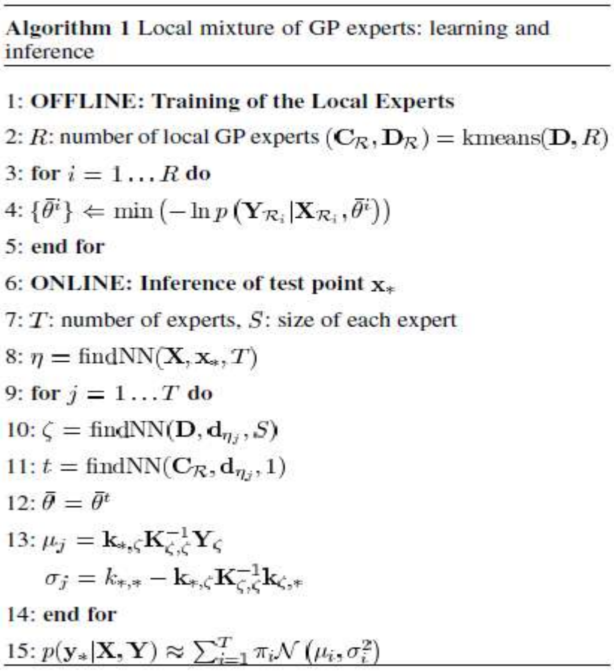
From another viewpoint the mean prediction is a weighted voting from N training outputs.

$\mu(x_*) = \Sigma^{T}_{n=1}\, w_n y_n$

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:

$\mu(x_*) = \Sigma^{T}_{i=1}\, \Pi_i\, k_{*,\zeta i}\, K^{-1}_{\zeta i,\, \zeta i}\, Y_{\zeta i} = \Sigma^{T}_{i=1}\, \Sigma^{S}_{j=1}\, \Pi_i\, \omega_{ij}\, y_{ij}$

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

```
OFFLINE: Training of the Local Experts
$R$: number of local GP experts $(\mathbf{C}_{\mathcal{R}}, \mathbf{D}_{\mathcal{R}}) = \mathrm{kmeans}(\mathbf{D}, R)$
for $i = 1 \ldots R$ do
    $\{\bar{\theta}^i\} \Leftarrow \min\left(-\ln p\left(\mathbf{Y}_{\mathcal{R}_i} | \mathbf{X}_{\mathcal{R}_i}, \bar{\theta}^i\right)\right)$
end for
ONLINE: Inference of test point $\mathbf{x}_*$
$T$: number of experts, $S$: size of each expert
$\eta = \mathrm{findNN}(\mathbf{X}, \mathbf{x}_*, T)$
for $j = 1 \ldots T$ do
    $\zeta = \mathrm{findNN}(\mathbf{D}, \mathbf{d}_{\eta_j}, S)$
    $t = \mathrm{findNN}(\mathbf{C}_{\mathcal{R}}, \mathbf{d}_{\eta_j}, 1)$
    $\bar{\theta} = \bar{\theta}^t$
    $\mu_j = \mathbf{k}_{*,\zeta}\mathbf{K}^{-1}_{\zeta,\zeta}\mathbf{Y}_\zeta$
    $\sigma_j = k_{*,*} - \mathbf{k}_{*,\zeta}\mathbf{K}^{-1}_{\zeta,\zeta}\mathbf{k}_{\zeta,*}$
end for
$p(\mathbf{y}_* | \mathbf{X}, \mathbf{Y}) \approx \sum_{i=1}^{T} \pi_i \mathcal{N}\left(\mu_i, \sigma_i^2\right)$
```

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

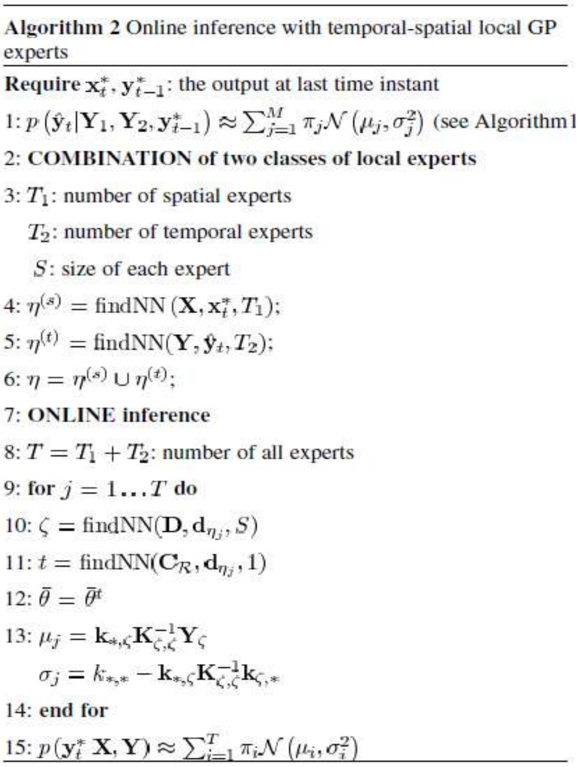
With the knowledge of spatial experts, we introduce temporal experts as an extension in order to deal effectively with multimodality. It is nothing but combined GP Experts model. In this model the local experts learn the relationship between the output space and input space. In the same fashion, the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

The proposed process is described by:

□ □ □

$P(y_t|y_{t-1},x_t) = \int p(y_t|y_t,x_t)p(y_t|y_{t-1})\, dy_t.$

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.



```
Algorithm 2 Online inference with temporal-spatial local GP experts
Require $\mathbf{x}_t^*, \mathbf{y}_{t-1}^*$: the output at last time instant
$p(\hat{\mathbf{y}}_t|\mathbf{Y}_1, \mathbf{Y}_2, \mathbf{y}_{t-1}^*) \approx \sum_{j=1}^{M} \pi_j \mathcal{N}(\mu_j, \sigma_j^2)$ (see Algorithm1
COMBINATION of two classes of local experts
$T_1$: number of spatial experts
    $T_2$: number of temporal experts
    $S$: size of each expert
$\eta^{(s)} = \text{findNN}(\mathbf{X}, \mathbf{x}_t^*, T_1)$;
$\eta^{(t)} = \text{findNN}(\mathbf{Y}, \hat{\mathbf{y}}_t, T_2)$;
$\eta = \eta^{(s)} \cup \eta^{(t)}$;
ONLINE inference
$T = T_1 + T_2$: number of all experts
for $j = 1 \ldots T$ do
$\zeta = \text{findNN}(\mathbf{D}, \mathbf{d}_{\eta_j}, S)$
$t = \text{findNN}(\mathbf{C}_{\mathcal{R}}, \mathbf{d}_{\eta_j}, 1)$
$\bar{\theta} = \bar{\theta}^t$
$\mu_j = \mathbf{k}_{*,\zeta} \mathbf{K}_{\zeta,\zeta}^{-1} \mathbf{Y}_\zeta$
    $\sigma_j = k_{*,*} - \mathbf{k}_{*,\zeta} \mathbf{K}_{\zeta,\zeta}^{-1} \mathbf{k}_{\zeta,*}$
end for
$p(\mathbf{y}_t^* \mathbf{X}, \mathbf{Y}) \approx \sum_{i=1}^{T} \pi_i \mathcal{N}(\mu_i, \sigma_i^2)$
```

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.

Set Partition	Action	S1	S2	S3	Total
Training set	Walking	613	438	393	1444
	Box	97	81	507	685
	Jog	228	397	348	973
Test set	Walking	386	433	267	1086
	Box	126	110	271	507
	Jog	85	393	396	874

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.

Set Partition	Action	S1	S2	S3	S4	Total
Training Set	Walk	250	218	350	367	1185
	Jog	250	251	320	372	1193
	Jump	150	97	240	249	736
	Skip	150	124	238	242	754
	Wave	150	119	240	246	755
	Stretch	150	126	240	239	755
Validate Set	Walk	233	211	240	335	1019
	Jog	250	208	320	369	1147
	Jump	150	120	215	241	726
	Skip	150	124	240	246	760

	Wave	150	85	240	250	725
	Stretch	150	127	240	231	748
Test Set	Walk	250	218	350	361	1179
	Jog	250	251	320	355	1176
	Jump	146	120	239	242	747
	Skip	150	133	240	241	764
	Wave	150	86	240	241	717
	Stretch	149	118	240	234	741

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.

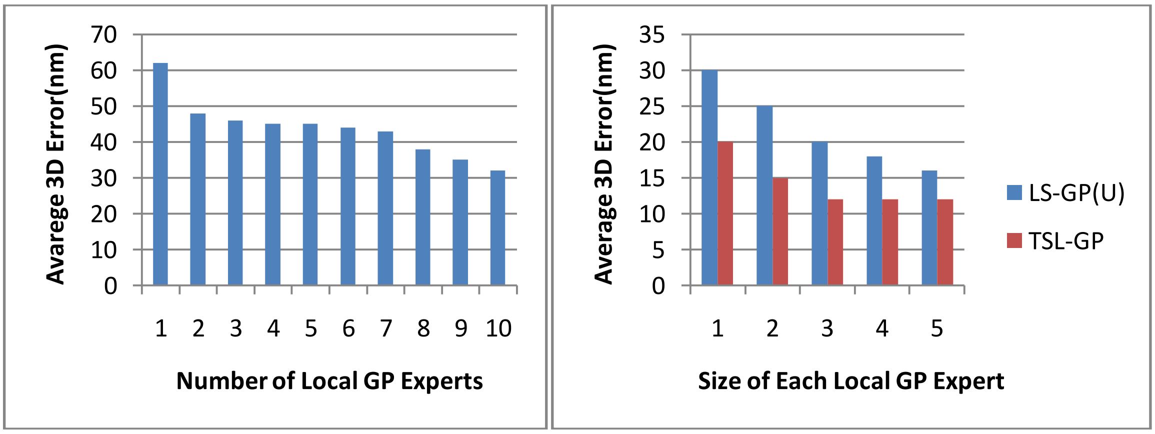


Fig. 4 – Impact of number of experts on the performance

As seen in fig. 4, the size of each local GP expert 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.

Curve Comparison of Joint Angles

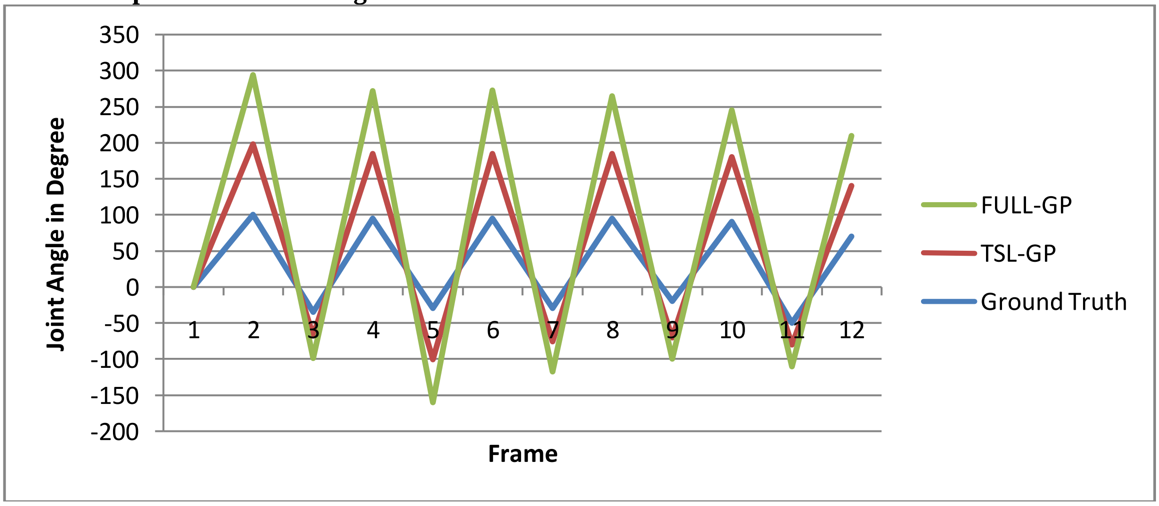


Fig. 5 – Left Shoulder of subject S2 in walking action

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.

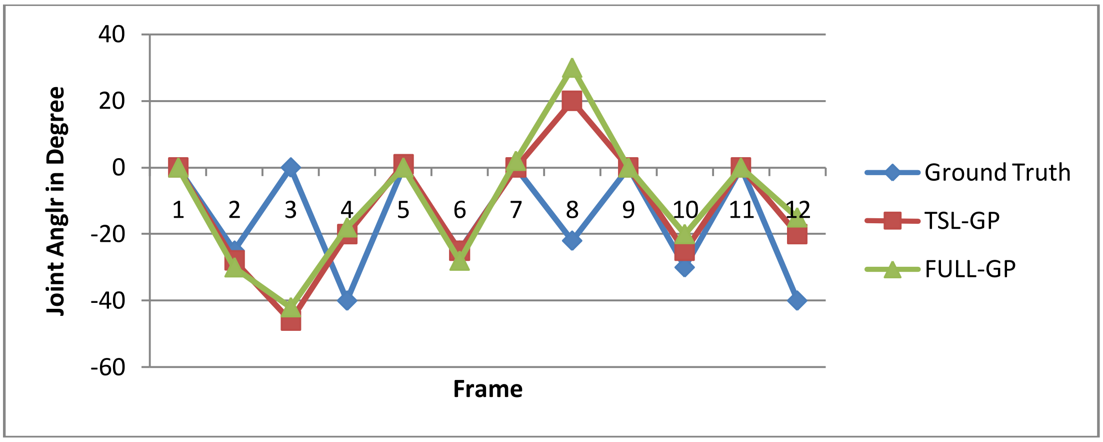


Fig. 6 – Right hip of subject S3 in job action

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.

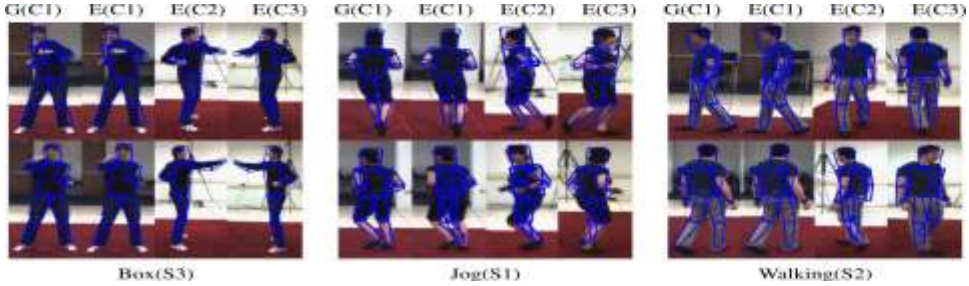


Fig. 7 – Sample 3-D pose estimation results

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.

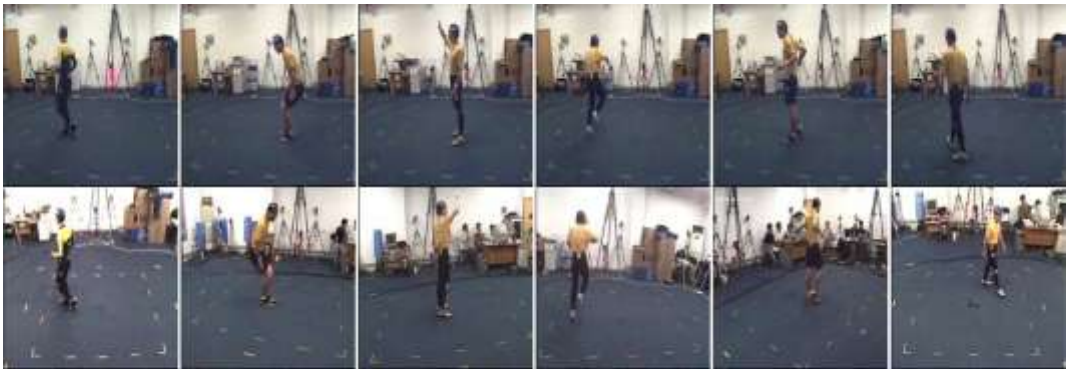


Fig. 8 – Sample images in the PEAR database

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.

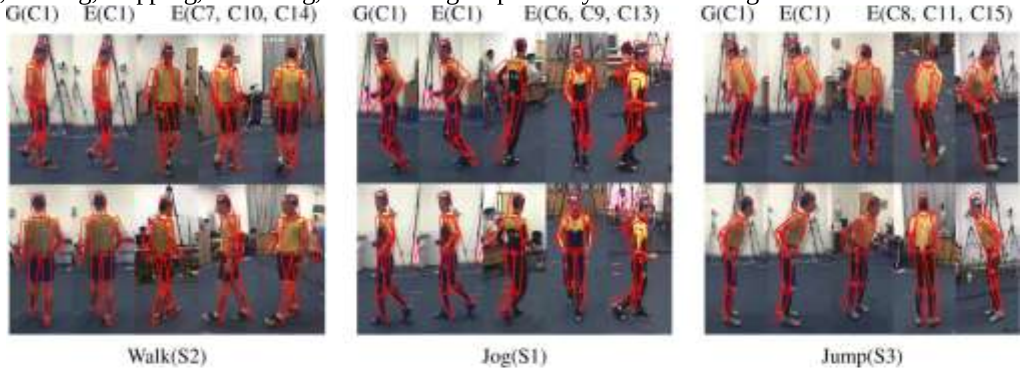


Fig. 9 – Sample 3-D pose estimation results

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.

[2] C. Sminchisescu, A. Kanaujia, Z. Li, and D. Metaxas, "Discriminative density propagation for 3-D human motion estimation," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2005, vol. 1, pp. 390–398.

[3] H. Ning, X. Wei, Y. Gong, and T. S. Huang, "Discriminative learning of visual words for 3-D human pose estimation," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2008, pp. 1–8.

[4] A. Bissacco, M.-H. Yang, and S. Soatto, "Fast human pose estimation using appearance and motion via multi-dimensional boosting regression," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2007, pp. 1–8.

[5] G. Shakhnarovich, P. Viola, and T. Darrell, "Fast pose estimation with parameter-sensitive hashing," in Proc. IEEE Int. Conf. Comput. Vis., 2003, vol. 2, pp. 750–757.

[6] X. Zhao, H. Ning, Y. Liu, and T. S. Huang, "Discriminative estimation of 3-D human pose using Gaussian processes," in Proc. 19th Int. Conf. Pattern Recognit., 2008, pp. 1–4.

[7] A. Elgammal and C. S. Lee, "Inferring 3-D body pose from silhouettes using activity manifold learning," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2004, vol. 2, pp. 681–688.

[8] C. E. Rasmussen and C. K.Williams, Gaussian Processes for Machine Learning. Cambridge, MA: MIT Press, 2006.

[9] X. Zhao, Y. Fu, H. Ning, Y. Liu, and T. S. Huang, "Human pose regression through multiview visual fusion," IEEE Trans. Circuits Syst. Video Technol., vol. 20, pp. 957–966, 2010.

[10] R. Urtasun, D. J. Fleet, A. Hertzmann, and P. Fua, "Priors for people tracking from small training sets," in Proc. IEEE Int. Conf. Comput. Vis., 2005, vol. 1, pp. 403–410.

[11] X. Zhao, Y. Fu, and Y. Liu, "Temporal-spatial local Gaussian process experts for human pose estimation," in Proc. 9th Asian Conf. Comput. Vis., 2009, vol. 5994, pp. 364–373.

[12] C. E. Rasmussen and Z. Ghahramani, "Infinite mixtures of Gaussian process experts," in Proc. Advances in Neural Information Processing Systems, 2002, pp. 881–888.

[13] V. Tresp, "Mixtures of Gaussian processes," in Proc. Advances in Neural Information Processing Systems, 2001, pp. 654–660.

[14] R. Urtasun, S. Epfl, D. J. Fleet, and P. Fua, "3D people tracking with Gaussian process dynamical models," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2006, vol. 1, pp. 238–245.

[15] J. M.Wang, D. J. Fleet, and A. Hertzmann, "Gaussian process dynamical models," in Proc. Advances in Neural Information Processing Systems, 2005, pp. 1441–1448.

[16] R. Urtasun and T. Darrell, "Local probabilistic regression for activityindependent human pose inference," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2008, vol. 2, pp. 1–8.

[17] L. Sigal and M. J. Black, "Humaneva: Synchronized Video and Motion Capture Dataset for Evaluation of Articulated Human Motion," Brown University, Providence, RI, 2006, Tech. Report CS-06-08.

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.