Human Motion Tracking by Temporal-Spatial Local Gaussian Process Experts

Abstract:

Human pose estimation via motion tracking systems can be considered as a regression problem within a discriminative framework. It is always a challenging task to model the mapping from observation space to state space because of the high-dimensional characteristic in the multimodal conditional distribution. In order to build the mapping, existing techniques usually involve a large set of training samples in the learning process which are limited in their capability to deal with multimodality. We propose, in this work, a novel online sparse Gaussian Process (GP) regression model to recover 3-D human motion in monocular videos. Particularly, we investigate the fact that for a given test input, its output is mainly determined by the training samples potentially residing in its local neighborhood and defined in the unified input-output space. This leads to a local mixture GP experts system composed of different local GP experts, each of which dominates a mapping behavior with the specific covariance function adapting to a local region. To handle the multimodality, we combine both temporal and spatial information therefore to obtain two categories of local experts. The temporal and spatial experts are integrated into a seamless hybrid system, which is automatically self-initialized and robust for visual tracking of nonlinear human motion. Learning and inference are extremely efficient as all the local experts are defined online within very small neighborhoods. Extensive experiments on two real-world databases, Human Eva and PEAR, demonstrate the effectiveness of our proposed model, which significantly improve the performance of existing models.

EXISTING SYSTEM:

- VISION BASED human motion tracking has been a fundamental open problem, with pervasive real-world applications [1], such as surveillance, rehabilitation, diagnostics, and human computer interaction.

- Among the large amount of studies in this field, the discriminative approach [2] has been prevalent due to its feasibility of fast inference in real-world scenarios and flexibility of adapting to different learning methods.
- Suffering from the intrinsic visual-to-pose ambiguity, however, all the discriminative approaches have the same difficulty of effectively modelling multimodal conditional distributions with small-size training data in a high-dimensional space.

PROPOSED SYSTEM:

- We propose a novel mixture of local GP experts' model in this work, which incorporates both temporal and spatial information.
- Theoretically, it is insufficient to effectively handle multimodality only by spatial information since the problem of monocular human motion estimation itself is ill-posed.
- Introducing temporal information into the model is reasonably necessary. But existing discriminative methods are short of temporal estimation framework.
- One exception is the parametric model proposed in which temporal smoothness constraints are added into the BME model. It is also worth noting that in, the Gaussian Process Dynamical Model (GPDM) is used to model the dynamics of human Motions.
- As the original GPDM is designed to find a low-dimensional latent space with associated dynamics, it is introduced to capture the motion priors in the latent state space.

Hardware Requirements

- SYSTEM : Pentium IV 2.4 GHz
- HARD DISK : 40 GB
- FLOPPY DRIVE : 1.44 MB
- MONITOR : 15 VGA colour
- MOUSE : Logitech.
- RAM : 256 MB
- KEYBOARD : 110 keys enhanced.

Software Requirements

- Operating system :- Windows XP Professional
- Front End   :- Microsoft Visual Studio .Net 2008
- Coding Language : - C# 2008.

Modules:

1. Video source selection
2. Analyzing Video
3. Extracting Frames
4. Track the Objects
5. Reconstruct the Frames with motion identifiers.

Modules Description:

Video source selection

In our project video signal input can be receive through the following 3 ways:

1. From Local Hard drive
2. Live video url from internet.
3. Capture Devices (Web camera, TV tuner card etc..,)

Analyzing Video

The Avi format videos are only supported in our project. Avi media Library in .net Framework 2.0 is used. There are many inner classifications are available in avi format. Before extracting frames support for Tracking is Fixed first.

Extracting Frames

Every video is converted as frames for object tracking. In live video internet urls there is no need to frame extraction. Because they are already available as Frames.

Track the Objects

Frames are like an image. Pixels are classified in an array. Horizontal and vertical Object matching is taken to track the variations in a pixels are identified. They are noted in a new array.

Reconstruct the Frames with motion identifiers.

Finally, based on a new array value Frames are constructed with Motion identifying red marks. From the frames new video is reconstructed.

REFERENCE:

Xu Zhao, Yun Fu and Yuncai Liu, "Human Motion Tracking by Temporal-Spatial Local Gaussian Process Experts", IEEE Transactions on Image Processing, Vol. 20, No.4, April 2011.