

Sistem Pengenal Aksi Spatio-Temporal pada Video Multi-Objek dengan Dynamic Positioning Aerial Camera = Spatio-Temporal Action Recognition System on Multi-Object Video with Dynamic Positioning Aerial Camera

Aldy Raja, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=9999920544110&lokasi=lokal>

Abstrak

Klasifikasi aksi multi-objek berdasarkan video RGB aerial merupakan tantangan kompleks yang dapat berguna untuk pengembangan sistem keamanan. Terdapat dua pendekatan jaringan saraf tiruan yang umum digunakan dalam sistem pengenal berbasis kerangka, Convolutional Neural Network (CNN) dan Graph Convolutional Network (GCN). Pendekatan CNN lebih efektif dalam mempelajari fitur spatio-temporal, lebih kuat terhadap noise dalam estimasi pose, dan dapat menangani skenario multi-objek dengan komputasi yang lebih ringan. Penelitian ini meliputi pengembangan pengenal aksi manusia dengan pendeteksi spatio-temporal berbasis kerangka menggunakan pendekatan 3D Convolutional Neural Network (3D-CNN). Pendeteksi spatio-temporal memungkinkan sistem untuk mengenali tiap-tiap aksi yang simultan dilakukan oleh multi-objek dalam satu rekaman video. Percobaan dilakukan menggunakan sejumlah pre-trained dataset dan menggunakan dataset video RGB aerial primer yang dilatih terhadap model pengenal aksi berbasis video frontal, dengan menerapkan metode transfer learning. Proses tranfer learning dilakukan dengan dataset khusus untuk menghasilkan model pelatihan yang memiliki akurasi tinggi. Pelatihan memberi keluaran berupa model jaringan saraf tiruan dengan nilai akurasinya. Pengujian dilakukan menggunakan data video untuk mengetahui ketepatan model. Dari model yang diperoleh, akan dilakukan analisis terhadap keberhasilan dan keakuratan metode dalam mengenali aksi manusia.

.....Multi-object action recognition based on aerial RGB video is a complex challenge that can be useful for security system development. There are two commonly used artificial neural network approaches in skeleton-based recognition systems, Convolutional Neural Network (CNN) and Graph Convolutional Network (GCN). CNN approach is more effective in learning spatio-temporal features, more robust to noise in pose estimation, and can handle multi-object scenarios with lighter computation. This research involves developing a human action recognition with skeleton-based spatio-temporal detection using a 3D Convolutional Neural Network (3D-CNN) approach. Spatio-temporal detection allows the system to recognize each simultaneous action performed by multiple objects in a single video footage. Experiments were conducted using a number of pre-trained datasets and using a primary aerial RGB video dataset trained on a frontal video-based action recognition model, by applying the transfer learning method. The transfer learning process is performed with a specific dataset to produce a high-accuracy training model. The training outputs an artificial neural network model with its accuracy value. Testing is done using video data to determine the accuracy of the model. From the model obtained, the success and accuracy of the method in recognizing human actions will be analyzed.