Spatio-Temporal Pattern Discovery for Pose-based Action Recognition

presented by

Junwu Weng
Ph.D. Student in IMI & EEE

Supervisor : Junsong Yuan

23rd February 2018
Outline

• Introduction

• Spatio-Temporal Naïve Bayes Nearest Neighbour
  – Motivation
  – Visual Representation
  – Naïve Bayes Nearest Neighbour (NBNN)
  – Spatio-Temporal NBNN
Introduction
Data Source

RGB
Depth
Pose

Texture: rich
Action Info.: redundant

Pose: barren

Semantic, concise
Problem Statement

• Pose-based Action Recognition
  – Input: pre-segmented 3D action sequence
  – Output: corresponding action class label

Label: Waving Right Hand
Application

1) Surveillance

2) Human-Computer Interaction
Spatio-Temporal Naïve Bayes Nearest Neighbor
Motivation

• Non-parametric model like NBNN has not been well explored in this field
  - NBNN has been successfully applied in image recognition
  - 3D action recognition is a light-weight problem

• Recognition of a certain action only related to movement of a subset of joints (spatial) and to a few certain frames (temporal)
Visual Representation

- Given a sequence of 3D poses

\[ p_j \in \mathbb{R}^{3J} \]

\[ v_j = p_{j+1} - p_j \]

Pose:

Velocity:

\[ \hat{x}_p = \frac{1}{z} [(p_1)^T \ldots (p_l)^T] \]

\[ \hat{x}_v = \frac{1}{z'} [(v_1)^T \ldots (v_1)^T] \]

Frame

Stage - Local Window \((l\text{ poses})\)

\[ x^i = [ (\hat{x}_p^i)^T, (\hat{x}_v^i)^T ] \]

Stage Descriptor \((i)\)

Video

Action Representation

\[ V = \{ x^1, \ldots, x^N \} \]
Naïve Bayes Nearest Neighbor

Class Set: \[ S_c = \bigcup_{k=1}^{N_c} V_k \]

Query Sample: \[ V_q = \{x^1, ..., x^N\} \] (testing sample)

Classification:

\[ \hat{c} = \arg \max_c p(V_q | c) = \arg \max_c \sum_{i=1}^{N} \log p(x^i | c) \]

\[ \hat{c} = \arg \min_c \sum_{i=1}^{N} \| x^i - NN_c(x^i) \|^2 \]
Naïve Bayes Nearest Neighbor

• From a set to a matrix (in temporal order)

\[ V = \{x^1, \ldots, x^N\} \quad \Rightarrow \quad X = [x^1, \ldots, x^N] \]

Nearest Neighbor Matrix:

\[ X^{NN}_c = [NN_c(x^1), \ldots, NN_c(x^N)] \]

NN Distance Matrix:

\[ X_c = (X - X^{NN}_c) \odot (X - X^{NN}_c) \]

\[ \hat{c} = \arg \min_c \sum_{i=1}^N \|x^i - NN_c(x^i)\|^2 = \arg \min_c \text{sum}(X_c) \]

\text{sum()} : Summation of Elements in } X_c
Spatio-Temporal NBNN

**NBNN :**

\[
\hat{c} = \arg \min_c \sum_{i=1}^{N} \| x^i - NN_c(x^i) \|^2 = \arg \min_c \text{sum}(X_c)
\]

\text{sum( )} : Summation of elements in \( X_c \)

**NBNN+SVM :**

\[
\hat{c} = \arg \min_c w^T x_c
\]

\( w^T \) Weights learnt by linear SVM

\( x_c \) Vectorized \( X_c \)

1) Too many parameters
2) Easy to over-fitting
Spatio-Temporal NBNN

**NBNN:**
\[ \hat{c} = \arg \min_c \sum_{i=1}^{N} \| x^i - NN_c(x^i) \|^2 = \arg \min_c \sum(X_c) \]
\[ \text{sum}(\cdot) : \text{Summation of elements in } X_c \]

**ST-NBNN:**
\[ \hat{c} = \arg \min_c (u^S_c)^T X_c u^t_c = \arg \min_c f_c(X_c) \]
- \( u^S_c \) Spatial Weights
- \( u^t_c \) Temporal Weights
Spatio-Temporal NBNN – Weight Learning

Objective Function:
(Support Tensor Machine)

\[
\min_{u^s_c, u^t_c} \frac{1}{2} \|u^s_c (u^t_c)^T\|^2 + \lambda \sum_{i=1}^{K} \xi_i
\]

\[
s.t \quad \sum_{1}^{N} u^t_c(i) = N, \quad u^t_c \geq 0
\]

\[
\xi_i \geq \max \left(0, 1 - c_i f_c(X^i_c)\right)^2
\]

\[
\xi_i \geq 0, \quad i = 1, ..., K
\]

Two Steps (iteratively):

Fix \( u_t \) and Update \( u_s \): \((u_t = [1, ..., 1])\)

\[
\min_{u^s_c} \frac{1}{2} \beta_1 \|u^s_c\|^2 + \lambda \sum_{i=1}^{K} \max \left(0, 1 - c_i f_c(X^i_c)\right)^2
\]

\[
\beta_1 = \|u^t_c\|
\]

Fix \( u_s \) and Update \( u_t \):

\[
\min_{u^t_c} \frac{1}{2} \beta_2 \|u^t_c\|^2 + \lambda \sum_{i=1}^{K} \max \left(0, 1 - c_i f_c(X^i_c)\right)^2
\]

\[
\beta_2 = \|u^s_c\|
\]

1) \( u_s, u_t \) are dependent on each other
2) Cannot be solved independently
Spatio-Temporal NBNN – Overview
Spatio-Temporal NBNN – Result

- Comparison with Baselines and State-of-the-arts

<table>
<thead>
<tr>
<th>Method</th>
<th>MSR</th>
<th>UTK</th>
<th>UCB</th>
</tr>
</thead>
<tbody>
<tr>
<td>NBNN-N</td>
<td>91.7</td>
<td>95.5</td>
<td>88.0</td>
</tr>
<tr>
<td>NBNN+SVM</td>
<td>92.4</td>
<td>94.0</td>
<td>100.0</td>
</tr>
<tr>
<td>Best Method</td>
<td>94.8</td>
<td>98.2</td>
<td>100.0</td>
</tr>
<tr>
<td>Ours</td>
<td>94.8</td>
<td>98.0</td>
<td>100.0</td>
</tr>
</tbody>
</table>
Spatio-Temporal NBNN – Result

• Visualization of Key Stages and Key Joints
Thanks