NON-LINEAR KERNEL SPACE INVARIENT REPRESENTATION FOR VIEW-INVARIENT
MOTION TraJECTORY RETRIEVAL AND CLASSIFICATION

Xu Chen, Dan Schonfeld and Ashfaq Khokhar

Department of Electrical and Computer Engineering, University of Illinois at Chicago
{xchen27, dans, ashfaq}@uic.edu

ABSTRACT
View-invariant representation has been shown to be a powerful tool in classification and retrieval of motion events due to camera motions. Traditional null space representation is invariant only for linear transformations and does not yield high accuracy for camera with non-linear motions. In this paper, we propose a novel general framework for non-linear kernel space invarient representation (NKSI), which is invariant to non-linear transformations due to camera motions with standard perspective transformation. We first derive NKSI and then propose an efficient classification and retrieval system relying on NKSI for archiving and searching motion events consisting of motion trajectories. The simulation results demonstrate superior performance of the proposed systems over traditional approaches.

Index Terms— Classification, non-linear, kernel space, retrieval, trajectory.

1. INTRODUCTION

The advent of diverse application areas including sign language gesture recognition, Global Positioning System (GPS), Car Navigation System (CNS), animal mobility experiments, sports video trajectory analysis and automatic video surveillance has recently raised many interesting research questions in object motion trajectory-based recognition. Object trajectories captured from different viewpoints lead to completely different representations. Therefore, a robust view invariance classification and retrieval algorithm is very desirable in the presence of linear and nonlinear transformations due to camera motions.

Among view-invariant systems, majority of them represent affine view-invariance in a linear space [1], thus limiting their applicability to data that may have only linear transformations, e.g. weak perspective transformation model. However, weak perspective model is only suitable and valid for data capturing environments that have small focal length, small field of view, and small depth variation. If the depth of an object has large variations, representing the feature points of the object with the average depth, assumed in the weak perspective model, will lose significant information and therefore cause inaccurate results for classification and retrieval. Moreover, if the x and y coordinates are not small compared to the depth, namely the field of view is not small enough, weak perspective model is also invalid for mapping 3D data onto 2D images. In this case, the weak perspective transformation does not hold and the standard perspective transformation is required. The standard perspective model will result in nonlinear mapping of the 3D environment onto 2D images.

In this paper, we propose a novel fundamental mathematical framework for non-linear invariant representation and use this framework for the important application of view-invariant classification and retrieval of motion events involving multiple motion trajectories. Our main contributions are as follows: (i) as far as we know, it is the first time we introduce the invariant representation with non-linear transformations, namely, non-linear kernel space invariant representation (NKSI), (ii) we demonstrate robustness and superiority of NKSI as a powerful tool in classification and retrieval of motion events over traditional approaches, especially with standard perspective transformation.

The rest of the paper is organized as follows: In Section 2, the mathematical formulation and the invariant nature of non-linear kernel space invariant representation (NKSI) are introduced. Section 3 presents simulation results using real-life trajectories from the Australian Sign Language database. Finally, in Section 4, we provide a brief summary of the results.

2. NON-LINEAR KERNEL SPACE INVARIANTS

2.1. Linear Null Space Invariants (LSN)

As in [1], let \( Q_i = (x_i, y_i) \) be a single 2-D point, \( i = 0, 1, \ldots, n - 1 \), among \( n \) ordered non-linear points in \( R^3 \). Consider the following arrangement of the \( n \) 2-D points in a \( 3 \times n \) matrix \( M \):

\[
M = \begin{pmatrix}
x_0 & x_1 & \cdots & x_{n-1} \\
y_0 & y_1 & \cdots & y_{n-1} \\
1 & 1 & \cdots & 1
\end{pmatrix}
\]

(1)

the null space \( H^{n-3} \) can be represented as:

\[
H^{n-3} = \{ q = (q_0, q_1, \ldots, q_{n-1})^T, \text{s.t.} Mq = (0, 0, 0)^T \}
\]

(2)
As it can be seen from equation (2) that, any linear transformation $T$ applied to $M$ can be written as $TM$, therefore, $TMq = 0$ is satisfied, which guarantees the invariance of null space when the data $M$ undergo linear transformations. However, the computation of null space invariants for non-linear operators is still an unresolved and challenging problem. We address this challenge by by first expanding the non-linear operators into a high-dimensional linear space with changing variables and subsequently computing the null space of the high-dimensional linear space. We find that the resulting space is invariant to the non-linear operators no matter how the coefficients of the non-linear operator change. The resulting space is referred to as non-linear kernel space invariants (NKSI). The main difference of NKSI with the null space for linear operators is that the dimensions of NKSI depend on the forms of non-linear operators. It resolves the important problem of motion event recognition when the data undergo non-linear transformations due to camera motions. There are various methods to expand the non-linear functions onto linear space.

2.2. Non-Linear Kernel Space Invariants (NKSI)

In this paper, we mainly discuss the use of Taylor expansion for determining NKSI. The approach can be easily extended to other expansion methods (e.g., Chebyshev polynomial and Lagrange polynomial). We first provide a theoretical analysis of the basic invariance property of the proposed non-linear invariant space representation. Let us assume we have a non-linear transformation $f(x, y) = p_1 \times f_1(x, y_1) + p_2 \times f_2(x, y_1) + \ldots + p_k \times f_k(x, y_1)$, where the coefficients $p_k$ are random variables and $f_i(x, y_1)$ are non-linear transformations of the raw data $x, y_1$, with deterministic forms. For example, in the case of $k = 2$, the original raw data matrix $M$ with $n$ samples is transformed into a new raw data matrix $\tilde{M}$ with the dimension $3 \times n$ as:

$$\tilde{M} = \begin{pmatrix} f_1(x_0, y_0) & \ldots & f_1(x_{n-1}, y_{n-1}) \\ f_2(x_0, y_0) & \ldots & f_2(x_{n-1}, y_{n-1}) \end{pmatrix}$$

Once we have obtained the new raw data $\tilde{M}$, the non-linear kernel space invariants (NKSI) can be computed as:

$$q_i^0 = \det \begin{pmatrix} f_1(x_1, y_1) & f_1(x_2, y_2) & f_1(x_3, y_3) \\ f_2(x_1, y_1) & f_2(x_2, y_2) & f_2(x_3, y_3) \end{pmatrix}$$

$$q_i^1 = -\det \begin{pmatrix} f_1(x_0, y_0) & f_1(x_2, y_2) & f_1(x_3, y_3) \\ f_2(x_0, y_0) & f_2(x_2, y_2) & f_2(x_3, y_3) \end{pmatrix}$$

$$q_i^2 = \det \begin{pmatrix} f_1(x_0, y_0) & f_1(x_1, y_1) & f_1(x_3, y_3) \\ f_2(x_0, y_0) & f_2(x_1, y_1) & f_2(x_3, y_3) \end{pmatrix}$$

Moreover, it is easy to extend the mathematical formulation of (3) for the case of 3D points. As it is known, the standard perspective object-to-image transformation $\pi_{A, \xi}$ has the form:

$$\begin{pmatrix} u' \\ v' \end{pmatrix} = \pi_{A, \xi} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \frac{\lambda}{gx + hy + kz + \xi_3 + \lambda} \begin{pmatrix} x \\ y \\ z \end{pmatrix} + \xi$$

where $\lambda$ is the focal length and the transformation:

$$A = \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & k \end{pmatrix}$$

Therefore, we have:

$$u_i' = \frac{\lambda(ax_i + by_i + cz_i + \xi_1)}{gx_i + hy_i + kz_i + \xi_3 + \lambda}$$

$$v_i' = \frac{\lambda(dx_i + ey_i + fz_i + \xi_2)}{gx_i + hy_i + kz_i + \xi_3 + \lambda},$$

where $i = 0, \ldots, N - 1$.

2.3. Basis Decomposition

From the equations (10) and (11), the standard perspective transformation is a non-linear transformation. Generally, we can formulate the expansions as:

$$u_i' \approx \sum_{p=0}^{N} p_{u,k} \Phi_{u,k}(x_i, y_i), v_i' \approx \sum_{p=0}^{N} p_{v,k} \Phi_{v,k}(x_i, y_i),$$

where $\Phi_{u,k}(x_i, y_i), \Phi_{v,k}(x_i, y_i)$ is a basis which is a function of $x_i$ and $y_i$, $p_{u,k}, p_{v,k}$ are the corresponding coefficients for the basis. It is well known that Karhunen-Loeve Transform (KLT) [2] gives the optimal expansions where $\Phi_{u,k}(x_i, y_i)$ here represents the eigenfunction of the transformation $f_u(x_i, y_i)$, $f_v(x_i, y_i)$. The computation of optimal $\Phi_{u,k}(x_i, y_i), \Phi_{v,k}(x_i, y_i)$ requires that the covariance matrix of the transformation function is known. If covariance matrix of the non-linear function is not available, we consider the widely used expansion, such as Taylor expansion at the original point:

$$u_i' \approx \sum_{w=0}^{N} q_{u,k} x_i^m y_i^n u_i, v_i' \approx \sum_{w=0}^{N} q_{v,k} x_i^m y_i^n,$$
where $m + n = w$. It has to be noted that we could always normalize the coordinates of all the trajectories to the range $[0, 1]$, therefore the high order terms $q_{u,k}x^m y^m, q_{v,k}x^m y^n$ where $m + n > N$, is small enough and can be ignored. In this case, the transformed raw data matrix becomes:

$$\tilde{M} = \begin{bmatrix} x_0 & \ldots & x_{n-1} \\ y_0 & \ldots & y_{n-1} \\ x_0^2 & \ldots & x_{2n-1} \\ y_0^2 & \ldots & y_{2n-1} \\ \vdots & \ldots & \vdots \\ 1 & \ldots & 1 \\ 1 & \ldots & 1 \end{bmatrix}$$

Since we have obtained the transformed raw data matrix $\tilde{M}$, we can compute the NKSI based on $\tilde{M}$. Besides Taylor expansion, we also test the performance of NKSI in classification and retrieval with Chebyshev polynomial and Lagrange polynomial. Specifically, the Chebyshev polynomial of the first kind can be represented as:

$$T_0(x) = 1, T_1(x) = x, T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x) \quad (14)$$

Once we have obtained the NKSI, there are various classification and retrieval algorithms that can be applied on NKSI. We choose a highly discriminative method, referred to as principal component null space analysis (PCNSA), described in [3].

3. View Invariant Motion Trajectory Retrieval and Classification

In order to implement and evaluate the proposed classification and retrieval system, we have used trajectories from the Australian Sign language (ASL) data set obtained from University of California at Irvine’s Knowledge Discovery in Databases (UCI-KDD) archive [4] and CAVIAR dataset [5]. The trajectories in the data set are obtained by registering the hand coordinates at each successive instant of time using a Power Glove interfaced to the system. In our simulations, we have used 40 different classes representing signing of 40 different words in the data set. Each class has 69 trajectories recorded at different instances. For each trajectory, we apply non-linear transformations under standard perspective models to make 5 transformed versions.

Since in real life trajectories may have different lengths, we normalize the length by taking the Fourier Transform and choosing the biggest $n=18$ coefficients and then taking the Inverse Fourier Transform so that all the trajectories are of size 32 before invariant matrix calculations. In the experiment, we vary the number of terms in Taylor expansions to evaluate the trade off for NKSI in the retrieval. It can be easily shown that in the equation (13), there are totally $\frac{N(N+1)}{2}$ terms. Fig.1 illustrates that the classification accuracy increases when the number of terms in Taylor expansions increase. We further demonstrate the classification accuracy of NKSI with increasing number of trajectories in one class in

Fig. 1. Accuracy for motion trajectory classification with an increasing number of Taylor expansion terms for NKSI.

Fig. 2. Accuracy for motion trajectory classification with increasing number of motion trajectories in one class for NKSI for linear and non-linear transformations.

3.1. Comparison

We compare the performance of our framework with traditional approaches based on null space invariant [1], curvature scale space (CSS) and centroid distance function (CDF) [6], normalized edit distance (NED) [7] in terms of classification and retrieval accuracy. Since the null space invariant based approach does not work with non-linear transformations, we show the comparison of other state of the art approach with NKSI under the standard perspective transforma-
We use the first 25 terms of Taylor expansion to compute the NKSI. As it can be seen from the results shown in Table 1, NKSI provides significantly higher classification accuracy due to its invariance nature for non-linear transformations. The approach of edit distance (NED) provides the worst performance since the distance measure will not be accurate with non-linear transformations. The comparison of the performance of classification accuracy with different expansion approaches is shown in Table 2.

### Table 1. Comparison of the classification accuracy for CSS, CDF and NED, with our approach NKSI in terms of different number of classes in CAVIAR dataset

<table>
<thead>
<tr>
<th>number of classes</th>
<th>CSS</th>
<th>CDF</th>
<th>NED</th>
<th>NKSI</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>0.675</td>
<td>0.632</td>
<td>0.563</td>
<td><strong>0.831</strong></td>
</tr>
<tr>
<td>30</td>
<td>0.632</td>
<td>0.613</td>
<td>0.512</td>
<td><strong>0.817</strong></td>
</tr>
<tr>
<td>35</td>
<td>0.623</td>
<td>0.592</td>
<td>0.482</td>
<td><strong>0.803</strong></td>
</tr>
<tr>
<td>40</td>
<td>0.586</td>
<td>0.574</td>
<td>0.453</td>
<td><strong>0.792</strong></td>
</tr>
<tr>
<td>45</td>
<td>0.553</td>
<td>0.521</td>
<td>0.419</td>
<td><strong>0.775</strong></td>
</tr>
</tbody>
</table>

### Table 2. Comparison of the classification accuracy for Taylor expansion, Chebyshev polynomial and Lagrange polynomial, with our approach NKSI with different number of terms in the expansion for CAVIAR dataset

<table>
<thead>
<tr>
<th>number of classes</th>
<th>Chebyshev</th>
<th>Lagrange</th>
<th>Taylor</th>
</tr>
</thead>
<tbody>
<tr>
<td>21</td>
<td>0.658</td>
<td>0.713</td>
<td>0.792</td>
</tr>
<tr>
<td>28</td>
<td>0.683</td>
<td>0.732</td>
<td>0.803</td>
</tr>
<tr>
<td>36</td>
<td>0.704</td>
<td>0.753</td>
<td>0.825</td>
</tr>
<tr>
<td>45</td>
<td>0.725</td>
<td>0.782</td>
<td>0.851</td>
</tr>
</tbody>
</table>

### 4. CONCLUSION

We proposed both the theoretical framework and practical application for non-linear invariant space as a powerful view-invariant representation for recognition and retrieval of motion event data sets. The proposed NKSI provides a promising solution for computing invariants for the original video data due to non-linear transformations caused by camera motion. We investigated the invariant property of NKSI and subsequently determined exact solutions for NKSI for arbitrary non-linear transformations with different basis. When the covariance matrix of the non-linear transformation is available, Karhunen-Loeve Transform with NKSI provides perfect solutions. When the covariance matrix of the non-linear transformation is not available, we evaluate the performance of NKSI with Taylor expansion, Chebyshev polynomial and Lagrange polynomial. Simulation results demonstrate the superiority of NKSI in the classification and retrieval of motion trajectories over traditional approaches.

### 5. REFERENCES


