Lie Group distance based generic 3-d vehicle Classification

Pradeep Yarlagadda
Brown University
Providence, RI 02906

Ozge Ozcanli
Brown University
Providence, RI 02906

Joseph Mundy
Brown University
Providence, RI 02906

***Abstract*—This paper introduces a 3-d representation of vehicles as a space of scale and orientation transformations that define the shape of individual vehicle instances. This shape space forms a group, where the similarity of different vehicle observations can be evaluated using a distance measure defined by Lie group theory. A generic class of vehicles (e.g. SUV) is represented by a set of curves on the Lie group manifold, called geodesics. The classification of any given vehicle instance is achieved by finding the class with the smallest Lie distance between the geodesics and the vehicle shape. Vehicle recognition is carried out on 3-d LIDAR point clouds. The performance of the Lie classifier is evaluated against two other approaches and found to provide superior recognition performance, particularly with respect to the ability to generalize from a small number of labeled prototypes.**

I. INTRODUCTION

Lie group theory is the fundamental representation of a space of transformations. The three central elements of the Lie group framework are: Lie distance; intrinsic mean; and principal geodesics. Lie distance is a measure of the similarity of two transformations. The intrinsic mean represents the 'average' of a set of transformations, i.e. the transformation that minimizes the Lie distance to all the transformations in the set. A geodesic is a 1-d subspace of transformations that is the shortest path between two transformations. A principal geodesic is one that accounts for the maximum variation in the set of transformations along the path, analogous to principal component of a covariance matrix.

Fletcher et al. [4] use Lie group theory to measure the statistics of shapes on a Lie group manifold. They encode the shape of organs such as the human kidney as a set of 3-d surface elements called a M-rep. The intrinsic variability in organ shape is represented by curves of minimal distance (geodesics) on the Lie group manifold. Klassen et al. [7] also use geodesic paths to span shape spaces and demonstrate their approach to generate a continuous family of shape prototypes within a category. Lie distance in transform space is exploited to cluster objects according to their shape. Tuzel et al. [5] exploit the linear behavior of Lie group elements in the tangent plane to the group manifold in order to implement a mean-shift algorithm for clustering transformations. The modes correspond to global transformation of objects due to 3-d orientation. [6] use the Lie group framework for building co-variance descriptors for detecting humans in video scenes. Extensive work on 3-d vehicle instance recognition in LIDAR has been carried out by Huber et al [1]. Ferryman et al. [9] use appearance model eigen-space representation for vehicle tracking in videos.

In this paper the Lie group framework is applied to the recognition of generic 3-d object categories, such as vehicles. The categories are defined by a set of transformations of vehicle shape and thus similarity between vehicle instances can be measured using Lie distance metric. The key contributions are: 1) A novel hierarchical strategy for generic object recognition based on the Lie group distance metric; 2) the first implementation of a complete object recognition system based on the Lie group framework; 3) an experimental validation on LIDAR vehicle database of the effectiveness of the Lie distance metric to measure the similarity of shapes and to generalize within a shape category.

II. SPOKE MODEL AND LIE GROUP

The vehicle database used in this paper was constructed from the LIDAR scan of a parking lot area consisting of vehicles. The parking lot and the sample 3-d vehicle point clouds extracted from the scan are shown in Figure 1 a. The sparsity of points in the vehicle clouds can be observed from the figure which makes it difficult to extract useful information, for example, learning the appearanace model mentioned in [9]. Instead, the vehicle point clouds are fitted with a shape representation called the spoke model in an automatic fitting procedure. The spoke models are subsequently used in a robust classification framework which is presented under the approach section. The spoke model and example vehicle point clouds fitted with the spoke model are shown in Figure 1 b. The spoke model is aligned with the principal axes co-ordinate frame of the point cloud. Each spoke has two controlling parameters namely, the 2-d rotation on the plane parallel to major axis of the vehicle and the radius of the spoke. Let r_i, n_i denote the radius and the normal vector of spoke S_i in the spoke model. Let k denote the number of spokes in the model.

Let $X = \begin{pmatrix} U_1 \\ . \\ . \\ U_k \end{pmatrix}$,$U_i = \begin{pmatrix} r_i \\ n_i \end{pmatrix}$. X denotes the configuration of all the spokes. Let $Y = T(X)$ denote the configuration obtained by a transformation of X. This can be written in the matrix form $Y = TX$ where

978-1-4244-2175-6/08/$25.00 ©2008 IEEE

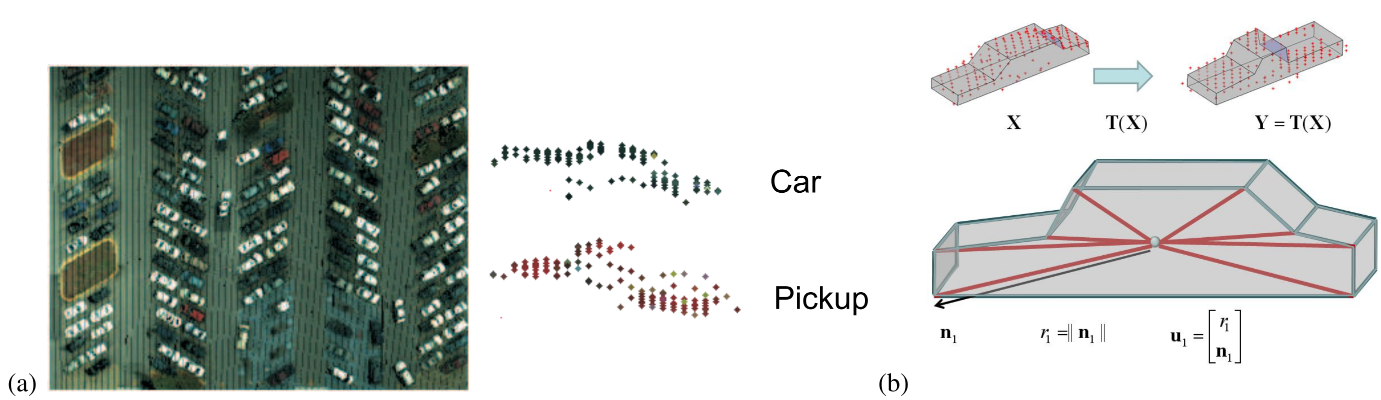

Figure 1 (a) shows the parking lot area consisting of vehicles and the extracted vehicle point clouds from the LIDAR scan. (b) shows the spoke model vehicle representation and examples of point clouds fitted with spoke model.

$$T = \begin{pmatrix} M_1 & . & 0 \\ . & . & . \\ 0 & . & M_k \end{pmatrix}, M_i = \begin{pmatrix} e^{\alpha_i} & 0 \\ 0 & R_i \end{pmatrix}$$

e^{α_i} denotes the scale acting on radius r_i. R_i denotes the 2-d rotation acting on normal vector n_i. X and Y represent the spoke model configurations of two vehicle instances. By varying T, different vehicle instances can be represented as transformations of X.

A. Exponential map and log map

A group is defined as a set of elements together with a binary operation (multiplication) satisfying the closure, associative, identity and the inverse axioms. A Lie group G is a group that is also a differentiable manifold. The tangent space of group G at the identity e, T_e, is called the Lie algebra g. The exponential map $\exp$ is a mapping from Lie algebra elements to Lie group elements. The logarithmic map $\log$ takes group elements onto the tangent plane. The Lie group distance between two points is defined as $d(x_1, x_2) = \| \log(x_1^{-1} x_2) \|$ where $\|.\|$ is the Frobenius norm of the resulting algebra element. A detailed treatment on Lie groups can be found in [8]

B. Intrinsic mean and principal geodesics

The spoke transformation matrix T is the Cartesian product of transformation matrices M_i acting on the individual spokes. Each M_i is a Cartesian product $,\Re$ x $SO(2)$ which forms a Lie group. [1] Since the Cartesian product of Lie group elements is a Lie group, T forms a Lie group. The Lie algebra element of T is obtained by performing component-wise log operation on each of the M_i.

$$\log(T) = \begin{pmatrix} \log(M_1) & . & 0 \\ . & . & . \\ 0 & . & \log(M_k) \end{pmatrix} \quad (1)$$

$$\log(M_i) = \alpha_i \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \theta_i \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \quad (2)$$

[1] Note that the Lie group $,\Re$ x $SO(2)$ is isomorphic to C^*, the multiplicative group of complex numbers. However, the approach presented in this paper works for any Lie group in general.

Equation (2) expresses the Lie algebra element of an individual spoke S_i in terms of the generator matrices for scaling and 2-d rotation.

The intrinsic mean μ of a set, S, of spoke transformation matrices $T_1, T_2, ..., T_n$ is defined as

$$\mu = \arg\min \sum_{k=1}^{n} \| \log(T^{-1} T_k) \| \quad (3)$$

The generators of the Lie algebra can be used to generate curves on the Lie group manifold. Consider the following 1-parameter Lie algebra element of spoke S_i.

$$A_{v_i}(t) = t\alpha_i \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + t\theta_i \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \quad (4)$$

α_i and θ_i are the scaling and the rotation generator coefficients of spoke S_i. v_i denotes that the Lie algebra element is defined at a fixed (α_i, θ_i), which represents the tangent to a geodesic curve, parameterized by t. The 1-parameter Lie algebra element of the spoke model is then given by

$$A_v(t) = \begin{pmatrix} A_{v_1}(t) & . & 0 \\ . & . & . \\ 0 & . & A_{v_n}(t) \end{pmatrix} \quad (5)$$

v denotes that the Lie algebra element is defined at a fixed $(v_1, v_2, ..., v_n)$, which represents the tangent to a geodesic curve, parameterized by t.

The parameter t in equation 5 sweeps out a 1-parameter sub-group, $H_v(t)$ of the Lie group G of spoke transformations. By choosing different v, curves corresponding to different 1-parameter sub-groups can be traced on the group manifold. For any $g \in G$, the distance between g and $H_v(t)$ is defined as

$$d(g, H_v) = \min \; d(g, \exp[\; A_v(t) \;]), t \in \Re \quad (6)$$

d is the Lie group distance between group elements. The concept of projection onto the 1-parameter sub-group can be defined as follows. First find,

$$t^* = \arg\min \; d(g, \exp[\; A_v(t) \;]), t \in \Re \quad (7)$$

then $\exp[\ A_v(t^*)\ ]$ is the projection of g onto H_v.

Analogous to the principal components of a vector space, there exist 1-parameter subgroups called the principal geodesic curves [4] which explain the variability of the data points lying on the manifold. Suppose that $g_1, g_2, ..., g_m \in G$, then the first principal geodesic is defined as H_{v^*}, where

$$v^* = \arg\min \sum_{i=1}^{n} d^2(\mu^{-1} g_i, H_v), v \in \Re^{2n} \tag{8}$$

Once v^* is determined, the projection of $\mu^{-1} g_i$ on H_{v^*} is computed. Each projection is denoted by p_i. Each of the quantities $\mu^{-1} g_i$ are multiplied by p_i^{-1} and are used in equation (8) to determine v^* corresponding to the next principal geodesic, and so on.

III. Approach

The input to the Lie group based classifier is a small set of labeled samples Y_j from each vehicle category,C_j. n_Y denotes the number of training samples in a category. The intrinsic mean μ_j and the principal geodesics H_{v^*} are computed for each vehicle category,C_j using the samples, $Y_j^m \in Y_j$, $1 \leq m \leq n_Y$

Once the intrinsic mean and geodesics are available for each C_j, recognition of an unlabeled sample x, can be carried out by finding the category with the closest intrinsic mean to x. The closest category is found by,

$$j^* = \arg\min \| \log(\mu_j^{-1} x) \|, j \in 1, 2, ..., N \tag{9}$$

N is the number of categories.

Thus, recognition can be based solely on the intrinsic mean of a category. However, it can be expected that there will be significant variation in shape over a category and so the intrinsic mean μ_j, will not in general be an effective basis for classification. Instead a hierarchical procedure is carried out, where some number of categories whose means are closest to x are chosen for further analysis using the geodesics of each candidate category, C_j

The first step is to remove the intrinsic mean from the sample. Define $x_j^0 = \mu_j^{-1} x$ which denotes the residual obtained by removing the intrinsic mean of category C_j from x. Next, x_j^0 is projected onto the first principal geodesic of C_j. Let p_j^1 denote the projection of x_j^0 on the first principal geodesic of C_j. $(A_v)_j^1$ denotes the 1-parameter Lie algebra element which corresponds to the first principal geodesic of C_j. It should be noted that

$$\begin{aligned} p_j^1 &= \exp[\ (A_v)_j^1(t_1^*)\ ] && (10) \\ t_1^* &= \arg\min\ d(x_j^0, \exp[\ (A_v)_j^1(t)\ ]), t \in \Re && (11) \end{aligned}$$

The Lie group distance between the residual x_j^0 and the projection p_j^1 is denoted as d_j^1. x_j^k denotes the residual obtained by removing the projection of x_j^{k-1} onto the k^{th} principal geodesic of category C_j namely, p_j^k from x_j^{k-1}.

$$\begin{aligned} x_j^k &= (p_j^k)^{-1} x_j^{k-1} && (12) \\ p_j^k &= \exp[\ (A_v)_j^k(t_{k^*})\ ] && (13) \\ t_k^* &= \arg\min\ d(x_j^{k-1}, \exp[\ (A_v)_j^k(t)\ ]) && (14) \end{aligned}$$

$(A_v)_j^n$ denotes the 1-parameter Lie algebra element of n^{th} principal geodesic of C_j. d_j^n denotes the Lie group distance between x_j^{n-1} and p_j^n,

$$d_j^n = \| \log[\ (p_j^n)^{-1} x_j^{n-1}\ \| \tag{15}$$

The vector of distances, d_j^n, can be calculated by finding the Lie distance between the residuals and the projections of the residuals on the principal geodesics and updating the residuals by multiplying with the inverse of the projections in an iterative manner.

Removing the intrinsic mean from the sample and successively projecting the residual onto the principal geodesics brings the residue close to the identity spoke configuration. The identity configuration is one in which all the spokes collapse onto a single spoke having a unit radius. For an unknown sample, successive projections onto the correct category geodesics brings the sample closer to identity compared to projections on the other categories.

It should be the case that for the true category, j^* , of sample there will be an n such that $d_{j^*}^n$ is smaller than the distance for any other category. The set of feasible categories is successively reduced by a hierarchical examination of principal geodesics for the candidate categories at each stage which is summarized below.

Algorithm

Given N categories, C_j

1) Find the distances to the intrinsic mean of each category. $d_j^0 = \| \log(\mu_j^{-1} x) \|$, $j \in 1, 2, ..., N$
2) Select the $m_0 < N$ categories, C_{j_1} which are the categories corresponding to the first m_0 values of distances d_j^0 sorted in ascending order.
3) For each j_1, compute the distances $d_{j_1}^1$ and reduce the number of candidates to $m_1 < m_0$ selecting the categories corresponding to the first m_1 values of $d_{j_1}^1$ sorted in ascending order.
4) Repeat step 3 for additional geodesics until there is a unique remaining category.

The choice of the number of feasible candidates remaining at each stage, m_i, is a tradeoff of computation and recognition accuracy. For the unlabeled instances lying close to intrinsic mean of a category, the Lie distance from the intrinsic mean is sufficient for correct classification. By eliminating the categories hierarchically, it is made sure that close ambiguities are always resolved at a very fine level.

IV. Experiments and results

In this section, the vehicle classification results obtained on the LIDAR vehicle dataset are presented. To compare the

(a)

(b)

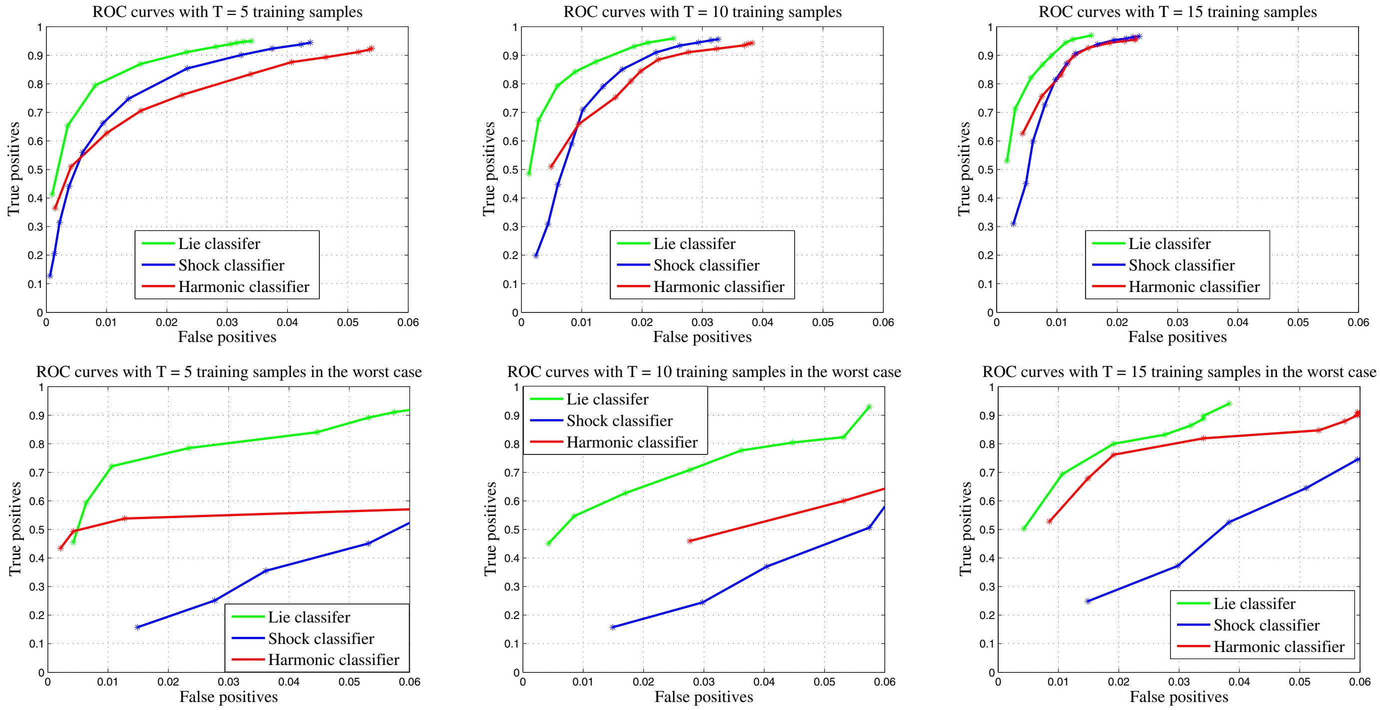

Figure 2 (a) shows the average ROC curves obtained on the 3-d LIDAR vehicle dataset using Lie group based classifier, shock graph based classifier and the spherical harmonic based classifier with the training sample set size varying as T = 5,10,15. (b) shows the worst-case ROC curves.

performance of Lie group based vehicle classifier with existing shape matching approaches, shock graphs [3] and rotation invariant spherical harmonics [2] are used to classify the same data set. The classification strategy for shock graphs and spherical harmonics based classifiers is to match the query vehicle instance to training samples of various categories and assign the class label corresponding to the sample with closest distance.

The vehicle database consists of 226 cars, 156 SUVs and 88 trucks. An initial set of T = 5 training samples from each category is used to do the classification of the database. To account for the variability in selecting the training samples, R = 100 sets of randomly drawn samples for each category are used as prototypes for vehicle classification. The rationale behind this experimental design is to eliminate the effect of poorly chosen training prototypes on recognition performance. The experiment was repeated for T = 10, 15 training samples and R = 100 sets of randomly drawn T samples. ROC curves are computed based on classification results averaged over the R sets of classification results as well as the worst set out of R sets and are shown respectively in figure 2 a and figure 2 b.

The Lie group classifier exhibits superior performance in all cases. Of particular interest is the demonstrated ability to generalize from worst-case training prototypes.

V. Conclusions and Future work

A novel hierarchical vehicle classification framework based on Lie group distance has been demonstrated. The validity of the approach is established by superior classification results obtained on a dataset of 3-d vehicle point clouds.

The proposed framework can be extended to account for occlusion by treating occlusion as a projection onto a sub-manifold of reduced dimension, accounting for missing parts of the shape.

References

[1] Huber D, Kapuria A, Donamukkala R and Hebert M. ‘Parts-based 3-D object classification’. *CVPR 2004*.
[2] Kazhdan M, Funkhouser T and Rusinkiewicz S. ‘Rotation invariant spherical harmonics representation of 3-d shape descriptors’. *SGP 2003*.
[3] Sebastian T, Klein P and Kimia B. ‘Recognition of shapes by editing their shock graphs’. *PAMI2004*.
[4] Fletcher T, Lu C and Joshi S. ‘Statistics of Shape via Principal Geodesic Analysis on Lie Groups’. *CVPR 2003*.
[5] Tuzel O, SubbaRao R and Meer P. ‘Nonlinear Mean Shift for clustering over Analytic Manifolds’. *CVPR 2006*.
[6] Tuzel O, Porikli F and Meer P. ‘Human Detection via Classification on Riemannian Manifolds’. *CVPR 2007*.
[7] Klassen E, Srivastava A,Mio W and Joshi S. ‘Analysis of Planar Shapes Using Geodesic Paths on Shape Spaces’. *PAMI 2004*.
[8] Rossmann W. ‘Lie Groups: An introduction through linear groups’. *Oxford Press 2002*.
[9] Ferryman J, Worrall A and Maybank S. ‘Learning Enhanced 3-d Models for Vehicle Tracking’. *BMVC 1998*.