



Review

A Literature Review: Geometric Methods and Their Applications in Human-Related Analysis

Wenjuan Gong ¹, Bin Zhang ², Chaoqi Wang ¹, Hanbing Yue ¹, Chuantao Li ¹, Linjie Xing ³, Yu Qiao ^{3,*}, Weishan Zhang ¹ and Faming Gong ¹

¹ The College of Computer Science and Communication Engineering, China University of Petroleum (East China), Qingdao 257061, China; wenjuangong@upc.edu.cn (W.G.); wcq86451612@163.com (C.W.); icyeyueHB@163.com (H.Y.); china_lichuantao@163.com (C.L.); zhangws@upc.edu.cn (W.Z.); gffaming@163.com (F.G.)

² The Beijing University of Posts and Telecommunications, Beijing 100876, China; fealie28@163.com

³ Key Laboratory of Human-Machine Intelligence-Synergy Systems, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China; csljxing@gmail.com

* Correspondence: yu.qiao@siat.ac.cn

Received: 1 April 2019; Accepted: 6 June 2019; Published: 23 June 2019



Abstract: Geometric features, such as the topological and manifold properties, are utilized to extract geometric properties. Geometric methods that exploit the applications of geometrics, e.g., geometric features, are widely used in computer graphics and computer vision problems. This review presents a literature review on geometric concepts, geometric methods, and their applications in human-related analysis, e.g., human shape analysis, human pose analysis, and human action analysis. This review proposes to categorize geometric methods based on the scope of the geometric properties that are extracted: object-oriented geometric methods, feature-oriented geometric methods, and routine-based geometric methods. Considering the broad applications of deep learning methods, this review also studies geometric deep learning, which has recently become a popular topic of research. Validation datasets are collected, and method performances are collected and compared. Finally, research trends and possible research topics are discussed.

Keywords: geometric methods; human action recognition; human pose estimation; human body shape analysis

1. Introduction

With the emergence of low-cost RGB-D cameras, human bodies can be digitized at a lower cost [1–3], and their actions can also be easily captured [4,5]. In 3D spaces (for point cloud models or meshes), studying the geometric attributes becomes natural. The geometric attributes (for example, the number of holes and the geometric adjacency of objects) of data are extracted, and methods for studying geometric attributes are proposed.

The notion of “geometric methods” is used in this review and refers to methods that study the geometric attributes of data, methods with geometric constraints, or generalized methods with spatial or temporal information. When dealing with continuous 3D models, certain geometries, like topology, Riemann manifold, conformal geometry, etc., are better choices. They are capable of describing the properties of the geometric object from the perspective of the geometric object. In a Euclidean space, the global coordinates are cumbersome for describing attributes along the object surface. For example, the geodesic distance is a better description for two points on a geometric object than the Euclidean distance; the geodesic distance is from the perspective of the points on the surface, and it considers the distance one point needs to traverse on the surface.

Based on the scope of geometric properties that are extracted and the way in which geometric properties are encoded, the review proposes to classify geometric methods into: object-oriented geometric methods, in which the geometric properties of the object surfaces are explored, feature-oriented geometric methods, in which the object features are extracted and the geometric properties of the feature space are explored, and routine-based geometric methods, in which geometric information is extracted following certain routines. For the first category, mathematical concepts and theorems are directly developed and applied. For the second category, geometric theorems can also be developed and applied, but under most circumstances, well-developed geometric concepts and theorems are utilized instead.

Geometric methods are advantageous in several aspects. In many application domains, the data reside on a manifold, e.g., the Grassmannian manifold [6,7], the hypersphere [8,9], or the manifold of symmetric positive definite (SPD) matrices [10–12]. Furthermore, using geometric methods results in the concise representation of data. For example, a sphere in a 3D Euclidean space is represented as a 2D geometry in the topology. Furthermore, geometric methods, such as Lie algebra, provide semantic meaning to data operations. For example, adding or subtracting two human body poses denoted by joint positions or limb angles in the Euclidean space has no meaning, while an addition in a Lie group results in a semantically-meaningful pose.

Comparatively, Euclidean distance is not suitable for certain computations, for example, comparing temporal sequences, which is critical for automatic video processing. It is not easy to discriminate sequences among classes. Some methods align the data before using a Euclidean metric, e.g., through dynamic time warping (DTW) [13], specialized kernels or a Fourier hierarchical pyramid [14]; other methods transform the data before using them, e.g., covariance features [15]. None of these methods consider the implicit dynamics of the sequences and the lower dimensional space where the features lie. Geometric spaces provide a possibility for solving this problem.

In this review, attributes and theories in non-Euclidean geometric spaces are explored. These geometric methods and their applications in human-related analysis are collected and studied. “Human-related analysis” (HRA) in this review denotes human shape analysis (HSA), human pose-related analysis (HPA), and human action-related analysis (HAA). HSA includes human shape matching, human shape segmentation, etc.; HPA includes human pose estimation, human posture estimation, human pose modeling, etc.; and HAA contains human action recognition, human activity recognition, etc. Geometric methods are effective solutions for human-related analysis.

Geometric methods were initially utilized in shape analysis and surface registration, which involves understanding the relationship between two geometric objects, for example in finding shape correspondences between two objects [16–20]. Furthermore, geometric methods are utilized in object recognition. For example, moduli space [21] provides a geometric solution for 3D face recognition. Other examples include describing properties in a local area instead of on a single point, like functional correspondences [22], or exploring geometric properties through linear algebra representations, like difference operators [23] among 3D shapes. Shape analysis methods for general objects can be generalized to HSA. Geometric methods can also be utilized for computing correlations between the human shape and another human attribute, e.g., between the shape and the age using manifold regression [24]. Notable works on geometric methods for HPA include Lie group representations of human poses [25]. For HAA, exemplary solutions include: methods of localizing humans in images and tracking and analyzing human motion trajectories and methods that directly extract spatial, temporal, or spatiotemporal patterns from image sequences [26,27].

Recently, many researchers focus on dealing with geometric data using deep learning methods. This research area is named as geometric deep learning and it attempts to generalize deep learning methods to non-Euclidean structured data such as graphs and manifolds. There are mainly two streams of methods: extrinsic methods and intrinsic methods. Extrinsic methods treat 3D data similar as 2D data but with one more dimension. One example of extrinsic methods is volumetric CNNs [28]. This representation is not invariant to deformations. In an intrinsic representation, the filter considers

local geometric properties and is applied along the object surface. There are also methods that extend deep learning to manifolds through embedding. Since there is not a canonical embedding for a general manifold, the researchers in [11,12,29,30] proposed solutions for the special case of Riemannian manifolds.

Despite the wide applications of geometric methods especially in HRA, there are few literature reviews. Related works are studied extensively in this literature review. Because geometric concepts require math foundations, mathematical knowledge of geometric spaces, such as basic concepts, geometric properties, and geometric measurements, are firstly introduced. The contributions of this paper include:

- Geometric methods and their applications in human-related analysis are extensively studied.
- Geometric methods are studied based on the scope in which they are applied, and we classify them into: feature-oriented geometric methods, object-oriented geometric methods, and routine-based geometric methods.
- Geometric methods and their performances on standard datasets are collected so that researchers who are interested in this topic can identify the state of the art.

The remainder of the paper is organized as follows. Section 2 introduces basic geometric concepts. Section 3 explores variant types of geometric methods. Section 4 explores specific geometric methods for human-related analysis. Section 5 introduces deep learning-based geometric methods. Section 6 studies generalized geometrics for human-related analysis. Section 7 collects a validation dataset for human-related analysis. Section 8 compares the performances of related works. Section 9 concludes the review and discusses future research trends. Figure 1 presents an overall view of the paper.

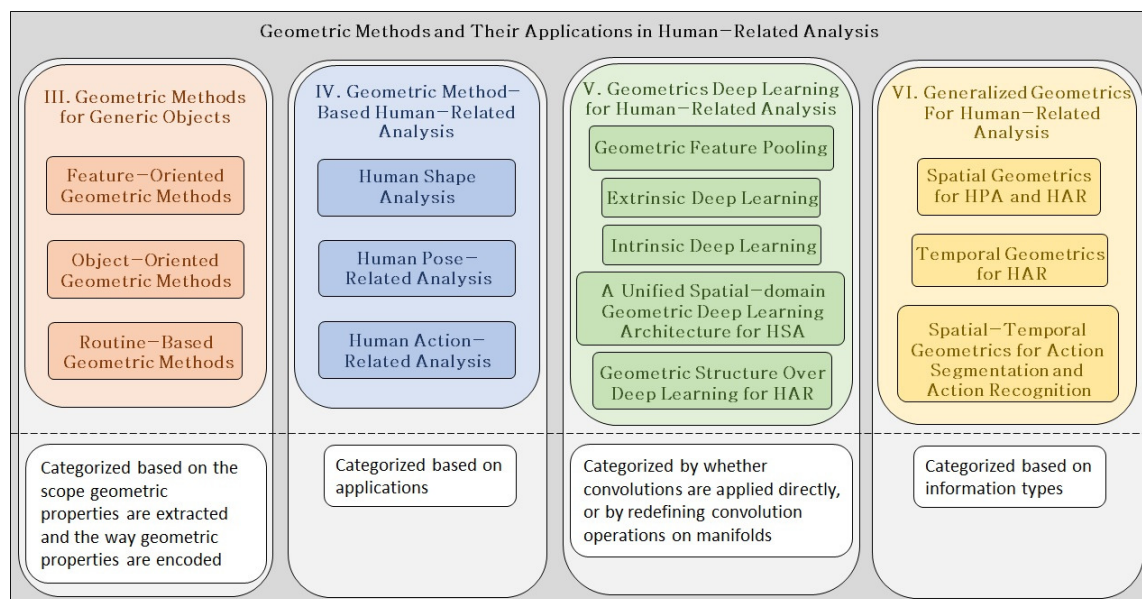


Figure 1. Overall view of this paper. This review is mainly composed of four modules: geometric methods for generic objects; geometric method-based human-related analysis; geometric deep learning for human-related analysis; and generalized geometrics for human-related analysis. Each module has its subsections, each of which is a class of methods based on its categorization standards. HSA, human shape analysis.

2. Basic Geometric Concepts

In this section, important concepts in the topology and manifolds that are widely used in geometric methods are introduced. Concepts of manifolds are usually difficult to conceive of and can be defined in various ways. This review selects among the different definitions of each concept and chooses the one that is easier to understand. It is more conceivable to consider topological concepts as being

developed from set theories, so set theories are firstly introduced. Many concepts of manifolds are developed based on topological concepts, so manifold concepts are introduced following topological concepts. Figure 2 shows the components of this section. Before presenting detailed definitions, the mathematical symbols are listed below.

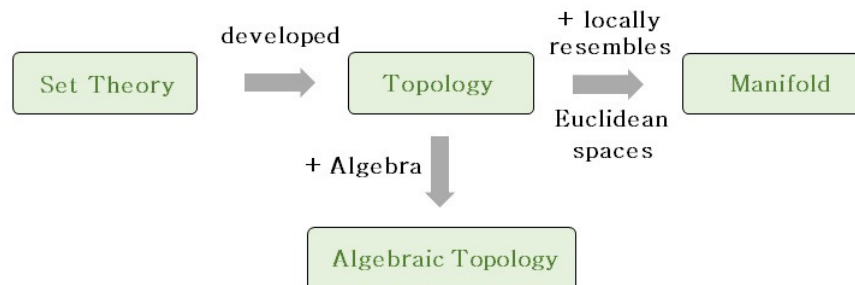


Figure 2. Main components of Section 2. This section is composed of four modules: set theory concepts; topology concepts developed from set theory; algebraic topology concepts (topology plus algebra); and manifold concepts (a topology that locally resembles Euclidean spaces).

2.1. Set Theory Concepts

An easier way to interpret topology is to consider topological concepts to be developed from set theories. Selected concepts from set theories are introduced in this section. Other concepts, like equivalence relation, equivalence class, and covering, are introduced in the Appendix A.

2.1.1. Metric

A metric or distance function on a set X is a real-valued function d defined on the Cartesian product $X \times X$ such that for all $x, y, z \in X$:

- $d(x, y) \geq 0$ with equality iff $x = y$.
- $d(x, y) = d(y, x)$.
- $d(x, y) + d(y, z) \leq d(x, z)$.

2.1.2. Quotient Vector Space

The quotient of a vector space by a subspace can be defined based on the equivalence class. Let V be a vector space over a field k and $W \subseteq V$ be a subspace. An equivalence relation \sim_W on V can be denoted by $v \sim_W v'$ if and only if $v - v' \in W$, where \sim_W is an equivalence relation. The quotient is denoted by V/\sim_W , and V/\sim_W is itself a vector space over k , with the addition and scalar multiplication rules satisfying $[v] + [v'] = [v + v']$ and $\kappa[v] = [\kappa v]$. V/\sim_W can also be denoted by V/W , which is referred to as the quotient space of V by W .

2.2. Topological Concepts

Topology is independent of any particular coordinate representation, while objects' representation in Euclidean spaces is certain. For example, every point in a three-dimensional Euclidean space is determined by three coordinates, while in topology, a global coordinate system does not exist. The self perspective is the essence of the conciseness in topological representations. It retains more general features, like the number of holes in the geometry while ignoring some fine details, like the distance functions. Specifically, topological properties of a shape are invariant under certain deformations: they do not change if the shape is stretched or compressed, but change under other deformations, like "tearing" or "adjoining". Topological concepts are selectively introduced in this section. Please refer to the Appendix for definitions of closed sets, the interior and closure of a set, limit points, continuous functions, quotient maps, Hausdorff space, and metrics.

2.2.1. Topology

Here, the geometric view of the topology developed from surfaces and neighborhoods is adopted. According to [31], for each point x of a set X , the neighborhoods of x are a non-empty collection of subsets of X and satisfy four axioms:

- x lies in each of its neighborhoods.
- The intersection of two neighborhoods of x is a neighborhood of x .
- If N is a neighborhood of x and if U is a subset of X that contains N , then U is a neighborhood of x .
- If N is a neighborhood of x and if $\overset{\circ}{N}$ denotes the set $\{z \in N | N \text{ is a neighborhood of } z\}$, then $\overset{\circ}{N}$ is a neighborhood of x (the set $\overset{\circ}{N}$ is called the interior of N).

The assignment of a collection of neighborhoods is called a topology on the set X . A topology defined with neighborhoods is easy to conceive of, but hard to work with. On the contrary, the topology based on openness is defined. A subset O of X is open if it is a neighborhood of each of its points. A topological space is then a set X together with a collection of open subsets Ω that satisfies the four conditions:

- The empty set \emptyset is in Ω .
- X is in Ω .
- The intersection of a finite number of sets in Ω is also in Ω .
- The union of an arbitrary number of sets in Ω is also in Ω .

2.2.2. Homeomorphism

A function $h : X \rightarrow Y$ is called a homeomorphism if it is one-one onto continuous and has a continuous inverse. When such a function exists, X and Y are called homeomorphic (or topologically equivalent) spaces. Figure 3 shows a homeomorphism between a sphere and a tetrahedron. The illustration shows a regular tetrahedron T projected onto a sphere with center \hat{T} using radial projections from the center.

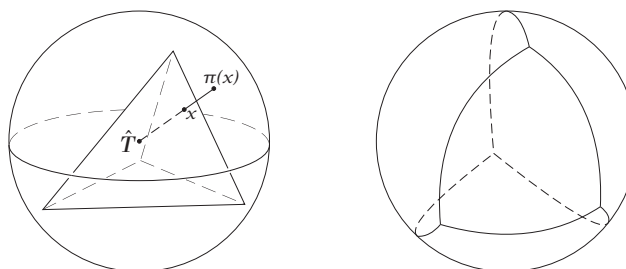


Figure 3. Radial projection from a tetrahedron T onto a sphere with center \hat{T} . An example is shown as follows: a point x on a surface of the tetrahedron projected onto its corresponding point $\pi(x)$ on the sphere with the radial projection function π .

2.2.3. Quotient Space

A quotient space is a set together with a topology. If X is a space and A is a set and if $p : X \rightarrow A$ is a surjective map, then there exists exactly one topology Ω on A relative to which p is a quotient map, and it is called the quotient topology induced by p .

A quotient space (also called an identification space) is, intuitively speaking, the result of identifying or “gluing together” certain points of a given topological space. Figure 4 shows an example of obtaining the two-sphere S^2 by gluing the circle S^1 to a single point.

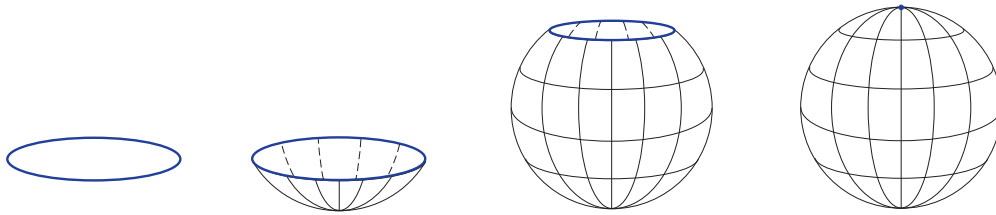


Figure 4. An example of creating a quotient space by gluing. Gluing the boundary of a circle onto a single point. The two-sphere S^2 is obtained by gluing the circle S^1 to a single point.

2.3. Algebraic Topology Concepts

Algebraic topology combines algebra with geometry by defining algebraic operations on geometric objects. The fundamental idea of algebraic topology is to develop methods for distinguishing between two topological spaces or two maps. The topological group is introduced. Please refer to the Appendix for the definitions of orbit space, homotopy, the fundamental group, and homology.

A topological group is a topological space with a binary operation and the inverse map, both being continuous. For example, G is a topological group if the multiplication map $(g, h) \mapsto gh$ from $G \times G$ to G and the inverse map $g \mapsto g^{-1}$ from G to G are both continuous. One extremely useful topology group is the general linear group. For example, the general linear group over \mathbb{R} , denoted by $GL(n, \mathbb{R})$, is the group of $n \times n$ invertible matrices with real entries.

2.4. Manifold Concepts

A manifold is both a Hausdorff space and a topological space that locally homeomorphic to the Euclidean space, that is, we can find a continuous bijective invertible mapping between a local area on the manifold and a local area in the Euclidean space. Furthermore, analysis can be carried out by imposing smooth structures on a manifold (similar as a differential Euclidean space). It is not sufficient to develop analysis on the manifold, but with certain methods (like parallel transport), tangent spaces at different points on the manifold are related. Essential concepts in the manifold are introduced in this section. Please refer to the Appendix for definitions of atlas, smooth manifold, section, vector bundle, fiber bundle, the tangent bundle of a vector bundle, vertical bundle, vector bundle homomorphism, vector bundle isomorphism, and connection.

2.4.1. Topological Manifold

Assume that \mathcal{M} is a topological space; \mathcal{M} is a topological manifold of dimension n if it has the following properties:

- \mathcal{M} is a Hausdorff space.
- \mathcal{M} is second countable: there exists a countable basis for the topology of \mathcal{M} .
- \mathcal{M} is locally Euclidean of dimension n : for each $p \in \mathcal{M}$, we can find an open set $U \in \mathcal{M}$ containing p , an open set $U' \in \mathbb{R}^n$, and a homeomorphism $\varphi : U \rightarrow U'$ (i.e., a continuous bijective map with the continuous inverse).

2.4.2. Chart

A chart, or a coordinate chart, on a manifold \mathcal{M} is a pair (U, φ) , where U is an open subset of \mathcal{M} and $\varphi : U \rightarrow \tilde{U}$ is a homeomorphism from U to an open subset $\tilde{U} = \varphi(U) \in \mathbb{R}^n$. Figure 5 illustrates an example of a coordinate chart.

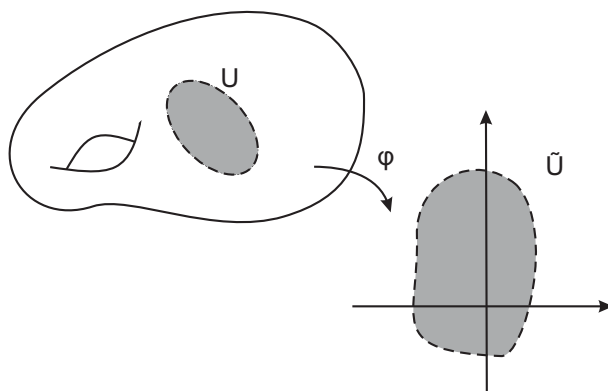


Figure 5. An example of a coordinate chart. The figure illustrates an example of a coordinate chart from U to \tilde{U} .

Let \mathcal{M} be a topological n -manifold. If (U, φ) and (V, ψ) are two charts such that $U \cap V = \emptyset$, then the composite map $\psi \circ \varphi^{-1} : \varphi(U \cap V) \rightarrow \psi(U \cap V)$, also called the transition map from φ to ψ , is a composition of homeomorphisms and is a homeomorphism.

Two charts, (U, φ) and (V, ψ) , are said to be smoothly compatible if either $U \cap V = \emptyset$ or the transition map $\psi \circ \varphi^{-1}$ is a diffeomorphism.

2.4.3. Tangent Space/Tangent Bundle

Let \mathcal{M} be a smooth manifold, $p \in \mathcal{M}$ and \mathcal{M}_p be the set of all smooth real-valued functions, each of which is defined on some open neighborhood of p . A tangent vector to \mathcal{M} at p is a map $v : \mathcal{M}_p \rightarrow \mathbb{R}$ such that:

- $v(\lambda f + \mu g) = \lambda v(f) + \mu v(g)$.
- $v(fg) = v(f)g(p) + f(p)v(g)$.

for all $f, g \in \mathcal{M}_p$, $\lambda, \mu \in \mathbb{R}$. The set of all tangent vectors to \mathcal{M} at p is denoted by $T_p\mathcal{M}$. It is called the tangent space to \mathcal{M} at p . Figure 6 shows an exemplary tangent space.

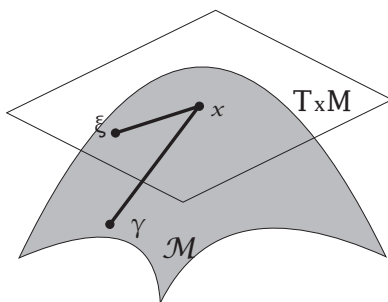


Figure 6. Illustration of a tangent space. $T_x\mathcal{M}$ is the tangent space of the manifold \mathcal{M} at point x .

The tangent bundle of a manifold \mathcal{M} is defined as the disjoint union of all tangent spaces to points of \mathcal{M} : $TM = \bigcup_{x \in \mathcal{M}} T_x\mathcal{M}$. Figure 7 shows an example of a tangent bundle of a circle. The figure illustrates the tangent bundle of a circle viewed from the side and from the top or bottom. Exemplary tangent spaces and their intersections with the circle are shown.

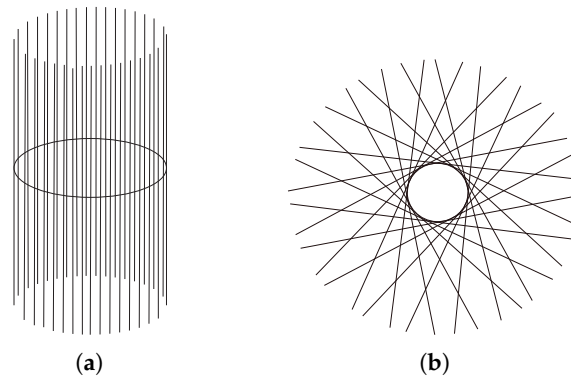


Figure 7. Illustration of a tangent bundle of a manifold. The figure illustrates the tangent bundle of a circle (a) viewed from the side and (b) viewed from the top or bottom.

2.4.4. Parallel Transport

Let \mathcal{M} be a smooth manifold with a vector bundle connection ∇ ; let $c : I \rightarrow \mathcal{M}$ be a differentiable curve from an interval I into \mathcal{M} ; and let $V_0 \in T_{c(t_0)}\mathcal{M}$ be a vector tangent to \mathcal{M} at $c(t_0)$ for some $t_0 \in I$. A vector field V is said to be a parallel transport of V_0 along c provided that $V(t)$ ($t \in I$) is a vector field for which $V(t_0) = V_0$. The notion of a parallel transport on a manifold \mathcal{M} clarifies the idea of translating a vector field V along a differentiable curve to attain a new vector field V' , which is parallel to V . Figure 8 shows an illustration of parallel transports under Levi-Civita connections. A Levi-Civita connection is a torsion-free metric connection preserving a given (pseudo-)Riemannian metric.

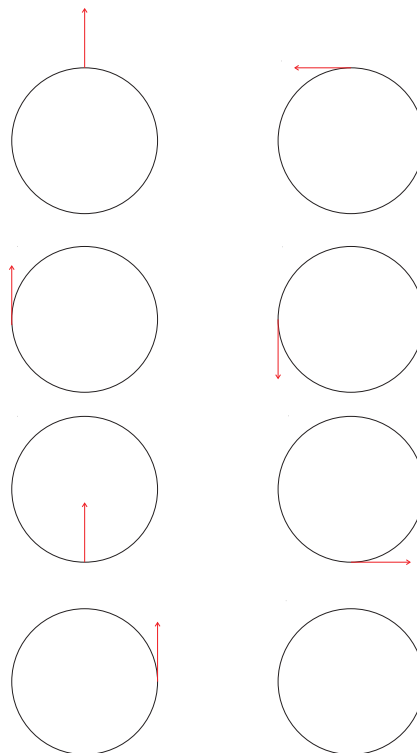


Figure 8. Examples of parallel transports. The figure illustrates two examples of parallel transports under Levi-Civita connections on four sampling positions. The transport on the left side is given by the metric $ds^2 = dr^2 + r^2 d\theta^2$. The transport on the right side is given by the metric $ds^2 = dr^2 + d\theta^2$.

2.5. Lie Group and Lie Algebra

A Lie group is a group G that is also an analytic manifold such that for $\sigma, \tau \in G$, the mapping $(\sigma, \tau) \rightarrow \sigma\tau^{-1}$ of the product manifold $G \times G$ into G is analytic. Lie algebra is a vector space \mathfrak{g} over

a field F with an operation $[\cdot, \cdot] : \mathfrak{g} \times \mathfrak{g} \rightarrow \mathfrak{g}$, which we call a Lie bracket, such that the following axioms are satisfied:

- Bilinearity: $[ax + by, z] = a[x, z] + b[y, z]$, $[z, ax + by] = a[z, x] + b[z, y]$ for all scalars a, b in F , and all elements x, y, z in \mathfrak{g} .
- Skew-symmetry or alternativity: $[x, x] = 0$, which implies $[x, y] = -[y, x]$ for all $x, y \in \mathfrak{g}$.
- Jacobi Identity: $[x, [y, z]] + [y, [z, x]] + [z, [x, y]] = 0$.

3. Geometric Methods for Generic Objects

In this section, various geometric methods are introduced. Reviewed methods are categorized based on the scope of the geometric properties that are extracted and the way in which the geometric properties are encoded. Some methods encode geometric attributes in features (see Section 3.1); some methods utilize concepts and theories from the topology and manifold and extract geometric properties on objects (see Section 3.2); and some methods extract geometric properties following certain procedures and denote the objects with structured representations, e.g., graph structures (see Section 3.3). There are also methods belonging to multiple categories. For example, the positive definite manifold-based methods in Section 3.1.2 belong to “feature-oriented geometric methods”, and they also belong to “object-oriented geometric methods”. This review selects a logically more appropriate categorization in the case mentioned above. In the following section, geometric methods are studied based on this method of categorization.

Other methods to incorporate geometric information, like regression-based methods [32–35], manifold diffeomorphisms [36], and others, are also utilized in applications like image processing. These methods are working on 2D objects and are difficult to extend to human-related analysis on 3D data, so they are not elaborated in this review.

3.1. Feature-Oriented Geometric Methods

A feature space is the space where an object is projected as a feature point. This section explores the geometric properties of the parameter spaces. Utilizing geometrics in a feature space can be implemented through exploiting neighboring properties of feature points, or through studying geometric attributes and geometric properties in the space.

3.1.1. Distance-Based Methods

Similarities among features extracted from the raw data can be calculated. Distances between sample pairs are extracted and are used to denote geometric attributes. Distances are constructed using similarity measures. The authors in [12] generalized from the case of vector space inputs to the case of a manifold. Distances on manifolds were calculated as geodesic distances between the data [37].

3.1.2. Positive Definite Manifold-Based Methods

Covariance matrices are used to capture representative features [38]. Covariance matrices describe the correlation between sampled data points. They are positive definite (PD) matrices and lie on PD manifolds. Temporal sequences are also capable of being embedded in the PD manifold. For example, the authors in [15] built a temporal hierarchy of covariance descriptors for human action classification. Works on computing distances on the PD manifold include [39–42].

To analyze covariance descriptors, Euclidean geometry is often not appropriate; thus, methods using non-Euclidean metrics have been proposed, e.g., [42,43]. In particular, Gram and Hankel matrices [44,45] and Bregman divergences [29,38,46–48] have been successfully applied in a number of covariance descriptor-based applications. Methods considering dynamic information have also been proposed [44,45], in which dynamic information is denoted with Hankel matrices and sequences are compared using the Hankel subspaces angle. Other examples include [49], in which the authors extended a locally aggregated descriptor (VLAD) to Riemannian manifolds.

In the special case of infinite dimensions, the authors in [50] extended covariance matrices into a Hilbert space.

3.1.3. Kernels over a Manifold

Kernels provide mathematical formulations for covariance matrices. The applications of this type of method include dictionary learning and sparse coding [29,30,51].

Usually, kernels over a manifold are implemented over the Riemannian manifold because the original manifold is required to have distance measures. A Riemannian metric on a manifold M is a smoothly-varying inner product $\langle \cdot, \cdot \rangle$ on the tangent space $T_x M$ at each point $x \in M$. A Riemannian manifold is a manifold equipped with a Riemannian metric. Some works embed Riemannian manifolds into the reproducing kernel Hilbert space (RKHS). RKHS is a linear space, so it is simple and effective representation. There are also other types of kernels, for example the geodesic exponential kernel in [52], which provides a kernel-based solution for the general Riemannian manifolds.

3.1.4. Moduli Space

For the specific task of classification, moduli space is a natural solution. Moduli spaces can be thought of as geometric solutions to geometric classification problems. Such spaces are the space of equivalence classes of complex structures, where two complex structures are deemed “the same” if they are equivalent by conformal mapping [53]. Two equivalent objects may look very different; but in a moduli space, equivalent objects have the same description, while inequivalent objects have different descriptions.

3.2. Object-Oriented Geometric Methods

In object-oriented methods, the geometric attributes of an object are extracted and studied.

3.2.1. Tangent Space-Based Methods

Tangent spaces (defined in Section 2.4.3) are associated with each point on a manifold. Some of the tangent space-based methods utilize mappings between the tangent space of the manifold and the manifold. An exponential map is a map from the tangent bundle of the manifold to the manifold. In addition, a logarithmic map is its reverse map. The exponential and logarithmic maps are illustrated in Figure 9. The authors in [12] used the Riemannian exponential and logarithmic maps to define a sparse representation on Riemannian manifolds. The formulation is a generalization of the linear sparsity condition to manifolds.

3.2.2. Conformal Geometry-Based Methods

Computational conformal geometry is an interdisciplinary field combining computing and conformal geometry. A conformal mapping is an angle-preserving mapping, and computational conformal geometry designs its algorithms in computing. The authors in [53] presented a thorough description of the theoretical foundations, as well as the practical algorithms of computational conformal geometry. A widely-used application of conformal geometry is in matching two object models. For example, the authors in [16] utilized it to find shape correspondences between two objects. It conformally maps the interior of an n -gon P bijectively to that of another n -gon Q . This mapping can be utilized to embed 3D meshes onto a 2D plane. However, when this map is extended to the boundary, it does not necessarily map the vertices of P to those of Q . For many applications, it is important to identify the “best” vertex-preserving mapping between two polygons, i.e., one that minimizes the maximum angle distortion. It can be considered as conformal geometric methods implemented in a greedy way. Such maps exist, are unique, and are known as extremal quasiconformal maps or Teichmüller maps.

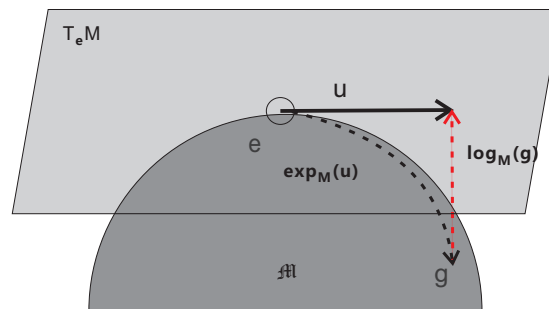


Figure 9. Illustration of the exponential and the logarithmic maps. The example point of g on the manifold M is mapped to a point on the tangent plane $T_e M$ using a logarithmic map $\text{Log}_M(g)$. The exponential map $\exp_M(u)$ is the reverse of the logarithmic map.

3.2.3. Principal Geodesic Analysis

Principal geodesic analysis (PGA) is an extension of principal component analysis (PCA) to manifolds. PGA has applications in shape analysis [54], and probabilistic PGA was utilized [55] to solve human activity recognition.

Since the objective function in the PGA algorithm is highly non-linear and generally difficult to solve efficiently, researchers who first introduced PGA [56] proposed a linear approximation. Exact computation can also be obtained under certain constraints. For example, the authors in [57] presented an exact computation of the PGA of data on the rotation group $SO(3)$. For constrained manifolds, like the constant curvature Riemannian manifolds in [58], optimization in PGA could be computed efficiently. The authors in [59] also proposed an exact PGA computation method without any linearization for data with a large variance.

3.3. Routine-Based Geometric Methods

Following certain routines, geometric information can also be encoded. Reducing representation dimensions, representing objects with a graph, and topological data analysis are all utilized to encode geometric information.

3.3.1. Dimension Reduction-Based Methods

Dimension-reduced representations (also called embeddings) are utilized to study feature space properties [60]. Considering the geometric properties of the feature representation, some non-linear dimension reduction algorithms have been utilized, e.g., the Laplacian eigenmaps (LE) framework, which recovers the low-rank structure of the manifold in a projected space. Laplacian eigenmaps [61] use graphs to find the embedding of the data in a low-dimensional space.

Furthermore, additional structures from low-dimensional data can be utilized as prior knowledge to enhance the representability of the models [62–64]. Discrete graphs are also utilized to incorporate data manifold information into the dimensionality reduction framework [65–73].

3.3.2. Graph-Based Methods

Graphs are concise representations for structural data. Graphs consist of units and connections. Units are connected if certain criteria are met. One wide application of graphs is to construct a mesh model from point clouds, in which units are connected if the distance between a pair is below a threshold. After graphs are constructed, clustering is usually utilized to explore the geometrically-adjacent attributes. One method of clustering a point cloud is single linkage clustering and its extensions [74,75]. In the single-linkage clustering methods, a graph is constructed with the vertex set as the set of points in the cloud and the edges as point connections if their distance is less than a threshold.

Under the assumption that high-dimensional data samples lie on or close to a smooth low-dimensional manifold, and the manifold can be approximated discretely as a graph, graphs can also be utilized to describe the low-dimensional intrinsic structure of the high-dimensional data. The emerging field of signal processing on graphs also facilitates the graph representation of signals [76].

3.3.3. Topological Data Analysis

In a broader perspective, topological data analysis (TDA) is an approach to analyzing data using topological methods [77,78] and is closely related to persistent homology, an adaptation of homology (defined in Appendix A.3.4) to point cloud data.

The TDA mentioned here refers to certain procedures for extracting topological properties from point cloud data. For example, the authors in [79] analyzed the geometric adjacency properties of an object and represented the object as a graph composed of nodes denoting key parts of the object. The graph considered the topological properties of the object. Topological properties are denoted by a topological network, i.e., a collection of nodes and a collection of edges connecting some of the nodes. Figure 10 shows the pipeline of the proposed method. TDA summarizes the data in a way that keep its global structure and local details to some degree, which is missing in other analysis methods, such as principal component analysis (PCA), multidimensional scaling (MDS), and cluster analysis.

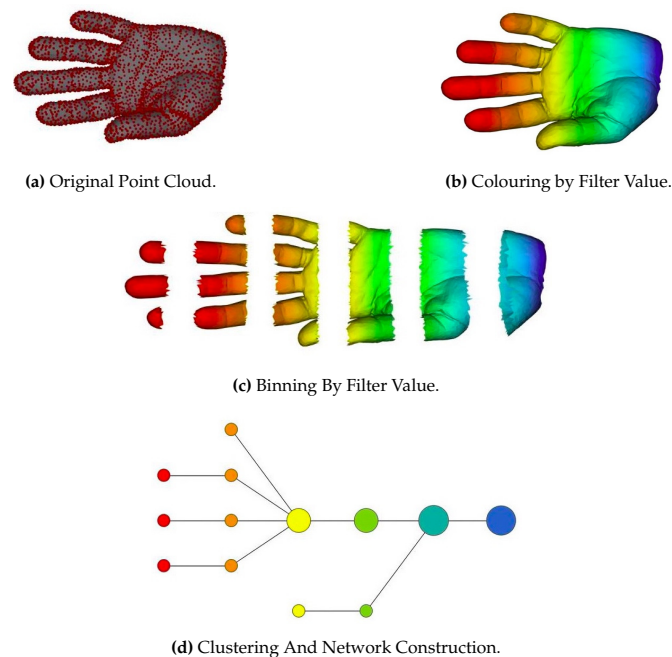


Figure 10. Illustration of a topological data analysis (TDA) pipeline. (a) A 3D object (hand) represented as a point cloud. (b) A filter value is applied to the point cloud, and the object is now colored by the values of the filter functions. (c) The data points are binned into overlapping groups. (d) Each bin is clustered and a center of the cluster is calculated, and a network is built by connecting the cluster center sequentially. The figure is originally from [79].

4. Geometric Method-Based Human-Related Analysis

For articulated objects, like human bodies, extrinsic properties are not capable of describing their intrinsic properties, like shapes and symmetric properties. Although suffering from topological noise, isometry-preserving properties are widely used for human-related analysis, e.g., the methods from [80–85] are utilized for human shape analysis, and the method from [86] is utilized for human shape recognition. Geometric methods are also isometry-preserving methods. In this section, geometric methods applied in human-related analysis are explored. Aiming at various application scenarios, different geometric methods are utilized including those introduced in Section 3. The methods in

this section are classified based on the applications. There are also works with literature reviews for specific applications, e.g., mesh segmentation [87], shape analysis [88,89], or shape retrieval [90].

4.1. Human Shape Analysis

General shape analysis has a wider scope than HSA. Shape comparisons, computing shape summary statistics, mathematical modeling of shape variations, and shape synthesis are all included in general shape analysis. 3D human shape synthesis is plausible using general shape synthesis methods, and this review concentrates on analyzing the human models instead of editing them, so shape synthesis is not the focus of this review. Shape summary statistics and shape variation modeling-related methods are discussed in the human pose-related analysis subsection. In this section, shape comparisons are discussed.

In computer graphics, object shapes are usually compared through a metric, or the dissimilarity measure. Geodesics are important for computing distances between object samples in representation space (e.g., a shape space) or on the shape surface. Spectral analysis is one widely-used method for measuring 3D human shape geodesics. Spectral analysis is an analysis in terms of eigenvalues (e.g., heat kernel signature-based method in Section 4.1.1), frequency spectrum (e.g., the learned spectral descriptor-based method in Section 4.1.3), etc.

Furthermore, diffusion geometry has been studied and utilized to describe intrinsic geometric properties of objects. In diffusion geometry, the distances between points are denoted in a way so that this is transformed into a metric learning problem, and various kernels are used, including the heat kernel, the wave kernel, etc.

4.1.1. Heat Kernel-Based Methods

The behavior of a quantum particle on the manifold is modeled by the Schrödinger equation:

$$\left(i\Delta_{\mathcal{M}} + \frac{\partial}{\partial t}\right)\psi(x, t) = 0, \quad (1)$$

where $\psi(x, t)$ is the function capturing the particle behavior, and $\Delta_{\mathcal{M}}$ is the Laplace–Beltrami operator (LBO) of \mathcal{M} :

$$\Delta := \operatorname{div} \nabla, \quad (2)$$

which is the divergence of the gradient. The divergence is the extent to which some quantity is exiting an infinitesimal region of a space, and the gradient is a multi-variant version of the derivative. LBO is the generalization of the Laplacian on Riemannian manifolds.

Given an initial heat distribution $f : M \rightarrow R$, let $H_t(f)$ denote the heat distribution at time t : $H_t = e^{-t\Delta_{\mathcal{M}}}$. The heat kernel is based on the exponential function of the eigenvalues of the LBO [91]: $k_t(x, y) : \mathbb{R}^+ \times M \times M \rightarrow R$ and satisfies $H_t f(x) = \int_M k_t(x, y) f(y) dy$, where dy is the volume form at $y \in M$:

$$k_t(x, y) = \sum_{i=0}^{\infty} e^{-\lambda_i t} \phi_i(x) \phi_i(y). \quad (3)$$

The heat kernel signature (HKS) [92] is a dense descriptor constructed by considering the diagonal of the heat kernel:

$$k_t(x, x) = \sum_{k \geq 0} e^{-t\lambda_k} \phi_k^2(x). \quad (4)$$

It is also known as the autodiffusivity function. Additionally, the HKS of dimension Q at point x is defined by sampling the autodiffusivity function at some fixed times t_1, \dots, t_Q :

$$\mathbf{f}(x) = (k_{t_1}(x, x), \dots, k_{t_Q}(x, x))^T. \quad (5)$$

4.1.2. Wave Kernel Signature-Based Methods

The wave kernel signature (WKS) evaluates the probability of a quantum particle being located at a point of a manifold under a certain energy distribution. The probability of finding the particle at point x is given by:

$$p(x) = \lim_{T \rightarrow \infty} \int_0^T |\psi(x, t)|^2 dt = \sum_{k \geq 1} \pi^2(\lambda_k) \phi_k^2(x). \quad (6)$$

The definition depends on the initial frequency distribution $\pi(\lambda)$. For example, the authors in [93,94] considered a log-normal frequency distribution $\pi_\nu(\lambda) = \exp(\frac{\log \nu - \log \lambda}{2\sigma^2})$ with mean frequency ν and standard deviation σ . The Q -dimensional wave kernel signature (WKS) is defined as:

$$\mathbf{f}(x) = (p_{\nu_1}(x), \dots, p_{\nu_Q}(x))^T, \quad (7)$$

where $p_\nu(x)$ is the probability Equation (6) corresponding to the initial log-normal frequency distribution with mean frequency ν , and ν_1, \dots, ν_Q are some logarithmically-sampled frequencies.

4.1.3. Learned Spectral Descriptor-Based Methods

Under the proposition that the descriptor should consider the statistics of the corpus of shapes (for example, thin and fat human models) and those of the class of transformations (such as human pose variations), the authors in [95] proposed a learning scheme for the construction of optimized spectral descriptors and formulated the descriptor in a generic form:

$$\mathbf{f}(x) = \sum_{k \geq 1} \boldsymbol{\tau}(\lambda_k) \phi_k^2(x) \approx \boldsymbol{\tau}(\lambda_k) \phi_k^2(x), \quad (8)$$

where $\boldsymbol{\tau}(\lambda) = (\tau_1(\lambda), \dots, \tau_Q(\lambda))^T$ is a bank of transfer functions acting on the LBO eigenvalues, and the parametric transfer function:

$$\tau_q(\lambda) = \sum_{m=1}^M a_{qm} \beta_m(\lambda) \quad (9)$$

is defined in terms of the B-spline basis $\beta_1(\lambda), \dots, \beta_M(\lambda)$ and the parametrization coefficients a_{qm} ($q = 1, \dots, Q, m = 1, \dots, M$). Plugging Equation (9) into Equation (8), the q^{th} component of the spectral descriptor is represented as:

$$f_{q(x)} = \sum_{k \geq 1} \tau_q(\lambda_k) \phi_k^2(x) = \sum_{m=1}^M a_{qm} \underbrace{\sum_{k \geq 1} \beta_m(\lambda_k) \phi_k^2(x)}_{g_m(x)}, \quad (10)$$

where $\mathbf{g}(x) = (g_1(x), \dots, g_M(x))^T$ is a vector-valued function dependent only on the intrinsic geometry of the shape. Thus, Equation (8) is parametrized by the $Q \times M$ matrix $\mathbf{A} = (a_{qm})$ and can be written in matrix form as $\mathbf{f}(x) = \mathbf{A}\mathbf{g}(x)$. The main idea of [95] is to learn the optimal parameters \mathbf{A} by minimizing a task-specific loss, which reduces to Mahalanobis-type metric learning.

Figure 11 visualizes the distances computed from the three kernels mentioned in this section, and Figure 12 shows the computed correspondences between two human models using the three kernels.



Figure 11. Three kernel-based distance visualized on human models. Visualized distances between the reference point (pointed with red arrows in the first column of each sub-group) and other points on the model. On the left, the reference point is the right wrists, in the middle the belly, and on the right the chest. The first row shows the results from the heat kernel, the second row shows the results from the wave kernel, and the third row shows the results of the proposed kernel in [95]. Dark blue shows small distances; red represents large distances. ©2014 IEEE. Reprinted, with permission, from R. Litman, and A. M. Bronstein, Learning Spectral Descriptors for Deformable Shape Correspondence, in *IEEE Trans. Pattern Anal. Mach. Intell.*, 2014, 36, 170–180.



Figure 12. Heat kernel signature (HKS), wave kernel signature (WKS), and learned spectral descriptors for point matching between human models. Correspondences computed on TOSCA shapes with geodesic distance distortion below 10% of the shape diameter using the heat kernel signature, wave kernel signature, and learned spectral descriptor (from left to right) [95]. ©2014 IEEE. Reprinted, with permission, from R. Litman, and A. M. Bronstein, Learning Spectral Descriptors for Deformable Shape Correspondence, in *IEEE Trans. Pattern Anal. Mach. Intell.*, 2014, 36, 170–180.

4.2. Human Pose-Related Analysis

Pose space deformation methods are widely used in human pose morphing. Based on the pose space deformation methods, model reduction has proven useful to increase the performance of static pose-space deformation both with [96–98] and without dynamics [99]. Given morphing targets, some works [96] constructed a single pose-independent basis by performing PCA on the sets of bases computed at the underformed configuration. Others obtained the basis by performing PCA on full simulation data [97–99]. To accommodate large deformations, the basis can be improved using modal derivatives [100] or linear transformations of the basis [101].

Pose-space subspace methods are utilized in human pose representation to construct reduced-order models with pose-dependent bases [102]. Variant subspace is computed for each representative set and these subspace is further combined into a dynamic system.

In Euclidean space, adding two poses might result in a physically-infeasible pose. Methods for representing 3D human poses with Lie groups have been proposed to solve this problem [25]. Lie group theory provides a semantically meaningful space for adding and subtracting human poses.

4.3. Human Action-Related Analysis

Human actions are recognizable from both still images and videos (or image sequences). When dealing with videos (or image sequences), temporal information is beneficial to boost the action recognition accuracy.

4.3.1. Relative 3D Geometry-Based Methods for Human Action Recognition

Many of the skeleton-based approaches for human action recognition use joint locations and joint angles to represent human poses. For example, the authors in [103] introduced a family of skeletal representations for HAR. The family of the proposed features used the relative 3D rotations between various body parts. They were split into two groups: four transformation-based features and two rotation-based features. Using the proposed representations, human actions are modeled as curves in the R3DG feature space (illustrated in Figure 13). Action recognition is then performed by classifying these curves with a combined method of dynamic time warping, Fourier temporal pyramid representation, and support vector machines.

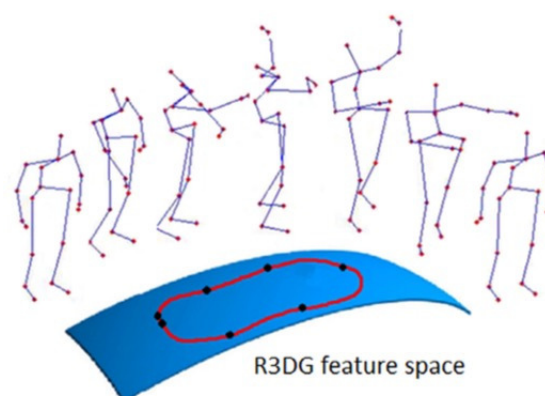


Figure 13. An action trajectory in R3DGfeature space. One point on the action trajectory is an R3DG feature of a pose [103]. Reprinted from Comput. Vis. Image Underst., Vol. 152, R. Vemulapalli, F. Arrate, and R. Chellappa, R3DG features: Relative 3D geometry-based skeletal representations for human action recognition, 155–166, Copyright 2016, with permission from Elsevier.

4.3.2. Matrix Embedding for 3D Human Action Recognition

Hankel matrices carry useful invariant properties, e.g., the rank of the Hankel matrix measures the complexity of the underlying dynamics [45]. Despite its advantages, Hankel matrices are not robust against noise. The authors in [104] embedded the sequences into a Riemannian manifold by using positive definite regularized Gram matrices of their Hankalets. Gram matrices inherit the rank and invariance properties of the associated Hankel matrices. Furthermore, Gram matrices are confined to the positive semi-definite (PSD) manifold and capture the underlying geometry better than directly comparing the sequences or Hankel matrices.

4.3.3. Graph-Based Human Action Recognition

Graph-based algorithms have been widely used for action recognition in conventional RGB videos [105–107]. Interesting works include graph representations for high-level features. For example,

the authors in [108] proposed a graph representation for skeleton-based 3D action recognition. A node of the graph is modeled as a motionlet, which is a semantic part of the trajectory of a joint. The edge is labeled as spatiotemporal relationships between connected motionlets. Constructed graphs are decomposed into substructures called subgraphs, and these subgraphs are compared based on a proposed graph kernel named the subgraph-pattern graph kernel (SPGK). The proposed kernel considers both spatial and temporal information. To circumvent the NP-hard problem of extracting all subgraph patterns from a graph, the authors reformulated the kernel using dynamic programming.

4.3.4. Lie Group-Based Human Action Recognition

Given human skeletal representations in a Lie group, human actions can be represented as curves in this Lie group. The authors in [109] used this type of method. First, a skeletal representation was proposed to explicitly model the 3D geometric relationships between various body parts using rotations and translations in the 3D space. The proposed skeletal representation lies in the Lie group $SE(3) \times \dots \times SE(3)$, which is a curved manifold. Using the proposed representation, human actions can be modeled as curves in this Lie group. Due to the difficulty of classifying curves in the Lie group, the action curves from the Lie group are mapped to its Lie algebra, which is a vector space. Then, classification is performed with a combined method of dynamic time warping, Fourier temporal pyramid representation, and linear SVM.

The authors in [110] used a similar pipeline of first representing skeletons with Lie groups and then classifying the actions, represented as curves, in Lie groups. Specifically, each skeleton is represented using the relative 3D rotations between various body parts. The skeletal representation is a point in the Lie group $SO(3) \times \dots \times SO(3)$. Then, using this representation, human actions are modeled as curves in this Lie group. The action curves are mapped onto its Lie algebra by combining the logarithm map with rolling maps, and classification is performed in the Lie algebra.

4.3.5. Dynamic Manifold Warping for Human Action Recognition

For temporal misalignment problems on a manifold, dynamic time warping algorithms are adapted for solving human action recognition problems. For example, the authors in [111] proposed a spatiotemporal manifold (STM) model to analyze human action trajectories with latent spatial structure. Action sequences were aligned with respect to latent parameters, which encoded a path as a point moving on a manifold from a starting point with a parameter value of zero to an ending point with a parameter value of one. In addition, a motion similarity metric was proposed for human action sequences, both in 2D and 3D.

5. Geometric Deep Learning for Human-Related Analysis

Deep learning has achieved remarkable performance breakthroughs in speech recognition, natural language processing, and computer vision. In particular, convolutional neural network (CNN) architectures perform well on many image analysis tasks such as classification [112], segmentation [113–115], regression [116], and synthesis tasks [117]. A convolution can be thought of as a template matching with filters, and convolution operations on a whole image are carried out by a sliding window procedure. In the case of processing images, one extracts a patch of pixels within a window, correlates it with a template, and moves the window to the next position. Recently, geometric deep learning [118,119] has been the focus of considerable research attention (<http://geometricdeeplearning.com/>, <https://sites.google.com/site/deepgeometry/>), while literature reviews on specific applications remain absent.

In this section, geometric deep learning methods and their applications in human-related analysis are studied extensively. Based on how geometric information is utilized, by directly applying traditional convolution operations to geometric objects or by redefining convolution operations and traversing methods on manifolds, geometric deep learning is classified into extrinsic deep learning methods and intrinsic deep learning methods. Figure 14 compares these two types of methods

implemented with CNN. Extrinsic CNN (the left subfigure in Figure 14) extends the traditional convolution operation from 2D to 3D and does convolution using the 3D templates shown as the cube in the figure. On the contrary, intrinsic CNN (the right subfigure in Figure 14) defines convolution on the manifold, i.e., along the object surface, and the dimensions of the convolution operations can be considered as 2D.

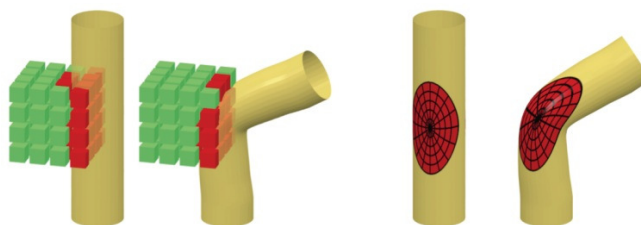


Figure 14. Illustrations of the differences between extrinsic CNN and intrinsic CNN. Intrinsic methods (right) work on the manifold rather than its Euclidean realization. The figure is originally from [120]. Reproduced with permission from Michael Bronstein, NIPS Proceedings; published by Neural Information Processing Systems Foundation, Inc., 2016.

Feature pooling is also an important module in the deep learning architecture, and it is crucial for dimension reduction. Therefore, Section 5.1 introduces feature pooling methods, and the rest of this section explains various ways to define convolutions.

5.1. Geometric Feature Pooling

Feature pooling is a key component for reducing representation dimensions. Two prevailing pooling techniques, namely average and max poolings, are not theoretically optimal due to the unrecoverable loss of the spatial information. The authors in [121] proposed generalizing previous pooling methods towards a weighted ℓ_p -norm spatial pooling function tailored for class-specific feature distributions. Specifically, the pooled features are weighted by the image location of a specific visual word. The original method was proposed under the bag of words (BoW) pipeline, but theoretically, it can be adapted to the deep learning architecture.

5.2. Extrinsic Deep Learning

Deep CNNs have recently been adapted to process 3D data by generalizing standard 2D convolutions to 3D. These methods of treating geometric data are called extrinsic methods. Their applications include processing 3D geometric shapes, for example, 3D object detection from RGB-D data [122], object classification of point clouds data [123], 3D object local feature matching [124], and 3D deformation flows [125].

5.2.1. Volumetric CNN for Shape Analysis

A natural extension to the classic CNN that processes 2D images is to process 3D data using a volumetric representation and perform 3D convolutions. The authors in [28] presented a 3D deep learning framework for modeling shapes using a voxel representation for 3D object shapes, called ShapeNets. The approach represents a geometric 3D shape as a probabilistic distribution in a voxel grid, and a convolutional deep belief network is used to learn the joint distribution of all voxels. The dataset and the source code are available (<http://3DShapeNets.cs.princeton.edu>). This generic shape analysis algorithm is applicable to human body models.

5.2.2. Geometric Constrained Extrinsic CNN for Human Shape Analysis

Instead of adapting the convolution operations in the network, geometric information can also be incorporated through other measures. Traditional classification neural networks tend to separate the surface points lying in different, but nearby classes, which results in ambiguous point

categories at the segmentation boundaries. To solve this problem, the authors in [126] proposed smoother feature representations. The CNN network consists of layers of descriptor extractions and a classification layer and removing the classification layer after training leaves the descriptor extraction network. This architecture is widely used for feature extraction. Extracted features are then fused with an ensemble of classification tasks. To ensure descriptor smoothness, the authors proposed randomizing the dense-label generation procedure. Specifically, multiple segmentations of the same person were considered (shown in Figure 15), and a classification problem was introduced for each. The source code and the dataset are available (<https://github.com/halimacc/DenseHumanBodyCorrespondences>).

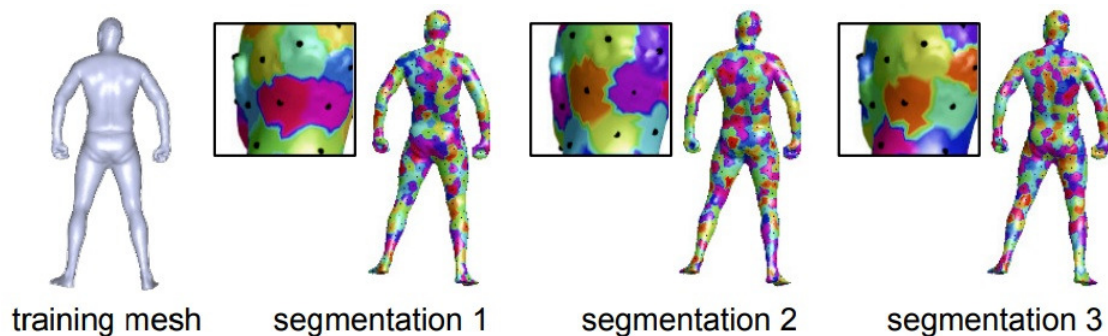


Figure 15. A Training Mesh Example with Its Multiple segmentations. To ensure smooth descriptors, the authors in [126] defined a classification problem for multiple segmentations of the human body. Points on the boundary might be assigned to nearby classes in different segmentation. ©2016 IEEE. Reprinted, with permission, from L. Wei, Q. Huang, D. Ceylan, E. Vouga, and H. Li, Dense Human Body Correspondences Using Convolutional Networks, in *Proceedings of the European Conference on Computer Vision*, Amsterdam, The Netherlands, 11–14 October 2016, 1544–1553.

5.3. Intrinsic Deep Learning

Alternatively, the convolution operations and how the convolution operates over the whole object are redefined on a manifold. This type of methods are called intrinsic methods.

5.3.1. Spatial-Domain Geometric CNN for Human Shape Analysis

A straightforward method for defining an intrinsic equivalent of a convolution is through the spatial domain. One method is to consider local receptive fields, in which the grid is replaced by a weighted neighborhood. Figure 16 shows an exemplary construction of a spatial-domain geometric CNN.

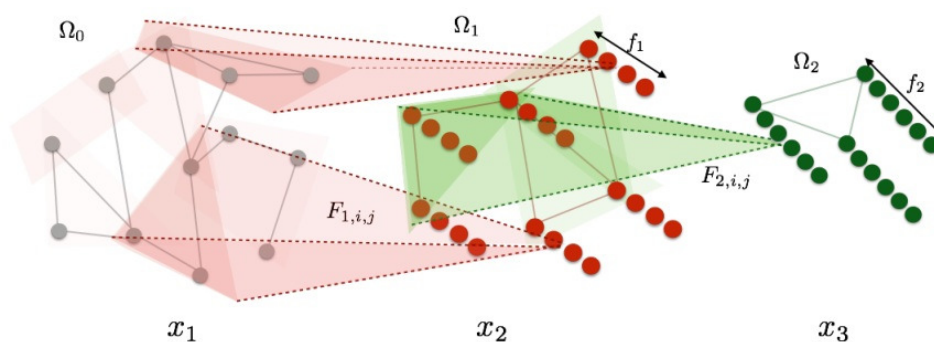


Figure 16. Spatial construction of geometric CNN. $K(K = 2$ in the example) scales are considered. Ω_k is defined as a partition of Ω_{k-1} into d_k clusters. Each layer of the network transforms a f_{k-1} -dimensional signal indexed by Ω_{k-1} into a f_k -dimensional signal indexed by Ω_k . The figure is originally from [127].

Another approach utilizes local polar coordinate systems. The authors in [128] defined the patch operator as a combination of Gaussian weights defined on a local polar system of coordinates (shown in Figure 17). After extracting the local geodesic coordinate system, the geodesic patch operator is defined as:

$$(D(x)f)(\theta, \rho) = \int_X f(x') w_\theta(x, x') w_\rho(x, x') dx', \quad (11)$$

$$w_\theta(x, x') = e^{-d_X^2(\Gamma(x, \theta), x')/2\sigma_\theta^2}, \quad (12)$$

$$w_\rho(x, x') = e^{-(d_X(x, x') - \rho)^2/2\sigma_\rho^2}, \quad (13)$$

where w_θ and w_ρ are the angular weight and the radial weight, respectively. An angular max pooling was used due to the difficulties of fixing the angular axes at each sampled point, leading to the following definition of the geodesic convolution:

$$(f \star w)(x) = \max_{\Delta\theta \in [0, 2\pi)} \int w(\theta + \Delta\theta, \rho) (D(x)f)(\theta, \rho) d\theta d\rho. \quad (14)$$

Furthermore, Fourier transform layers and covariance layers are also defined to transform signals into the frequency domain and inspect the global features from all input dimensions.

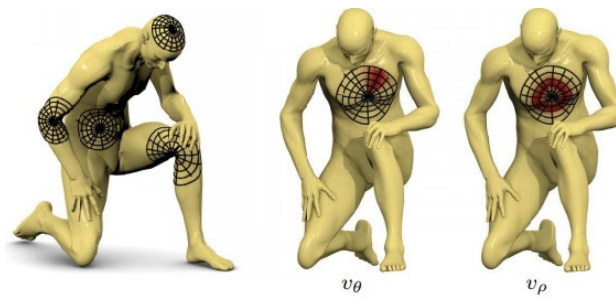


Figure 17. Visualized local geodesic polar coordinates. Left: examples of local geodesic patches, center and right: examples of angular weights and radial weights, v_θ and v_ρ , respectively (red denotes larger weights) [128]. ©2015 IEEE. Reprinted, with permission, from J. Masci, D. Boscaini, M. M. Bronstein, and P. Vandergheynst, Geodesic Convolutional Neural Networks on Riemannian Manifolds, in *Proceedings of the IEEE Workshop on 3D Representation and Recognition*, Santiago, Chile, 17 December 2015, 832–840.

5.3.2. Spectral Analysis-Based Intrinsic CNN

Another type of method generalizes the convolution operator with the spectrum analysis. A fundamental result of classical Euclidean signal processing states that the Fourier transform diagonalizes the convolution operator [119]. Then, convolutions may be extended to general manifolds by finding the corresponding basis. In the case of graph representations, the convolution operator can be carried out with the spectrum of its graph Laplacian. For example, in [127], convolution operations are defined as follows: for each layer $k = 1 \dots K$, an input vector x_k of size $|\Omega| \times f_{k-1}$ is transformed into an output x_{k+1} of dimensions $|\Omega| \times f_k$:

$$x_{k+1,j} = h \left(V \sum_{i=1}^{f_{k-1}} F_{k,i,j} V^T x_{k,i} \right) \quad (j = 1 \dots f_k), \quad (15)$$

where $F_{k,i,j}$ is a diagonal matrix, V is composed of the eigenvectors of the Laplacian, and h is a real-valued non-linear function. In addition, filters with constant spatial support are obtained by choosing specific sampling steps in the spectral domain.

Localized Spectral CNN for Human Shape Analysis

One drawback of spectral analysis is the difficulty in the spatial localization. Spectral analysis is global because the basis functions are global. There are studies that specialize in spatial localization through operations on the spectral domain. In [129,130], these operations were achieved through windowed Fourier transform on the spectral domain.

The windowed graph Fourier transform (WGFT) of a signal f [129,130] can be defined through the filtering signal g :

$$(Sf)(x, k) := \langle f, g_{x,k} \rangle, \quad (16)$$

where $g_{x,k}(n)$ is a windowed element centered at vertex x and frequency k :

$$\begin{aligned} g_{x,k}(n) &:= (M_k T_x g)(n) \\ &= N \chi_k(n) \sum_{l=0}^{N-1} \hat{g}(\lambda_l) \chi_l^*(x) \chi_l(n). \end{aligned} \quad (17)$$

Then, WGFT can be reformulated as:

$$(Sf)(x, k) = \sum_{l=0}^{N-1} \hat{g}_l \chi_l^*(x) \langle f, \chi_l \chi_k \rangle. \quad (18)$$

The WGFT $(Sf)(x, k)$ filters signal f at point x at frequency k with a window defined by \hat{g}_l .

By collecting its behavior over different frequencies, the content of signal f in a local support around x is extracted, thus reproducing the window extraction on images. The localized spectral convolution layer can thus be defined as:

$$\mathbf{f}_q^{out}(x) = \sum_{p=1}^P \sum_{k=1}^K w_{q,k,p} |(S\mathbf{f}_p)(x, k)|, \quad (19)$$

where \mathbf{f}_p ($p = 1, \dots, P$) is the input signal, $\mathbf{W} = (w_{q,k,p})$ is a $Q \times K \times P$ tensor representing the learnable weights, and \mathbf{f}_q^{out} ($q = 1, \dots, Q$) is the output signal.

5.3.3. Heat Diffusion CNN for Human Shape Analysis

The heat diffusion equation is also used for extending traditional CNN to a manifold. Heat diffusion measures heat diffused on a manifold. The heat propagation on a shape X is governed by the heat diffusion Equation (1). Given the initial heat distribution a delta function centered on x , the heat distribution on X after some time t is represented by the heat kernel $h_t(x, \cdot)$. The heat kernel, as formulated in Equation (3), is isotropic. Generalized heat diffusion is described by the anisotropic diffusion equation:

$$f_t(x, t) = -\text{div}_X(\mathbf{A}(x) \nabla_X f(x, t)), \quad (20)$$

where ∇_X and div_X denote the intrinsic gradient and divergence operators and $f(x, t)$ is the temperature at point x at time t . The thermal conductivity matrix $\mathbf{A}(x)$ specifies the heat conductivity properties at each point on shape X . The general diffusion model can be utilized for shape analysis [131].

The authors in [120] defined the thermal conductivity matrix as:

$$\mathbf{A}_{\alpha\theta}(x) = \mathbf{R}_\theta(x) \begin{pmatrix} \alpha & \\ & 1 \end{pmatrix} \mathbf{R}_\theta(x)^T, \quad (21)$$

where the matrix $\mathbf{R}_\theta(x)$ performs rotation of θ w.r.t. the reference direction (e.g., the maximum curvature) and $\alpha > 0$ is a parameter controlling the degree of anisotropy.

In the spectral domain, the anisotropic heat kernel is given by:

$$h_{\alpha\theta t}(x, x') = \sum_{k \geq 0} e^{-t\lambda_{\alpha\theta,k}} \phi_{\alpha\theta,k}(x) \phi_{\alpha\theta,k}(x'), \quad (22)$$

where $\phi_{\alpha\theta,k}(x)$ and $\lambda_{\alpha\theta,k}$ are the eigenfunctions and eigenvalues of the anisotropic Laplacian $\Delta_{\alpha\theta} = -\text{div}(\mathbf{A}_{\alpha\theta}(x)\nabla)$. In [120], such kernels were used as the weighting functions for the construction of the patch operator:

$$(D(x)f)(\theta, t) = \int_X h_{\alpha\theta t}(x, x') f(x') dx'. \quad (23)$$

Similar to the spectral analysis-based intrinsic CNN, heat diffusion CNN is composed of sequentially stacked layers, i.e., the output of the previous layer is used as the input to the subsequent layer, and the convolution operation is replaced by a layer tailored for heat diffusion.

5.4. A Unified Spatial-Domain Geometric Deep Learning Architecture for Human Shape Analysis

The authors in [132] proposed a unified geometric CNN generalizing the CNN to non-Euclidean domains. Instead of using fixed handcrafted weight functions, parametric kernels with learnable parameters were proposed. Particularly, a Gaussian kernel with learnable parameters was used:

$$w_j(\mathbf{u}) = \exp\left(-\frac{1}{2}(\mathbf{u} - \boldsymbol{\mu}_j)^T \boldsymbol{\Sigma}_j^{-1}(\mathbf{u} - \boldsymbol{\mu}_j)\right) \quad (24)$$

where $\boldsymbol{\Sigma}_j$ and $\boldsymbol{\mu}_j$ are learnable $d \times d$ and a $d \times 1$ covariance matrix and mean vector. Various non-Euclidean CNN methods previously proposed in the literature can be considered as particular instances of the proposed framework.

5.5. Geometric Structures over Deep Learning for Human Action Recognition

There are also studies on learning geometric structures over CNN. For example, the authors in [133] proposed a deep discriminative structured model, namely convolutional neural random fields (CNRFs), and applied it to the action recognition problem. In the proposed model, a spatiotemporal CNN was developed for feature learning from input image frames, and the CNN was combined with conditional random fields (CRFs) for capturing the interdependencies between outputs. The parameters from both CRF and CNN were learned in a joint fashion, which enabled structured prediction and feature learning.

6. Generalized Geometrics for Human-Related Analysis

General information denoting spatial or temporal distributions and attribute occurrences can be considered as geometric information in a generalized perspective. They are also useful information for boosting HRA. They are named generalized geometrics and are further classified into three sub-categories and introduced in this section.

6.1. Spatial Geometrics for Human Pose-Related Analysis and Human Action-Related Analysis

In HPA, spatial geometrics can be encoded as local structural features. The authors in [134] proposed a local joint structure as a complement for global features of individual body part locations and combined the two features for posture description. Local joint structures, specifically the triangle area of the three consecutively adjacent joints, were computed. It is a complement for body part locations in the sense that body part locations denote a single body part, while the proposed joint structure contains relative joint positions. Then, classification was performed with a combined method of dynamic time warping, Fourier temporal pyramid representation, and linear SVM.

Some works directly explored the neighboring properties. For example, the authors in [135] proposed a geometric correspondence feature named the Trisarea feature. It describes neighboring

properties between human body joints and is defined as the area of the triangle formed by three joints. This feature is utilized to identify human poses, of which variations over time capture the characteristics of human action.

Furthermore, in HAR, spatial geometrics can be encoded in features. For example, relative positions, distances between body joints, etc., are effective spatial geometric features. In [136], features from 3D skeleton data were processed separately by LSTM and CNN to conduct effective recognition with later fusion. Spatial features such as relative position, the distance between joints, and distances between joints and lines were explored, while temporal features such as the joint distances map and the joint trajectories map were studied. Spatial features were fed into LSTM, and temporal features were fed into CNN for recognizing actions.

Another way of encoding spatial geometrics is through modeling the co-occurrence of actions. The co-occurrence of actions was modeled in a probabilistic way without supervision in [137]. Videos containing human actions are considered as a sequence of short-term action clips (action words), and an activity is considered as a set of action topics indicating which actions are present in the video. A probabilistic model relating the action words and the action topics was proposed. It modeled long-range action relations that exist in the complex activity. The model was applied to unsupervised action segmentation and recognition and to detect forgotten actions, namely action patching.

6.2. Temporal Geometrics for Human Action Recognition

Temporal geometrics can be encoded by directly modeling the dynamics in a geometric space. For example, the authors in [138] proposed a second-order stochastic dynamical model in the state space (a Riemannian manifold) of articulated objects and derived equations of a Riemannian extended Kalman filter to perform the structure estimation from an image sequence captured by a camera from one perspective. The proposed model was proven by the authors to be locally weakly observable.

Furthermore, motion dynamics can be described in the original feature space. One widely-used measure is through scene flow. Scene flow describes the motion of 3D objects in the real world and implicitly describes the geometry of the 3D objects in motion. Scene flow can be considered as an optical flow fused from multiple cameras. The authors in [27] proposed the extraction and use of scene flow for action recognition from RGB-D data.

6.3. Spatial-Temporal Geometrics for Action Segmentation and Action Recognition

Action segmentation algorithms mine temporal segments containing actions from untrimmed videos. By incorporating a spatial component that represents the relationships between objects and a temporal component to capture object relationships across time, the method in [139] achieved better performances.

For action recognition problems, spatiotemporal information can be extracted through feature extraction and network extraction. The authors in [140] presented SkeletonNet, a deep learning framework for skeleton-based 3D action recognition. Cosine distance (CD) and normalized magnitude (NM) features were proposed and extracted from each frame of the skeleton sequence. Instead of treating the features of all frames as a time series, the authors fed extracted features to the proposed deep learning network, which contained two streams, one to extract the general features from the CD feature, while the other processed the NM feature. Outputs from the two streams were concatenated and processed by a fully-convolutional layer and then classified.

Furthermore, spatiotemporal information can be considered by modifying the deep learning network structure. For example, the authors in [141] proposed a differential gating scheme for a long short-term memory (LSTM) neural network and incorporated the spatial dynamics in action motions. The information gain was achieved by the derivative of states (DoS). The LSTM neural network utilizes three types of gating schemes for learning representations from long input sequences. The proposed method considered spatial information by incorporating DoS from the previous state into the input and forget gate and DoS from the current state into the output gate (as shown in Figure 18). Another

example is [142], in which the authors extended the RNN-based methods from temporal domains to spatiotemporal domains and applied them to analyze action-related information within the input data.

The spatiotemporal information can also be learned from channels other than RGB data. The authors in [26] combined spatiotemporal geometric features from depth images and joint positions to solve human action recognition problems. The method learned spatiotemporal features by constructing a 3D-based deep CNN (3D²CNN) for depth sequences. Depth images and joint positions were processed separately and fused in a later stage. Furthermore, spatiotemporal discrimination can be utilized to recognize human actions at different speeds. For example, the authors in [143] achieved this through considering spatiotemporal discrimination and action speed variations.

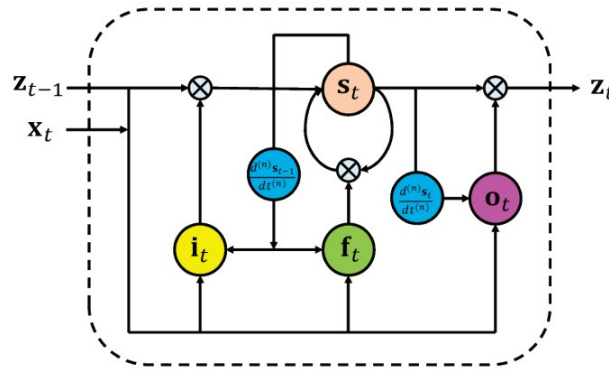


Figure 18. Architecture of the proposed dRNN model. In the memory cell, the input gate i_t and the forget gate f_t are controlled by the derivative of states (DoS) $\frac{d^{(n)}s_{t-1}}{dt^{(n)}}$ at $t - 1$, and the output gate o_t is controlled by the DoS $\frac{d^{(n)}s_t}{dt^{(n)}}$ at t [141]. ©2015 IEEE. Reprinted, with permission, from V. Veeriah, N. Zhuang, and G. Qi, Differential Recurrent Neural Networks for Action Recognition, in *Proceedings of the IEEE International Conference on Computer Vision*, Región Metropolitana, Chile, 11–18 December 2015, 4041–4049.

7. Validation Datasets

Publicly-available datasets for validating HRA are collected and categorized according to data types and applications. The datasets are classified based on their data type and their targeted applications: 3D human datasets, composed of 3D human models mainly for human shape analysis; 3D human action datasets with 3D data for human action analysis; RGB-D people datasets, composed of RGB-D data for people detection and people tracking; RGB-D human pose datasets for human pose analysis; and RGB-D human action datasets with RGB-D data for human action analysis.

7.1. 3D Human Datasets

In this section, public datasets on 3D humans are collected. These datasets are utilized to validate applications such as shape analysis, including deformable shape matching and shape retrieval. 3D human data with noise and partial 3D human model analysis are also considered.

7.1.1. KIDS Dataset

This dataset (<https://vision.in.tum.de/data/datasets/kids>) consists of two shape classes (“kid” and “fat kid”, as shown in Figure 19) under different poses, where the same poses are applied to both classes. The 3D shapes undergo nearly isometric and within-class deformations. All shapes in the dataset are given in OFF format and have around 60k vertices and consistent triangulations.



Figure 19. Examples from the Kidsdataset. The figure is originally from [94].

7.1.2. ShapeNet

ShapeNet is a well-maintained, large-scale dataset of 3D shapes. ShapeNet is composed of several subsets:

- (1) ShapeNetCore [144], including 55 common object categories (approximately 51,300 unique 3D models), 12 object categories of PASCAL 3D+, and a popular computer vision 3D benchmark dataset.
- (2) ShapeNetSem [145], including 12,000 models of 270 categories and annotated with manually-verified category labels, consistent alignments, real-world dimensions, estimates of their material composition at the category level, and estimates of their total volume and weight.

In ShapeNet, there are 35 subcategories (such as “adult, grownup”, “worker”, “child, baby”, etc.) and 2561 human-related models, namely “person, individual, someone, somebody, mortal, soul” in the “natural object” category.

7.1.3. TOSCA High-Resolution Dataset

The TOSCA [146] dataset (http://tosca.cs.technion.ac.il/book/resources_data.html) includes high-resolution 3D nonrigid shapes in variant poses. The dataset contains 80 object categories, including 11 cats, 9 dogs, 3 wolves, 8 horses, 6 centaurs, 4 gorillas, 12 females, and 2 males. Typically, the model has approximately 50,000 vertices.

7.1.4. Human 3.6M

This dataset (<http://vision.imar.ro/human3.6m/description.php>) is composed of 3.6 million 3D human poses with corresponding images. The dataset contains 11 professional actors, including 6 males and 5 females, and 17 scenarios, including “discussion”, “smoking”, “taking a photo”, “talking on the phone”, etc.

The dataset is composed of high-resolution 50-Hz videos from four calibrated cameras. The dataset has rich annotations, including accurate 3D joint positions, joint angles from a high-speed motion capture system, where 24 pixel-level body part labels for each configuration are given. It also provides accurate background subtraction and person bounding boxes.

Furthermore, the Human 3.6M dataset provides precomputed image descriptors, software for visualization and discriminative human pose prediction, and performance evaluation on a withheld test set [147,148].

7.1.5. H3D Database

H3D (Humans in 3D) (<https://www2.eecs.berkeley.edu/Research/Projects/CS/vision/shape/h3d/>) is a 3D human dataset with annotations. The annotations include the joints, other keypoints (“eyes”, “ears”, “nose”, “shoulders”, “elbows”, “wrists”, “hips”, “knees”, and “ankles”), and 3D poses inferred from the keypoints with a visibility Boolean for each keypoint. The dataset is also

annotated with regions (“upper clothes”, “lower clothes”, “dress”, “socks”, “shoes”, “hands”, “gloves”, “neck”, “face”, “hair”, “hat”, “sunglasses”, “bag”, “occluder”), and body types (“male”, “female”, or “child”). For detailed descriptions, please refer to [149]. Figure 20 shows an example of the data and their annotations.

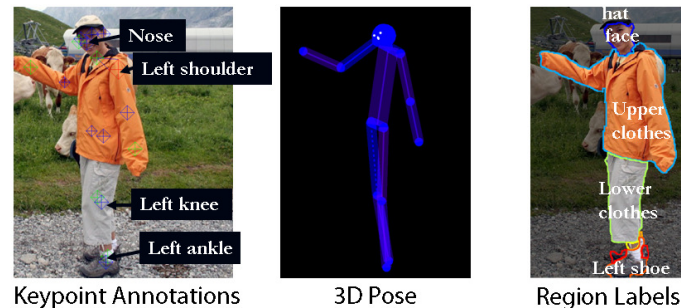


Figure 20. Contents from the H3D dataset. The figure is originally from the website [149].

7.1.6. 3D Shape Dataset with Noise

This dataset (<https://vision.in.tum.de/data/datasets/topkids>) consists of a collection of 3D shapes under deformations including topological changes [20]. The dataset has the ground-truth matching the null shape for all shapes, but not all vertices have a match due to topological changes.

7.1.7. Partial Shape Dataset

The Partial Shape Dataset (<https://vision.in.tum.de/data/datasets/partial>) includes two datasets, one is the cuts dataset with 456 partial shapes and the other is the holes dataset with 684 partial shapes. These two datasets exemplify different kinds of partiality: The cuts dataset contains shapes with a single cut; The holes dataset contains irregular holes and multiple cuts. Examples from the dataset are shown in Figure 21. The datasets provided can be used for deformable 3D shape matching and retrieval under partiality transformations [18].



Figure 21. Examples from the Partial Shape Dataset. The figure is originally from the website [18].

7.1.8. SHREC

The 3D Shape Retrieval Contest (<http://www.shrec.net>) evaluates the effectiveness of 3D shape retrieval algorithms. SHREC’18 is the tenth edition of the contest (<https://3dor2018.sites.uu.nl>). The contest contains tracks with various goals. Many tracks are related to the scope of this paper; for example, shape retrieval from 3D human shapes represented by triangular meshes (<https://vision.in.tum.de/~laehner/shrec2016/>), human shape retrieval from depth sensor data (<http://www.andreagiachetti.it/shrec16/>), and partial shape matching (http://tosca.cs.technion.ac.il/book/resources_data.html) based on the TOSCA high-resolution dataset [146] (http://tosca.cs.technion.ac.il/book/resources_data.html).

7.2. 3D Human Action Datasets

7.2.1. CMU Graphics Lab Motion Capture Database

The dataset (<http://mocap.cs.cmu.edu/>) contains 2605 motion capture trials of six categories, including “human interaction”, “interaction with environment”, “locomotion”, “physical activities & sports”, “situations & scenarios”, and “test motions”, and 23 subcategories, including “running”, “walking”, “jumping”, etc.

7.2.2. HumanEva Dataset

The HumanEva-I and HumanEva-II (<http://humaneva.is.tue.mpg.de/>) datasets were obtained from a motion capture system. The HumanEva-I dataset contains seven calibrated video sequences (four grayscale and three color) with synchronized 3D body poses. The dataset has 4 subjects of 6 actions, including “walking”, “jogging”, “gesturing”, etc. The dataset is split into training, validation, and testing sets. Also, the error measurements of the 2D and 3D poses are provided.

7.3. RGB-D People Datasets

7.3.1. RGB-D People Datasets

The RGB-D People Datasets (<http://www2.informatik.uni-freiburg.de/~spinello/RGBD-dataset.html>) contain people in RGB-D Kinect data with annotations. This datasets are composed of more than 3000 RGB-D frames. In the datasets, mostly are upright walking and standing persons. The persons are under different occlusion conditions. This dataset has been re-annotated in [150,151]. Examples from the dataset are shown in Figure 22.

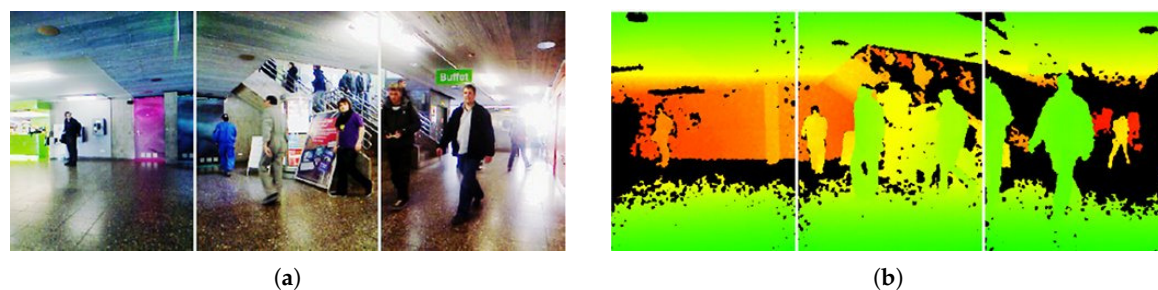


Figure 22. Examples from the RGB-D People Dataset. The figure shows the color image data (a) and the dense depth data (b) of three exemplar frames. The figure is originally from the website [150,151].

7.3.2. RGB-D Human Tracking Dataset

There are five validation videos with ground-truths, and 95 evaluation videos in the RGB-D Human Tracking Dataset (<http://tracking.cs.princeton.edu/dataset.html>). Captured by Kinect v1, each sequence has its RGB images and depth images. Captured videos contain moving objects such as humans, balls, and cars and are labeled with per-frame bounding boxes covering only the target object. The authors in [152] presented a quantitative comparison of various algorithms on this dataset. Examples and annotations from the dataset are shown in Figure 23.



Figure 23. Contents from the RGB-D Human Tracking Dataset. The figure is originally from the RGB-D Human Tracking Dataset website [152].

7.4. RGB-D Human Pose and Posture Datasets

Kinect Gesture Dataset

The Microsoft Research Cambridge-12 Kinect gesture dataset (<https://www.microsoft.com/en-us/download/details.aspx?id=52283&from=http%3A%2F%2Fresearch.microsoft.com%2Fen-us%2Fum%2Fcambridge%2Fprojects%2Fmsrc12%2F>) is composed of sequences of human movements. Human gestures are denoted by body part locations. The dataset contains 594 sequences and 719,359 frames performed by 30 people with 12 gestures. There are 6244 gesture instances in total in the dataset.

7.5. RGB-D Human Action and Activity Datasets

7.5.1. Human Daily Activity Dataset

The authors in [137] collected an RGB-D activity video dataset recorded by the Kinect v2, containing human daily activities composed of multiple actions interacting with different objects.

7.5.2. Cornell Activity Datasets

Cornell Activity Datasets CAD-60 and CAD-120 are two RGB-D human activity datasets (<http://pr.cs.cornell.edu/humanactivities/data.php>) containing skeleton and RGB-D data. RGB-D data have a resolution of 240×320 , of which the RGB data are saved as three-channel 8-bit PNG files, and the depth data are saved as single-channel 16-bit PNG files.

The CAD-60 dataset contains 60 RGB-D videos, performed by 4 subjects, including 2 males, 2 females, and 1 left-handed person, in 5 different environments, including “office”, “kitchen”, “bedroom”, “bathroom”, and “living room”, and of 12 activities, including “rinsing mouth”, “brushing teeth”, “wearing contact lens”, etc.

The CAD-120 dataset contains 120 RGB-D videos of long daily activities, 4 subjects (same as CAD-60), 10 high-level activities (“making cereal”, “taking medicine”, “stacking objects”, etc.), 10 sub-activity labels (“reaching”, “moving”, “pouring”, etc.), and 12 object affordance labels (“reachable”, “movable”, “pourable”, etc.).

7.5.3. 50 Salads Dataset

The dataset (<http://cvip.computing.dundee.ac.uk/datasets/foodpreparation/50salads/>) captures 25 people preparing two mixed salads each and contains over four hours of the annotated accelerometer and RGB-D video data. The RGB video data have a resolution of 640×480 pixels at 30 Hz and the depth maps a resolution of 640×480 pixels at 30 Hz, and the three-axis accelerometer data are at 50 Hz [153].

7.5.4. UR Fall Detection Dataset

This dataset (<http://fenix.univ.rzeszow.pl/~mkepski/ds/uf.html>) contains 70 (30 falls + 40 activities of daily living) sequences [154]. Fall events were recorded with two Microsoft Kinect cameras and corresponding accelerometric data. Examples from the dataset are shown in Figure 24.



Figure 24. Sample Images from the URfall dataset. The figure was captured from the demo video on the UR Fall Dataset website [154].

7.5.5. Tum Kitchen Dataset

The TUM Kitchen Dataset contains several subjects sitting by a table. Some perform activities simulating a robot, transporting items one-by-one; while others behave more human-like and grasp as many objects as they can in one performance. And for each subject performing reaching and grasping, there are two trials.

8. Performances of Related Works

The performances of various geometric methods for HRA studied in this review are compared in terms of estimation accuracy or estimation error and shown in Table 1. The methods are categorized based on their applications, i.e., HSA, HPA, or HAA. Due to the characteristics of the specific application, the number of methods in each category varies. For example, for HRA and HPA, many algorithms measure their performance by quality, while for HAA, many algorithms validate their performance based on recognition accuracy. In each category, the methods are listed in the chronological order of publication, and then in alphabetical order by the method names. For each validation dataset, the best recognition accuracy (in percent) or the minimum estimation error (in centimeter) among all experiment settings are listed.

From the table, we can see that the average precision of HRA and mostly HAA is quite high. Except for some difficult datasets (i.e., “PASCAL VOC 2011”, “PASCAL VOC 2012”, “ChaLearn LAP IsoGD”, “50 Salads”, and “JIGSAWS”), the recognition accuracy was above 80% for all validation datasets. The “Enhanced-LSTM-based method” and “Gram matrix-based method” achieved 100% accuracy on three of the validation datasets.

For HSA, the paper reviews the related works on human shape correspondence, human model symmetry analysis, and human shape recognition. For human shape correspondence, the authors in [83] represented a deformation field as a linear operator on real-valued functions on the shape and gave the state-of-the-art performance on human shape correspondence. An exemplary result is illustrated in Figure 25. Another exemplary work is from [95], and its visualized results are illustrated in Figure 12. Quantitative measurement of the method from [126] is shown in the human shape analysis section in Table 1.

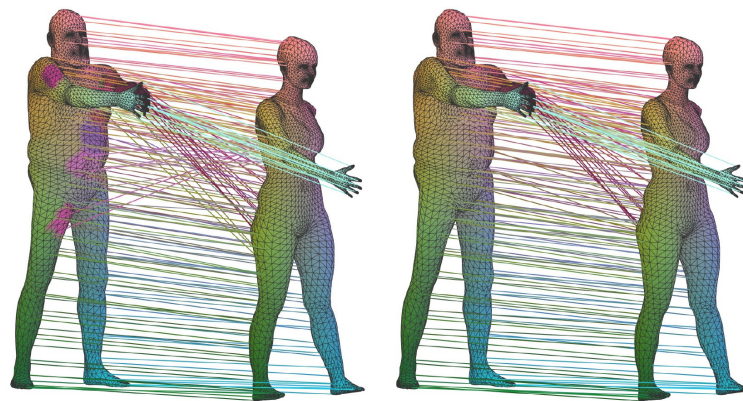


Figure 25. The figure shows exemplary point-to-point maps from one human body model to another. The overall performance of the proposed geometric method (right) is working better than the compared SHOT(left) method on the entire shape [83]. Republished with permission of ACM, from ACM Trans. Graph., E. Corman and M. Ovsjanikov, Vol. 38, 2019; permission conveyed through Copyright Clearance Center, Inc.

For human model symmetry analysis, the authors in [84] proposed a numerical framework for the analysis, addressing the problems of full and partial exact and approximate symmetry detection and classification. The exemplary results are illustrated in Figure 26. Note that the increase in regularity results in the shortening of the boundary at the expense of the symmetry of the part.

For human body shape recognition, a state-of-the art method [86] was proposed based on a geodesic distance matrix. A recognition rate of 100% was obtained on the TOSCA database. Some exemplary results are illustrated in Figure 27.

For HPA, the review studies human pose space modeling and human pose estimation. The method of using a pose-space subspace method [102] gives a good performance on modeling the human pose space. The proposed method uses secondary soft-tissue finite element method (FEM) dynamics computed under arbitrary rigged or skeletal motion. Experiment comparisons are illustrated in Figure 28. The performances of the human pose estimation method using geometric methods are shown in Table 1.

Table 1. Evaluation comparisons of geometric methods.

Applications	Year	Methods	Validation Datasets	Accuracy (%) or Error (cm)
Human Shape Analysis	2016	Dense correspondence-based method [126]	FAUST	2–2.35 cm
Human Pose Related Analysis	2017	SkeletonNet [140]	CMUMocap	89.46%
			NTURGB+D	81.16%
			SBUKinect interaction	93.47%
Human action related Analysis	2011	Spatio temporal manifold model-based method [111]	Mocap	90.00%
	2012	Bi-lingual Hangelets [45]	IXMAS	90.57%
	2012	Graph matching-based method [105]	KTH	89.3%
	2013	Directed acyclic graph kernel-based method [107]	UCFSport	85.2%
	2014	Fully-convolutional network-based method [113]	PASCAL VOC 2011	62.7%
	2014	Lie group-based method [109]	MSRAction3D	92.46%
			UTKinect-Action	97.08%
			Florence3D-Action	90.88%
	2014	Shape matching-based method [94]	TOSCA	90.00%
	2015	Deep deconvolution network-based method [114]	PASCAL VOC 2012	72.5%
	2015	Differential recurrent neural network-based method [141]	KTH-1	93.96%
			KTH-2	92.12%
			MSR Action3D	92.03%
	2016	3D DCNN-based method [26]	MSR Action3D	98.14%
	2016	Convolutional neural random fields [133]	Weizmann	98.88%
			Youtube	94.4%
			UCF50	86.5%
			WBJR	95.70%
	2016	Enhanced-LSTM-based method [142]	NTURGB+D	81.60%
			SBUInteraction	93.3%
			UT-Kinect	95.00%
			Berkeley MHAD	100.00%
			MSRAction3D	94.80%
	2016	Gram matrix-based method [104]	HDM05	88.0 ± 6.3%
			MSR-Action3D	96.97%
			MHAD	100%
			UTKinect	100%
	2016	Local joint structure and body part locations Feature-based method [134]	MSR Action3D	85.86%
			UTKinect-Action	96.49%
			Florence3D-Action	87.47%
	2016	Motionlet-graph-based method [108]	MSR Action3D	94.77%
			Florence 3D Actions	91.63%
			UTKinect Action	97.44%
	2016	Relative 3D geometry-based method [103]	Florence3D	92.16%
			G3D	92.12%
			MSR Action3D	90.69%
			MSRPairs	94.33%
			UTKinect-Action	97.20%
	2016	Rolling map-based method [110]	Florence3D	91.40%
			MSRPairs	94.67%
			G3D	90.94%
	2016	Segmental spatiotemporal CNN-based method [139]	50 Salads	72.00%
			JIGSAWS	74.22%
	2017	LSTM and CNN-based method [136]	NTU RGB+D	87.40%
	2017	Geometric feature pooling-based method [136]	HOIactivity dataset	89.6%
	2017	Scene flow to action map [27]	ChaLearn LAP IsoGD	36.27%
			Multi-modal and multi-view and interactive dataset	91.2%
	2017	Spatiotemporal feature-based method [143]	MSRAction3D	93.81%
			UTKinect-Action	97.47%

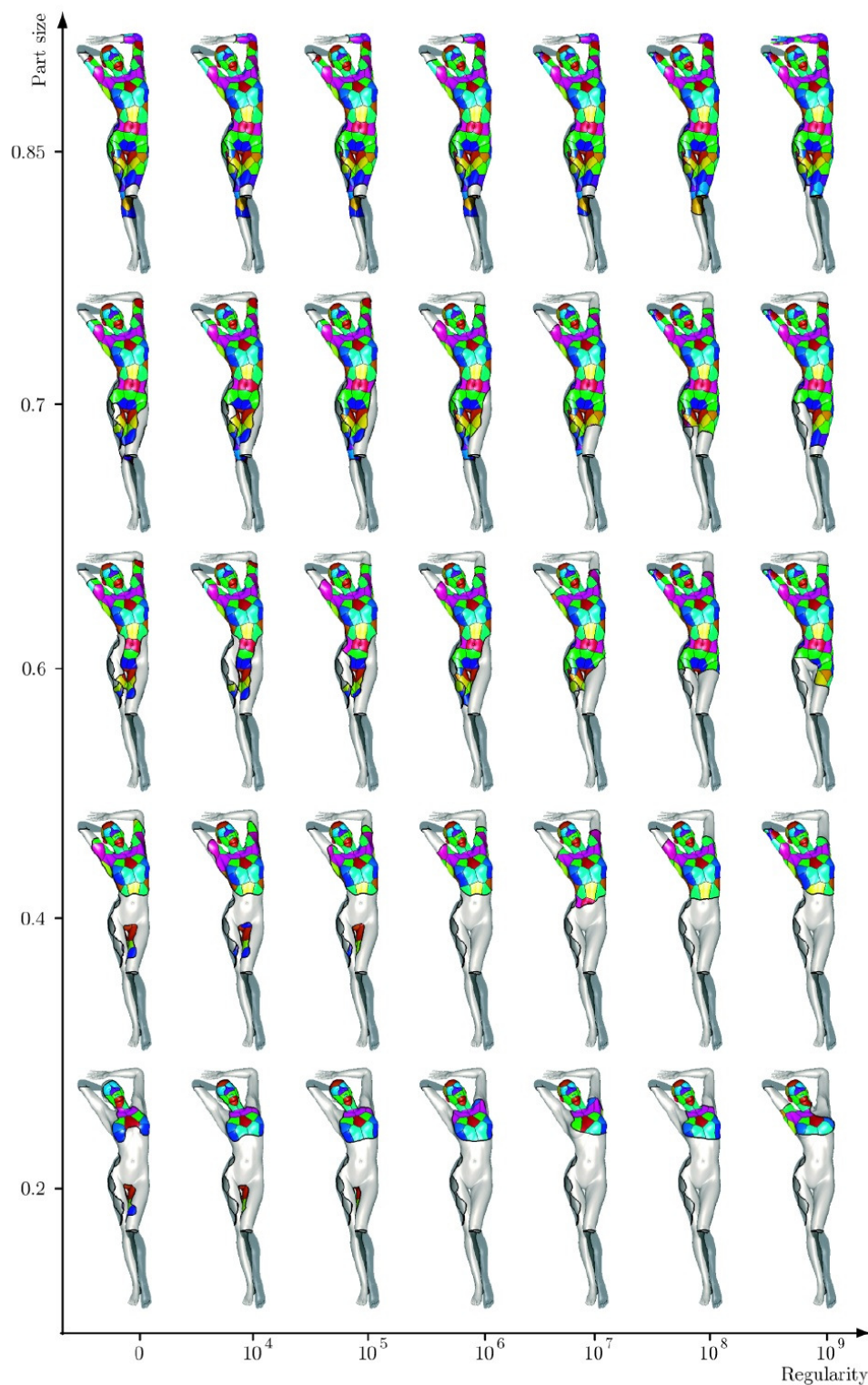


Figure 26. The figure shows exemplary results on partial symmetries of human body models. The partial human body models are obtained by removing certain body parts, and the removed body parts are marked in semitransparent dark gray. The experiments are carried out under various regularization coefficients (the horizontal axis) and various body part sizes (the vertical axis). Symmetric body parts are marked with the same color. Discarded body parts are marked in light gray [84]. Reprinted by permission from SPRINGER NATURE: Springer Nature, Int. J. Comput. Vis., Full and Partial Symmetries of Non-rigid Shapes, D. Raviv, A. M. Bronstein, M. M. Bronstein, R. Kimmel, Copyright 2010.

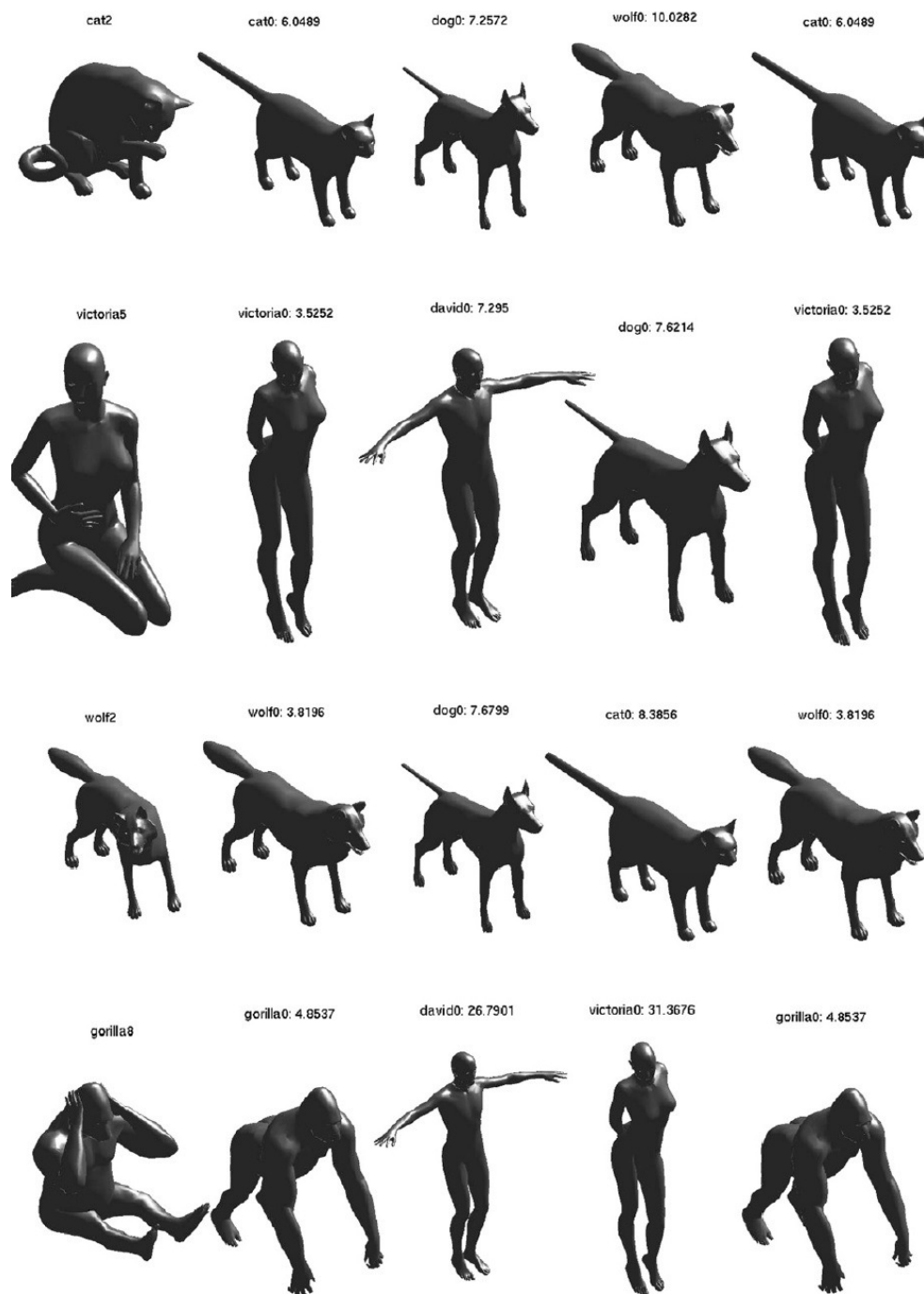


Figure 27. The figure shows exemplary shape recognition results. The first column denotes the query shape, and the second to the fourth columns show the three closest matches [86]. Reprinted from Pattern Recognition, Vol. 45, D. Smeets, J. Hermans, D. Vandermeulen, P. Suetens, Isometric Deformation Invariant 3D Shape Recognition, 2817–2831, Copyright 2012, with permission from Elsevier.

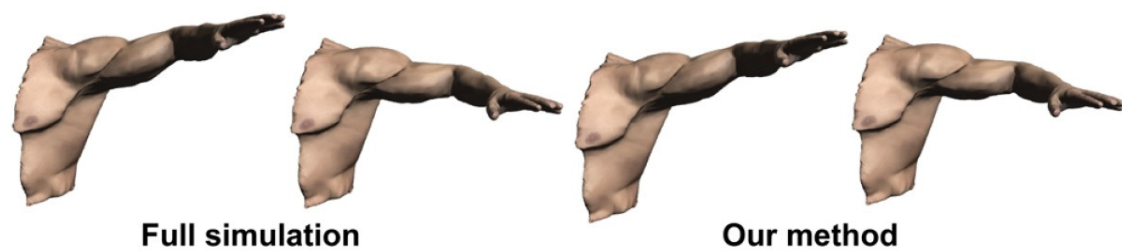


Figure 28. The proposed method produces a good approximation to the full simulation while being 60-times faster. The figure is originally from [102]. Reproduced with permission from Jernej Barbic, ACM Transactions on Graphics; published by ACM Digital Library, 2016.

9. Conclusions and Discussions

This review presented a comprehensive study on human-related analysis (HRA), including human shape analysis, human pose-related analysis, and human action-related analysis. It first introduced fundamental concepts in the topology and manifold as fundamental knowledge for geometric modeling with these theories. Then, geometric methods using these theories were introduced. Later, geometric methods applied for HRA were studied. Considering the great impact of deep learning and its potential in feature extraction and feature representation, the review also considered geometric deep learning, which has recently been a popular topic. Then, generalized geometric methods, which study general purpose geometric information for HRA, were explored. Validation datasets for verifying geometric HRA methods were collected, and the performances of various methods were collected, compared, and shown in a table.

For further research, one topic worth exploring is defining intrinsic deep learning algorithms on RGB-D data, specifically defining the convolution, the pooling, and the spatial shift operation on the RGB-D domain. There are very few works on this topic despite its wide applications. Another research topic worth exploring is learning geometric information and utilizing it as priors. 3D data are still comparatively more difficult to acquire than images or videos; thus, it would be helpful to utilize geometric information as priors for improving task performances, which use images or videos as inputs.

Author Contributions: Conceptualization, W.G. and Y.Q.; related works collection, C.W., H.Y., C.L., and L.X.; paper structure, W.Z. and F.G.; writing—original draft preparation, W.G. and B.Z.; writing—review and editing, C.W., H.Y., and C.L.; funding acquisition, Y.Q.

Funding: This work was supported in part by Shenzhen Basic Research Program (JCYJ20170818164704758).

Acknowledgments: We are extremely grateful for the reviewers who helped to improve the quality of the paper and the editors who helped us going through the process. This work is the result of a pleasant collaboration. It started with a three-month research stay of the first author in Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences (SIAT), where she got the idea and had fruitful discussions with Yali Wang and Tianqi Fan.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

Set Theory Symbols

\sim	A relation
$[x]$	The equivalence class x
V/\sim_W	The quotient space of V by W , also denoted as V/W

Topological Symbols

Ω	A topological space
G	A topological group
GL	The general linear group

Manifold Symbols

\mathcal{M}	A manifold
$T_x\mathcal{M}$	The tangent space of \mathcal{M} at x
$T\mathcal{M}$	The tangent bundle of \mathcal{M}
$T^*\mathcal{M}$	The cotangent bundle
$\Gamma(\mathcal{F})$	The section of the vector bundle \mathcal{F}
Exp	The exponential map
Log	The inverse exponential map, also denoted as Exp^{-1}
∇	A connection
∇^*	A dual connection
γ	A geodesic, i.e., a curve $\gamma : I \rightarrow \mathcal{M}$ such that $\nabla_{\gamma'}\gamma' = 0$.
ρ_g	$\mathcal{M} \times \mathcal{M} \rightarrow \mathbb{R}$, the geodesic distance function determined by g ,
(\mathcal{M}, g)	A Riemannian manifold equipped with a metric g
$(\mathcal{M}, \mathfrak{M})$	A smooth manifold of a pair of a topological manifolds \mathcal{M} and an atlas \mathfrak{M} on \mathcal{M}

Appendix A. Mathematical Concepts

Expanded geometric concepts are introduced in the Appendix. The concepts are categorized following the mathematical branch to which they belong. Appendix A.1 introduces expanded concepts from the set theory. Appendix A.2 explains expanded concepts from topology. Appendix A.3 introduces expanded concepts from algebraic topology, and Appendix A.4 introduces extended manifold concepts.

*Appendix A.1. Set Theory Concepts***Appendix A.1.1. Equivalence Relations**

A (binary) relation on a set X is a subset of $X \times X$. We often denote relations by \sim and write $x \sim x'$ to indicate that (x, x') is in the relation.

A relation \sim on a set X is an equivalence relation if the following three conditions hold:

- $x \sim x$ for all $x \in X$.
- $x \sim x'$ if and only if $x' \sim x$.
- $x \sim x'$ and $x' \sim x''$ implies $x \sim x''$.

Appendix A.1.2. Equivalence Class

The ways to denote equivalence classes are listed as follows:

- The equivalence class of $x \in X$, denoted by $[x]$, means the set $\{x' | x \sim x'\}$. The sets $[x]$ for all $x \in X$ form a partition of the set X .
- The set of equivalence classes under \sim can be denoted as X/\sim , and it is referred to as the quotient of X with respect to \sim .

Appendix A.1.3. Covering

A family $A_i, i \in I$ of subsets of X is a covering of X if $X = \bigcup_{i \in I} A_i$.

*Appendix A.2. Topological Concepts***Appendix A.2.1. Closed Sets/Interior and Closure of A Set/Limit Points**

A subset A of a topological space X is said to be closed if the set $X - A$ is open.

Given a subset A of a topological space X , the interior of A is defined as the union of all open sets contained in A , and the closure of A is defined as the intersection of all closed sets containing A .

If A is a subset of the topological space X and if x is a point of X , we say that x is a limit point (or “cluster point,” or “point of accumulation”) of A if every neighborhood of x intersects A at some point

other than x or x is a limit point of A if it belongs to the closure of $A - [x]$. The point x may or may not lie in A .

Appendix A.2.2. Continuous Function

Let X and Y be topological spaces. A function $f : X \rightarrow Y$ is continuous if for each point x of X and each neighborhood N of $f(x)$ in Y , the set $f^{-1}(N)$ is a neighborhood of x in X .

Formulated with the concept of openness, a function from X to Y is continuous if and only if the inverse image of each open set of Y is open in X .

Appendix A.2.3. Quotient Map

Let X and Y be topological spaces and $p : X \rightarrow Y$ be a surjective map. The map p is said to be a quotient map provided a subset U of Y that is open in Y if and only if $p^{-1}(U)$ is open in X .

Appendix A.2.4. Metric

A metric on a set X is a function $d : X \times X \rightarrow R$ having the following properties:

- $d(x, y) \geq 0$ for all $x, y \in X$, equality holds if and only if $x = y$.
- $d(x, y) = d(y, x)$ for all $x, y \in X$.
- (Triangle inequality) $d(x, y) + d(y, z) \leq d(x, z)$, for all $x, y, z \in X$.

A topology can be imposed on a set by defining a metric on a topology. Given a metric d on X , $d(x, y)$ is often called the distance between x and y . Given $t > 0$, consider the set $B_d(x, \epsilon) = \{y | d(x, y) < \epsilon\}$ of all points y whose distance from x is less than ϵ . It is called the ϵ -ball centered at x .

Appendix A.2.5. Hausdorff Space

For every pair of points $p, q \in M$ in a Hausdorff space, there are disjoint open subsets $U, V \subset M$ such that $p \in U$ and $q \in V$.

Appendix A.3. Algebraic Topology Concepts

Appendix A.3.1. Orbit Space

Let X be a space and G be a topological group; an action of G on X is a continuous map $a : G \times X \rightarrow X$ such that (denoting $a(g \times x)$ by $g \cdot x$):

- $e \cdot x = x$ for all $x \in X$.
- $g_1 \cdot (g_2 \cdot x) = (g_1 \cdot g_2) \cdot x$ for all $x \in X$ and $g_1, g_2 \in G$.

Define $x \sim g \cdot x$ for all x and g , and the resulting quotient space is denoted X/G and called the orbit space of the action a .

Appendix A.3.2. Homotopy

A homotopy between two maps $f : X \rightarrow Y$ and $g : X \rightarrow Y$ is a map $H : X \times I \rightarrow Y$ such that $H(x, 0) = f(x)$ and $H(x, 1) = g(x)$ for all $x \in X$. We then say that f and g are homotopic and write $f \simeq g$, or $H : f \simeq g$ or $f \stackrel{H}{\simeq} g$.

A special case is when f is a path in X . If $f : [0, 1] \rightarrow X$ is a continuous map such that $f(0) = x_0$ and $f(1) = x_1$, we say that f is a path in X from x_0 to x_1 . We also say that x_0 is the initial point, and x_1 the final point of the path f .

Appendix A.3.3. Fundamental Group

Let X be a space and x_0 be a point of X . A path in X that begins and ends at x_0 is called a loop based at x_0 . The set of path homotopy classes of loops based at x_0 with the operation $*$ is called the

fundamental group of X relative to the base point x_0 . It is denoted by $\pi_1(X, x_0)$. Two spaces that are homeomorphic have fundamental groups that are isomorphic.

Appendix A.3.4. Homology

Although homotopy is easy to define and conceptually very attractive, homotopy groups are very difficult to compute. Another invariant is homology, which, instead of being easy to define and difficult to compute, is difficult to define and easy to compute. The idea of counting occurrences of patterns directly is unworkable, but that of counting equivalence classes of occurrences of patterns under an equivalence relation is workable. In this section, basic notations for homology, i.e., the definition of homology, the normal subgroup, the Abelian Group, and the commutator subgroup, are introduced.

Appendix A.3.4.1. Normal Subgroup

Let G be a group and H be a subset of G . If H is a group under the operation of G , then we say that H is a subgroup of G .

A subgroup H of a group G is normal in G if and only if $gH = Hg$ for all g in G .

Appendix A.3.4.2. Abelian Group

A group is Abelian if $xy = yx$ for all group elements x and y .

Appendix A.3.4.3. Commutator Subgroup

Let G be a group. If $x, y \in G$, we denote the element $[x, y] = xyx^{-1}y^{-1}$ by $[x, y]$; it is called the commutator of x and y . The subgroup of G generated by the set of all commutators in G is called the commutator subgroup of G and is denoted by $[G, G]$. The subgroup $[G, G]$ is a normal subgroup of G , and the quotient group $G/[G, G]$ is Abelian.

Appendix A.3.4.4. Homology

If X is a path-connected space, let $H_1(X) = \pi_1(X, x_0)/[\pi_1(X, x_0), \pi_1(X, x_0)]$. $H_1(X)$ is the first homology group of X . Groups $H_n(X)$ are called the homology groups of X that are defined for all $n \geq 0$.

Appendix A.4. Manifold Concepts

Appendix A.4.1. Atlas

A family $(U_\alpha, u_\alpha)_{\alpha \in A}$ of charts on \mathcal{M} such that the U_α form a cover of \mathcal{M} is called an atlas. An atlas \mathcal{A} is called a smooth atlas if any two charts in \mathcal{A} are smoothly compatible with each other.

Appendix A.4.2. Smooth Manifold

A smooth structure on a topological n -manifold \mathcal{M} is a maximal smooth atlas. A smooth manifold is a pair $(\mathcal{M}, \mathcal{A})$, where \mathcal{M} is a topological manifold, and \mathcal{A} is a smooth structure on \mathcal{M} . The smooth structure is usually omitted, and \mathcal{M} is called a smooth manifold. Smooth structures are also called differentiable structures or C^∞ structures. Note that a C^0 manifold is a topological manifold.

Appendix A.4.3. Section

If $\pi : \widetilde{\mathcal{M}} \rightarrow \mathcal{M}$ is any continuous map, a section of π is a continuous map $\sigma : \mathcal{M} \rightarrow \widetilde{\mathcal{M}}$ such that $\pi \circ \sigma = Id_{\mathcal{M}}$. A local section is a continuous map $\sigma : U \rightarrow \widetilde{\mathcal{M}}$ defined on some open set $U \subset \mathcal{M}$ and satisfying the analogous relation $\pi \circ \sigma = Id_U$.

Appendix A.4.4. Vector Bundle/Fiber Bundle

A mapping $f : \mathcal{E} \rightarrow \mathcal{M}$ between manifolds is called a submersion at $x \in \mathcal{E}$ if the mapping of $T_x f : T_x \mathcal{E} \rightarrow T_{f(x)} \mathcal{M}$ is surjective or the rank of the mapping equals $\dim \mathcal{M}$.

A triple $(\mathcal{E}, p, \mathcal{M})$ is called a fibered manifold if $p : \mathcal{E} \rightarrow \mathcal{M}$ is a surjective submersion.

Let $p : \mathcal{E} \rightarrow \mathcal{M}$ be a smooth mapping between manifolds; a vector bundle chart on $(\mathcal{E}, p, \mathcal{M})$ means a pair (U, ψ) , where U is an open subset in \mathcal{M} and ψ is a fiber-respecting diffeomorphism as in the following diagram:

$$\begin{array}{ccc} \mathcal{E}|_U & \xrightarrow{\psi} & U \times S \\ & \searrow p \quad \swarrow pr_1 & \\ & U & \end{array}$$

Here, S is a fixed finite-dimensional real (unless otherwise specified) vector space called the standard fiber or the typical fiber, p is a surjective submersion, and $\psi : p^{-1}(U) \rightarrow U \times S$ such that $pr_1 \circ \psi = p$.

Two vector bundle charts (U_1, ψ_1) and (U_2, ψ_2) are called compatible, if $\psi_1 \circ \psi_2^{-1}$ is a fiber linear isomorphism, i.e., $(\psi_1 \circ \psi_2^{-1})(x, v) = (x, \psi_{1,2}(x)v)$ for some mapping $\psi_{1,2} : U_{1,2} := U_1 \cap U_2 \rightarrow GL(V)$.

A vector bundle atlas $(U_\alpha, \psi_\alpha)_{\alpha \in A}$ for $(\mathcal{E}, p, \mathcal{M})$ is a set of pairwise compatible vector bundle charts (U_α, ψ_α) such that $(U_\alpha)_{\alpha \in A}$ is an open cover of \mathcal{M} .

A vector bundle $(\mathcal{E}, p, \mathcal{M})$ consists of a manifold \mathcal{E} (the total space), a manifold \mathcal{M} (the base), and a smooth mapping $p : \mathcal{E} \rightarrow \mathcal{M}$ (the projection) together with an equivalence class of vector bundle atlases. Figure A1 shows an exemplary vector bundle, where a Möbius strip is a line bundle over the one-sphere.

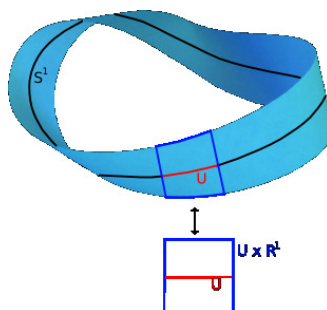


Figure A1. Vector bundle illustration (by Jakob.scholbach at English Wikipedia, CC BY-SA 3.0, <https://commons.wikimedia.org/w/index.php?curid=6082417>).

A fiber bundle $(\mathcal{E}, p, \mathcal{M}, S)$ consists of a manifold \mathcal{E} , a manifold \mathcal{M} , a vector space S , and a smooth mapping p satisfying the above conditions.

Appendix A.4.5. The Tangent Bundle Of A Vector Bundle

Let $(\mathcal{E}, p, \mathcal{M})$ be a vector bundle with fiber addition $+_{\mathcal{E}} : \mathcal{E} \times_{\mathcal{M}} \mathcal{E} \rightarrow \mathcal{E}$ and fiber scalar multiplication $m_t^{\mathcal{E}} : \mathcal{E} \rightarrow \mathcal{E}$. Then, the tangent bundle of the manifold \mathcal{E} , denoted by (TE, π_E, \mathcal{E}) , is itself a vector bundle with fiber addition denoted by $+_{TE}$ and scalar multiplication denoted by m_t^{TE} .

Appendix A.4.6. Vertical Bundle

Given two fibered manifolds $(\mathcal{M}, p, \mathcal{N})$ and $(\tilde{\mathcal{M}}, \tilde{p}, \tilde{\mathcal{N}})$, a morphism $(\mathcal{M}, p, \mathcal{N}) \rightarrow (\tilde{\mathcal{M}}, \tilde{p}, \tilde{\mathcal{N}})$ means a smooth map $f : \mathcal{M} \rightarrow \tilde{\mathcal{M}}$ transforming each fiber of \mathcal{M} into a fiber of $\tilde{\mathcal{M}}$.

Two maps $f, g : \mathcal{M} \rightarrow \mathcal{N}$ are said to determine the same r -jet (jets or holonomic jets) at $x \in M$, if for every curve $\gamma : \mathbb{R} \rightarrow M$ with $\gamma(0) = x$, the curves $f \circ \gamma$ and $g \circ \gamma$ have the r^{th} order contact at zero. In such a case, we write $j_x^r f = j_x^r g$ or $j^r f(x) = j^r g(x)$.

The elements of the manifold $T_k^r M := J_0^r(\mathbb{R}^k, \mathcal{M})$ are said to be the k -dimensional velocities of order r on \mathcal{M} , or (k, r) -velocities.

The projection $\pi_{r-1}^r : T^r M \rightarrow T^{r-1} M$ is a linear morphism of vector bundles. Its kernel is described by the following exact sequence of vector bundles over \mathcal{M} :

$$0 \rightarrow S^r T^* M \rightarrow T^r M \xrightarrow{\pi_{r-1}^r} T^{r-1} M \rightarrow 0$$

Let $(\mathcal{E}, p, \mathcal{M}, S)$ be a fiber bundle and $T_p : TE \rightarrow TM$ be the tangent mapping. Its kernel $\ker T_p =: VE$ is called the vertical bundle of \mathcal{E} . For example, if $(U_\alpha, \psi_\alpha : \mathcal{E}|U_\alpha \rightarrow U_\alpha \times V)_{\alpha \in A}$ is a vector bundle atlas for \mathcal{E} , such that (U_α, u_α) is also a manifold atlas for \mathcal{M} , then $(\mathcal{E}|U_\alpha, \psi'_\alpha)_\alpha \in A$ is an atlas for the manifold \mathcal{E} , where:

$$\psi'_\alpha := (u_\alpha \times \text{Id}_V) \circ \psi_\alpha : \mathcal{E}|U_\alpha \rightarrow U_\alpha \times V \rightarrow u_\alpha(U_\alpha) \times V \subset \mathbb{R}^m \times V.$$

Hence, the family $(T(\mathcal{E}|U_\alpha), T\psi'_\alpha : T(\mathcal{E}|U_\alpha) \rightarrow T(u_\alpha(U_\alpha) \times V) = u_\alpha(U_\alpha) \times V \times \mathbb{R}^m \times V)_{\alpha \in A}$ is the atlas describing the canonical vector bundle structure of (TE, π_E, \mathcal{E}) . The transition functions are in turn:

$$\begin{aligned} (\psi_\alpha \circ \psi_\beta^{-1})(x, v) &= (x, \psi_{\alpha\beta}(x)v) \text{ for } x \in U_{\alpha\beta} \\ (u_\alpha \circ u_\beta^{-1})(y) &= u_{\alpha\beta}(y) \text{ for } y \in u_\beta(U_{\alpha\beta}) \\ (\psi'_\alpha \circ (\psi'_\beta)^{-1})(y, v) &= (u_{\alpha\beta}(y), \psi_{\alpha\beta}(u_\beta^{-1}(y))v) \\ (T\psi'_\alpha \circ T(\psi'_\beta)^{-1})(y, v; \xi, w) &= (u_{\alpha\beta}(y), \psi_{\alpha\beta}(u_\beta^{-1}(y))v; \\ &\quad d(u_{\alpha\beta})(y)\xi, (d(\psi_{\alpha\beta} \circ u_\beta^{-1})(y))\xi)v + \psi_{\alpha\beta}(u_\beta^{-1}(y))w). \end{aligned}$$

Therefore, we see that for fixed (y, v) , the transition functions are linear in $(\xi, w) \in \mathbb{R}^m \times V$. This finding describes the vector bundle structure of the tangent bundle (TE, π_E, \mathcal{E}) .

For fixed (y, ξ) , the transition functions of TE are also linear in $(v, w) \in V \times V$. This gives a vector bundle structure on (TE, T_p, TM) . Its fiber addition will be denoted by $T(+_E) : T(\mathcal{E} \times_M \mathcal{E}) = TE \times_{TM} TE \rightarrow TE$ since it is the tangent mapping of $+_E$. Its scalar multiplication will be denoted by $T(m_t^E)$.

The space $\Xi \in TE : T_p \Xi = 0$ in $TM = (Tp)^{-1}(0)$ is denoted by VE and is called the vertical bundle over \mathcal{E} . The local form of a vertical vector Ξ is $T\psi'_\alpha \Xi = (y, v; 0, w)$; thus, the transition function looks like $(T\psi'_\alpha \circ T(\psi'_\beta)^{-1})(y, v; 0, w) = (u_{\alpha\beta}(y), \psi_{\alpha\beta}(u_\beta^{-1}(y))v; 0, \psi_{\alpha\beta}(u_\beta^{-1}(y))w)$.

Appendix A.4.7. Vector Bundle Homomorphism/Isomorphism

Let $(\mathcal{E}, p, \mathcal{M})$ and $(\mathcal{F}, q, \mathcal{N})$ be vector bundles. A vector bundle homomorphism $\varphi : \mathcal{E} \rightarrow \mathcal{F}$ is a fiber-respecting fiber linear smooth mapping (denoted by a diagram):

$$\begin{array}{ccc} \mathcal{E} & \xrightarrow{\varphi} & \mathcal{F} \\ p \downarrow & & \downarrow q \\ \mathcal{M} & \xrightarrow{\varphi} & \mathcal{N}. \end{array}$$

If φ is invertible, it is called a vector bundle isomorphism.

Appendix A.4.8. Connection

A connection on the fiber bundle $(\mathcal{E}, p, \mathcal{M}, S)$ is a vector-valued one-form (a one-form on a manifold \mathcal{M} is a smooth mapping of the total space of the tangent bundle of \mathcal{M} to \mathbb{R} whose restriction to each fiber is a linear functional on the tangent space, i.e., $\alpha : TM \rightarrow \mathbb{R}$, $\alpha_x = \alpha|_{T_x M} : T_x M \rightarrow \mathbb{R}$, where α_x is linear.) $\Phi \in \Omega^1(\mathcal{E}; VE)$ with values in the vertical bundle VE such that $\Phi \circ \Phi = \Phi$ and $Im\Phi = VE$; thus, Φ is a projection $TE \rightarrow VE$.

Connection defines parallel transport on the fiber bundle, so we can consider it as how fibers are connected over manifolds. If the fiber bundle is a vector bundle as shown in Figure A1, the connection is linear.

References

1. Dou, M.; Taylor, J.; Fuchs, H.; Fitzgibbon, A.; Izadi, S. 3D Scanning Deformable Objects with a Single RGBD Sensor. In Proceedings of the 2015 Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 493–501.
2. Li, H.; Vouga, E.; Gudym, A.; Luo, L.; Barron, J.T.; Gusev, G. 3D Self-Portraits. *ACM Trans. Graph.* **2013**, *32*, 1–9. [CrossRef]
3. Newcombe, R.A.; Fox, D.; Seitz, S.M. DynamicFusion: Reconstruction and Tracking of Non-Rigid Scenes in Real-Time. In Proceedings of the 2015 Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 343–352.
4. Li, H.; Adams, B.; Guibas, L.J.; Pauly, M. Robust Single-View Geometry and Motion Reconstruction. *ACM SIGGRAPH Asia* **2009**, *28*, 175.
5. Tevs, A.; Berner, A.; Wand, M.; Ihrke, I.; Bokeloh, M.; Kerber, J.; Seidel, H.-P. Animation Cartography-Intrinsic Reconstruction of Shape and Motion. *ACM Trans. Graph.* **2012**, *31*, 1–15. [CrossRef]
6. Cetinul, H.; Vidal, R. Intrinsic Mean Shift For Clustering on Stiefel and Grassmann Manifolds. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 1896–1902.
7. Chakraborty, R.; Vemuri, B.C. Recursive Fréchet Mean Computation On Grassmannian and Its Applications to Computer Vision. In Proceedings of the 2015 International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 2039–2047.
8. Salehian, H.; Chakraborty, R.; Ofori, E.; Vaillancourt, D.; Vemuri, B.C. An efficient recursive estimator of the Fréchet mean on hypersphere with applications to Medical Image Analysis. In Proceedings of the 2015 Mathematical Foundations of Computational Anatomy, Munich, Germany, 9 October 2015.
9. Srivastava, A.; Jermyn, I.; Joshi, S. Riemannian analysis of probability density functions with applications in vision. In Proceedings of the Computer Vision and Pattern Recognition, Minneapolis, MN, USA, 18–23 June 2007; pp. 1–8.
10. Fletcher, P.T.; Joshi, S. Riemannian Geometry for the Statistical Analysis of Diffusion Tensor Data. *Signal Process.* **2007**, *87*, 250–262. [CrossRef]
11. Sra, S.; Cherian, A. Generalized Dictionary Learning for Symmetric Positive Definite Matrices with Application to NN Retrieval. In *Machine Learning and Knowledge Discovery in Databases*; Springer: Berlin/Heidelberg, Germany, 2011; pp. 318–332.
12. Xie, Y.; Ho, J.; Vemuri, B. On a Nonlinear Generalization of Sparse Coding and Dictionary Learning. In Proceedings of the 2013 International Conference on Machine Learning, Atlanta, GA, USA, 16–21 June 2013; pp. 1480–1488.
13. Müller, M. *Information Retrieval For Music And Motion*; Springer: Berlin/Heidelberg, Germany, 2007.
14. Wang, J.; Liu, Z.; Wu, Y.; Yuan, J. Mining Actionlet Ensemble for Action Recognition with Depth Cameras. In Proceedings of the Computer Vision and Pattern Recognition, Providence, RI, USA, 16–21 June 2012; pp. 1290–1297.
15. Hussein, M.E.; Torki, M.; Gawayyed, M.A.; El-Saban, M. Human Action Recognition Using a Temporal Hierarchy of Covariance Descriptors on 3D Joint Locations. *Int. Jt. Conf. Artif. Intell.* **2013**, *86*, 639–644.

16. Goswami, M.; Gu, X.; PrithamPingali, V.; Gu, X. Computing Teichmüller Maps Between Polygons. *Found. Comput. Math.* **2017**, *17*, 1–30. [[CrossRef](#)]
17. Litany, O.; Rodolà, E.; Bronstein, A.M.; Bronstein, M.M.; Cremers, D. Non-Rigid Puzzles. *Comput. Graph. Forum* **2016**, *35*, 135–143. [[CrossRef](#)]
18. Rodolà, E.; Cosmo, L.; Bronstein, M.M.; Torsello, A.; Cremers, D. Partial Functional Correspondence. *Comput. Graph. Forum* **2017**, *36*, 222–236. [[CrossRef](#)]
19. Cosmo, L.; Rodolà, E.; Bronstein, M.; Sahillioğlu, Y. SHREC'16: Partial Matching of Deformable Shapes. In Proceedings of the International Conference on 3D Vision, Stanford, CA, USA, 25–28 October 2016.
20. Lähner, Z.; Rodolà, E.; Bronstein, M.M.; Cremers, D.; Burghard, O.; Cosmo, L.; Dieckmann, A.; Klein, R.; Sahillioğlu, Y. SHREC 16: Matching of Deformable Shapes with Topological Noise. In Proceedings of the 2016 International Conference on 3D Vision, Stanford, CA, USA, 25–28 October 2016.
21. Gu, X.D.; Zeng, W.; Luo, F.; Yau, S.T. Numerical Computation of Surface Conformal Mappings. *Comput. Methods Funct. Theory* **2012**, *11*, 747–787. [[CrossRef](#)]
22. Ovsjanikov, M.; Ben-Chen, M.; Solomon, J.; Butscher, A.; Guibas, L.J. Functional Maps: A Flexible Representation of Maps Between Shapes. *ACM Trans. Graph.* **2012**, *31*, 1–11. [[CrossRef](#)]
23. Rustamov, R.M.; Ovsjanikov, M.; Azencot, O.; Ben-Chen, M.; Chazal, F.; Guibas, L.J. Map-Based Exploration of Intrinsic Shape Differences and Variability. *Trans. Graph.* **2013**, *32*, 72. [[CrossRef](#)]
24. Davis, B.C.; Fletcher, P.T.; Bullitt, E.; Joshi, S. Population Shape Regression from Random Design Data. *Int. J. Comput. Vis.* **2010**, *90*, 255–266. [[CrossRef](#)]
25. Freifeld, O.; Black, M.J. Lie Bodies: A Manifold Representation of 3D Human Shape. In Proceedings of the 2012 European Conference on Computer Vision, Florence, Italy, 7–13 October 2012; Volume 7572, pp. 1–14.
26. Liu, Z.; Zhang, C.; Tian, Y. 3D-based Deep Convolutional Neural Network for Action Recognition with Depth Sequences. *Image Vis. Comput.* **2016**, *2016*, 93–100. [[CrossRef](#)]
27. Wang, P.; Li, W.; Gao, Z.; Zhang, Y.; Tang, C.; Ogunbona, P. Scene Flow to Action Map: A New Representation for RGB-D based Action Recognition with Convolutional Neural Networks. In Proceedings of the 2017 Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 416–425.
28. Wu, Z.; Song, S.; Khosla, A.; Yu, F.; Zhang, L.; Tang, X.; Xiao, J. 3D Shapenets: A Deep Representation for Volumetric Shapes. In Proceedings of the 2015 Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 1912–1920.
29. Harandi, M.T.; Sanderson, C.; Hartley, R.; Lovell, B.C. Sparse Coding And Dictionary Learning for Symmetric Positive Definite Matrices: A Kernel Approach. In Proceedings of the 2012 European Conference on Computer Vision, Florence, Italy, 7–13 October 2012; pp. 216–229.
30. Li, P.; Wang, Q.; Zuo, W.; Zhang, L. Log-Euclidean Kernels for Sparse Representation and Dictionary Learning. In Proceedings of the 2014 IEEE International Conference on Computer Vision, Paris, France, 27–30 October 2014; pp. 1601–1608.
31. Armstrong, M.A. *Basic Topology*; Springer: Berlin/Heidelberg, Germany, 1983.
32. Hong, Y.; Singh, N.; Kwitt, R.; Niethammer, M. Time-Warped Geodesic Regression. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Boston, MA, USA, 14–18 September 2014; Volume 17, pp. 105–112.
33. Du, J.; Goh, A.; Kushnarev, S.; Qiu, A. Geodesic Regression on Orientation Distribution Functions with Its Application to an Aging Study. *Neuroimage* **2014**, *87*, 416–426. [[CrossRef](#)] [[PubMed](#)]
34. Kim, H.J.; Adluru, N.; Collins, M.D.B.; Chung, M.K.; Bendlin, B.; Johnson, S.C. MGLM on Riemannian Manifolds with Applications to Statistical Analysis of Diffusion Weighted Images. In Proceedings of the 2014 Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 2705–2712.
35. Banerjee, M.; Chakraborty, R.; Ofori, E.; Vaillancourt, D.; Vemuri, B.C. Nonlinear Regression On Riemannian Manifolds And Its Applications To Neuro-Image Analysis. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 5–9 October 2015; Volume 9349, pp. 719–727.
36. Singh, N.; Niethammer, M. Splines For Diffeomorphic Image Regression. In Proceedings of the 2014 International Conference on Medical Image Computing and Computer-Assisted Intervention, Boston, MA, USA, 14–18 September 2014; pp. 121–129.

37. Chakraborty, R.; Banerjee, M.; Crawford, V.; Vemuri, B. An information theoretic formulation of the Dictionary Learning and Sparse Coding Problems on Statistical Manifolds. In Proceedings of the 2016 Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016.
38. Cherian, A.; Sra, S.; Banerjee, A.; Papanikolopoulos, N. Jensen-Bregman Logdet Divergence with Application to Efficient Similarity Search for Covariance Matrices. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 2161–2174. [[CrossRef](#)] [[PubMed](#)]
39. Arsigny, V.; Fillard, P.; Pennec, X.; Ayache, N. Logeuclidean Metrics For Fast And Simple Calculus on Diffusion Tensors. *Magn. Reson. Med.* **2006**, *56*, 411–421. [[CrossRef](#)] [[PubMed](#)]
40. Bhatia, R. *Positive Definite Matrices*; Princeton University Press: Princeton, NJ, USA, 2009.
41. Moakher, M.; Batchelor, P. Symmetric Positivedefinite Matrices: From Geometry to Applications and Visualization. In *Visualization and Processing of Tensor Fields*; Springer: Berlin/Heidelberg, Germany, 2006; pp. 285–298.
42. Pennec, X.; Fillard, P.; Ayache, N. A Riemannian Framework for Tensor Computing. *Int. J. Comput. Vis.* **2006**, *66*, 41–66. [[CrossRef](#)]
43. Jayasumana, S.; Hartley, R.; Salzmann, M.; Li, H.; Harandi, M. Kernel methods on the Riemannian manifold of symmetric positive definite matrices. In Proceedings of the 2013 Computer Vision and Pattern Recognition, Portland, OR, USA, 23–28 June 2013; pp. 73–80.
44. Li, B.; Ayazoglu, M.; Mao, T.; Camps, O.I.; Sznaiier, M. Activity Recognition Using Dynamic Subspace Angles. *Comput. Vis. Pattern Recognit.* **2011**, *42*, 3193–3200.
45. Li, B.; Camps, O.I.; Sznaiier, M. Cross-view activity recognition using hankellets. *Comput. Vis. Pattern Recognit. IEEE* **2012**, *2012*, 1362–1369.
46. Sivalingam, R.; Boley, D.; Morellas, V.; Papanikolopoulos, N. Tensor sparse coding for positive definite matrices. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *36*, 592–605. [[CrossRef](#)]
47. Sra, S. A new metric on the manifold of kernel matrices with application to matrix geometric means. In Proceedings of the 2012 Conference and Workshop on Neural Information Processing Systems, Lake Tahoe, NV, USA, 3–6 December 2012; pp. 144–152.
48. Wang, Z.; Vemuri, B.C. An affine invariant tensor dissimilarity measure and its applications to tensor-valued image segmentation. In Proceedings of the 2014 Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. I-228–I-233.
49. Faraki, M.; Harandi, M.T.; Porikli, F. More About Vlad: A Leap from Euclidean to Riemannian Manifolds. In Proceedings of the 2015 Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 4951–4960.
50. Harandi, M.; Salzmann, M.; Porikli, F. Bregman Divergences for Infinite Dimensional Covariance Matrices. In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 1003–1010.
51. Harandi, M.; Salzmann, M. Riemannian Coding and Dictionary Learning: Kernels to the Rescue. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 3926–3935.
52. Feragen, A.; Lauze, F.; Hauberg, S. Geodesic Exponential Kernels: When Curvature and Linearity Conflict. In Proceedings of the 2015 Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 3032–3042.
53. Gu, X.D.; Yau, S.-T. *Computational Conformal Geometry*; Higher Education Press: Beijing, China, 2008.
54. Boothby, W.M. *An Introduction to Differentiable Manifolds and Riemannian Geometry*; Academic Press: Cambridge, MA, USA, 1986; Volume 120.
55. Zhang, M.; Fletcher, P.T. Probabilistic Principal Geodesic Analysis. In Proceedings of the Conference and Workshop on Neural Information Processing Systems, Lake Tahoe, NV, USA, 5–10 December 2013.
56. Fletcher, P.T.; Lu, C.; Pizer, S.M.; Joshi, S. Principal Geodesic Analysis for the Study of Nonlinear Statistics of Shape. *IEEE Trans. Med. Imaging* **2004**, *23*, 995–1005. [[CrossRef](#)]
57. Said, S.; Coutry, N.; Bihan, N.; Sangwine, J. Exact Principal Geodesic Analysis for Data On SO(3). In Proceedings of the 2007 Signal Processing Conference, Poznan, Poland, 3–7 September 2007; pp. 1701–1705.
58. Chakraborty, R.; Seo, D.; Vemuri, B.C. An Efficient Exact-PGA Algorithm for Constant Curvature Manifolds. In Proceedings of the 2016 Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 3976–3984.

59. Sommer, S.; Lauze, F.; Hauberg, S.; Nielsen, M. Manifold Valued Statistics, Exact Principal Geodesic Analysis and the Effect of Linear Approximations. In Proceedings of the 2010 European Conference on Computer Vision, Crete, Greece, 5–11 September 2010; pp. 43–56.
60. Liu, Q.; Cao, X. Action Recognition Using Subtensor Constraint. In Proceedings of the 2012 European Conference on Computer Vision, Florence, Italy, 7–13 October 2012; pp. 764–777.
61. Belkin, M.; Niyogi, P. Laplacian Eigenmaps for Dimensionality Reduction and Data Representation. *Neural Comput.* **2003**, *15*, 1373–1396. [[CrossRef](#)]
62. Jenatton, R.; Obozinski, G.; Bach, F. Structured Sparse Principal Component Analysis. *J. Mach. Learn. Res.* **2009**, *9*, 131–160.
63. Mairal, J.; Bach, F.; Ponce, J. Sparse Modeling for Image and Vision Processing. *Eprint Arxiv* **2014**, *8*, 85–283. [[CrossRef](#)]
64. Wang, Y.-X.; Zhang, Y.-J. Nonnegative matrix factorization: A comprehensive review. *IEEE Trans. Knowl. Data Eng.* **2013**, *25*, 1336–1353. [[CrossRef](#)]
65. Cai, D.; He, X.; Han, J.; ; Huang, T.S. Graph Regularized Nonnegative Matrix Factorization for Data Representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2011**, *33*, 1548–1560.
66. Du, H.; Zhang, X.; Hu, Q.; Hou, Y. Sparse Representation-Based Robust Face Recognition by Graph Regularized Low-Rank Sparse Representation Recovery. *Neurocomputing* **2015**, *164*, 220–229. [[CrossRef](#)]
67. Gao, S.; Tsang, I.-H.; Chia, L.-T. Laplacian Sparse Coding, Hyper-Graph Laplacian Sparse Coding, and Applications. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 92–104. [[CrossRef](#)]
68. Jiang, B.; Ding, C.; Tang, J. Graph-Laplacian PCA: Closed-Form Solution and Robustness. *Comput. Vis. Pattern Recognit.* **2013**, *9*, 3492–3498.
69. Jin, T.; Yu, J.; You, J.; Zeng, K.; Li, C.; Yu, Z. Low-Rank Matrix Factorization with Multiple Hypergraph Regularizers. *Pattern Recognit.* **2014**, *48*, 1011–1022. [[CrossRef](#)]
70. Jin, T.; Yu, Z.; Li, L.; Li, C. Multiple Graph Regularized Sparse Coding and Multiple Hypergraph Regularized Sparse Coding for Image Representation. *Neurocomputing* **2014**, *154*, 245–256. [[CrossRef](#)]
71. Peng, Y.; Lu, B.-L.; Wang, S. Enhanced Low-Rank Representation Via Sparse Manifold Adaption for Semi-Supervised Learning. *Neural Netw.* **2015**, *65*, 1–17. [[CrossRef](#)]
72. Tao, L.; Ip, H.H.; Wang, Y.; Shu, X. Low Rank Approximation with Sparse Integration of Multiple Manifolds for Data Representation. *Appl. Intell.* **2015**, *42*, 430–446. [[CrossRef](#)]
73. Zhang, Z.; Zhao, K. Low-Rank Matrix Approximation with Manifold regularization. *Pattern Anal. Mach. Intell.* **2013**, *35*, 1717–1729. [[CrossRef](#)]
74. Sibson, R. SLINK: An optimally efficient algorithm for the single-link cluster method. *Comput. J.* **1973**, *16*, 30–34. [[CrossRef](#)]
75. Lu, X.; Yao, J.; Tu, J.; Li, K.; Li, L.; Liu, Y. Pairwise Linkage for Point Cloud Segmentation. In Proceedings of the 2016(III-3) ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences, Prague, Czech Republic, 12–19 July 2016; pp. 201–208.
76. Shuman, D.I.; Narang, S.K.; Frossard, P.; Ortega, A.; Vandergheynst, P. The Emerging Field of Signal Processing on Graphs: Extending High-Dimensional Data Analysis to Networks and Other Irregular Domains. *IEEE Signal Process. Mag.* **2013**, *30*, 83–98. [[CrossRef](#)]
77. Chazal, F.; Michel, B. An introduction to Topological Data Analysis: fundamental and practical aspects for data scientists. *arXiv* **2017**, arXiv:1710.04019.
78. Patania, A.; Vaccarino, F.; Petri, G. Topological analysis of data. *EPJ Data Sci.* **2017**, *6*, 7. [[CrossRef](#)]
79. Lum, P.; Singh, G.; Lehman, A.; Ishkanov, T.; Vejdemo-Johansson, M.; Alagappan, M.; Carlsson, J.; Carlsson, G. Extracting insights from the shape of complex data using topology. *Sci. Rep.* **2013**, *3*, 1236. [[CrossRef](#)]
80. Ovsjanikov, M.; Sun, J.; Guibas, L. Global Intrinsic Symmetries of Shapes *Comput. Graph. Forum* **2008**. [[CrossRef](#)]
81. Ovsjanikov, M.; Mrgot, Q.; Mxexmoli, F.; Guibas, L.J. One Point Isometric Matching with the Heat Kernel. *Comput. Graph. Forum* **2010**, *29*, 1555–1564. [[CrossRef](#)]
82. Solomon, J.; Ben-Chen, M.; Butscher, A.; Guibas, L. Discovery of Intrinsic Primitives on Triangle Meshes. *Comput. Graph. Forum* **2011**, *30*, 365–374. [[CrossRef](#)]
83. Corman, E.; Ovsjanikov, M. Functional Characterization of Deformation Fields. *ACM Trans. Graph.* **2019**, *38*, 8:1–8:19. [[CrossRef](#)]

84. Raviv, D.; Bronstein, A.M.; Bronstein, M.M.; Kimmel, R. Full and Partial Symmetries of Non-rigid Shapes. *Int. J. Comput. Vis.* **2010**, *89*, 18–39. [\[CrossRef\]](#)
85. Bronstein, A.M.; Bronstein, M.M.; Kimmel, R. Topology-Invariant Similarity of Nonrigid Shapes. *Int. J. Comput. Vis.* **2008**, *81*, 281. [\[CrossRef\]](#)
86. Smeets, D.; Hermans, J.; Vandermeulen, D.; Suetens, P. Isometric Deformation Invariant 3D Shape Recognition. *Pattern Recognit.* **2012**, *45*, 2817–2831. [\[CrossRef\]](#)
87. Shamir, A. A survey on Mesh Segmentation Techniques. *Comput. Graph. Forum* **2008**, *27*, 1539–1556. [\[CrossRef\]](#)
88. Loncaric, S. A Survey of Shape Analysis Techniques. *Pattern Recognit.* **1998**, *31*, 983–1001. [\[CrossRef\]](#)
89. Laga, H. *A Survey on Non-Rigid 3D Shape Analysis*; Academic Press Library in Signal Processing; Academic Press: Cambridge, MA, USA, 2018; Volume 6, pp. 261–304.
90. Tangelder, J.W.H.; Veltkamp, R.C. A Survey of Content Based 3D Shape Retrieval Methods. *Multimed. Tools Appl.* **2008**, *39*, 441. [\[CrossRef\]](#)
91. Coifman, R.R.; Lafon, S. Diffusion Maps. *Appl. Comput. Harmon. Anal.* **2006**, *21*, 5–30. [\[CrossRef\]](#)
92. Sun, J.; Ovsjanikov, M.; Guibas, L.J. A Concise and Provably Informative Multi-Scale Signature Based on Heat Diffusion. *Comput. Graph. Forum* **2009**, *28*, 1383–1392. [\[CrossRef\]](#)
93. Aubry, M.; Schlickewei, U.; Cremers, D. The Wave Kernel Signature: A Quantum Mechanical Approach to Shape Analysis. In Proceedings of the IEEE International Conference on Computer Vision Workshops, Barcelona, Spain, 6–13 November 2011; pp. 1626–1633.
94. Rodolà, E.; Bulò, S.R.; Windheuser, T.; Vestner, M.; Cremers, D. Dense Non-Rigid Shape Correspondence Using Random Forests. In Proceedings of the 2014 Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 4177–4184.
95. Litman, R.; Bronstein, A.M. Learning Spectral Descriptors for Deformable Shape Correspondence. *IEEE Trans. Pattern Anal. Mach. Intell.* **2014**, *36*, 170–180. [\[CrossRef\]](#)
96. Galoppo, N.; Otaduy, M.A.; Moss, W.; Sewall, J.; Curtis, S.; Lin, M.C. Controlling Deformable Material With Dynamic Morph Targets. In Proceedings of the ACM Siggraph Symposium on Interactived Graphics and Games, New Orleans, LA, USA, 4–6 August 2009; pp. 39–47.
97. Hahn, F.; Thomaszewski, B.; Coros, S.; Sumner, R.W.; Cole, F.; Meyer, M.; Deroose, T.; Gross, M. Subspace Clothing Simulation Using Adaptive Bases. *ACM Trans. Graph.* **2014**, *33*, 105. [\[CrossRef\]](#)
98. Teng, Y.; Meyer, M.; Deroose, T.; Kim, T. Subspace Condensation: Full Space Adaptivity for Subspace Deformations. *ACM Trans. Graph.* **2015**, *34*, 76:1–76:9. [\[CrossRef\]](#)
99. Kry, P.G.; James, D.L.; Pai, D.K. EigenSkin: Real Time Large Deformation Character Skinning in Hardware. In Proceedings of the 2002 Symposium on Computer Animation, San Antonio, TX, USA, 21–22 July 2002; pp. 153–160.
100. Barbič, J.; James, D.L. Real-time Subspace Integration For St. Venant-Kirchhoff Deformable Models. *ACM Trans. Graph.* **2005**, *24*, 982–990. [\[CrossRef\]](#)
101. Tycowicz, C.V.; Schulz, C.; Seidel, H.-P.; Hildebrandt, K. An Efficient Construction of Reduced Deformable Objects. *SIGGRAPH Asia* **2013**, *32*, 213. [\[CrossRef\]](#)
102. Xu, H.; Barbic, J. Pose-Space Subspace Dynamics. *ACM Trans. Graph.* **2016**, *35*, 1–14. [\[CrossRef\]](#)
103. Vemulapalli, R.; Arrate, F.; Chellappa, R. R3DG features: Relative 3D geometry-based skeletal representations for human action recognition. *Comput. Vis. Image Underst.* **2016**, *152*, 155–166. [\[CrossRef\]](#)
104. Zhang, X.; Wang, Y.; Gou, M.; Sznai, M.; Camps, O. Efficient Temporal Sequence Comparison and Classification Using Gram Matrix Embeddings on a Riemannian Manifold. In Proceedings of the 2016 Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 4498–4507.
105. Çeliktutan, O.; Wolf, C.; Sankur, B.; Lombardi, E. Real-Time Exact Graph Matching with Application in Human Action Recognition. In *Human Behavior Understanding*; Springer: Berlin/Heidelberg, Germany, 2012; pp. 17–28.
106. Gaur, U.; Zhu, Y.; Song, B.; Roy-Chowdhury, A. A “String of Feature Graphs” Model For Recognition of Complex Activities in Natural Videos. In Proceedings of the 2011 International Conference on Computer Vision, Tokyo, Japan, 25–27 May 2011; pp. 2595–2602.
107. Wang, L.; Sahbi, H. Directed Acyclic Graph Kernels for Action Recognition. In Proceedings of the 2013 IEEE International Conference on Computer Vision, Sydney, Australia, 1–8 December 2013; pp. 3168–3175.

108. Wang, P.; Yuan, C.; Hu, W.; Li, B.; Zhang, Y. Graph Based Skeleton Motion Representation and Similarity Measurement for Action Recognition. In Proceedings of the 2016 European Conference on Computer Vision, Amsterdam, The Netherlands, 11–14 October 2016; pp. 370–385.
109. Vemulapalli, R.; Arrate, F.; Chellappa, R. Human Action Recognition by Representing 3D Skeletons as Points in a Lie Group. In Proceedings of the 2014 Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 588–595.
110. Vemulapalli, R.; Chellappa, R. Rolling Rotations for Recognizing Human Actions from 3D Skeletal Data. In Proceedings of the Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 4471–4479.
111. Gong, D.; Medioni, G. Dynamic Manifold Warping for View Invariant Action Recognition. In Proceedings of the 2011 International Conference on Computer Vision, Tokyo, Japan, 25–27 May 2011; Volume 23, pp. 571–578.
112. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet Classification with Deep Convolutional Neural Networks. *Int. Conf. Neural Inf. Process. Syst.* **2012**, *60*, 1097–1105. [[CrossRef](#)]
113. Long, J.; Shelhamer, E.; Darrell, T. Fully Convolutional Networks for Semantic Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2014**, *2014*, 1.
114. Noh, H.; Hong, S.; Han, B. Learning Deconvolution Network For Semantic Segmentation. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 1520–1528.
115. Saito, S.; Li, T.; Li, H. Real-Time Facial Segmentation and Performance Capture from RGB Input. In Proceedings of the 2016 European Conference on Computer Vision, Amsterdam, The Netherlands, 11–14 October 2016; pp. 244–261.
116. Olszewski, K.; Lim, J.J.; Saito, S.; Li, H. High-fidelity facial and speech animation for VR HMDs. *ACM Trans. Graph.* **2016**, *35*, 221. [[CrossRef](#)]
117. Dosovitskiy, A.; Springenberg, J.T.; Brox, T. Learning to Generate Chairs With Convolutional Neural Networks. In Proceedings of the 2015 Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 1538–1546.
118. Bronstein, M.M.; Bruna, J.; LeCun, Y.; Szlam, A.; Vandergheynst, P. Geometric Deep Learning: Going beyond Euclidean Data. *IEEE Signal Process. Mag.* **2017**, *34*, 18–42. [[CrossRef](#)]
119. Masci, J.; Rodolà, E.; Boscaini, D.; Bronstein, M.M.; Li, H. Geometric deep learning. In Proceedings of the SIGGRAPH ASIA 2016 Courses, Macao, 5–8 December 2016.
120. Boscaini, D.; Masci, J.; Rodolà, E.; Bronstein, M. Learning Shape Correspondence with Anisotropic Convolutional Neural Networks. In Proceedings of the Neural Information Processing Systems, Barcelona, Spain, 5–10 December 2016; pp. 3189–3197.
121. Li, T.; Meng, Z.; Ni, B.; Shen, J.; Wang, M. Robust geometric ℓ_p -norm feature pooling for image classification and action recognition. *Image Vis. Comput.* **2016**, *55*, 64–76. [[CrossRef](#)]
122. Song, S.; Xiao, J. Deep Sliding Shapes for a Modal 3D Object Detection in RGB-D Images. In Proceedings of the 2016 Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 808–816.
123. Maturana, D.; Scherer, S. VoxNet: A 3D Convolutional Neural Network for Real-Time Object Recognition. In Proceedings of the Conference on Intelligent Robots and Systems IEEE, Hamburg, Germany, 28 September–2 October 2015; pp. 922–928.
124. Zeng, A.; Song, S.; Nißner, M.; Fisher, M.; Xiao, J. 3DMatch: Learning the Matching of Local 3D Geometry in Range Scans. In Proceedings of the 2017 Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017.
125. Yumer, M.E.; Mitra, N.J. Learning Semantic Deformation Flows with 3D Convolutional Networks. In Proceedings of the 2016 European Conference on Computer Vision, Amsterdam, The Netherlands, 11–14 October 2016.
126. Wei, L.; Huang, Q.; Ceylan, D.; Vouga, E.; Li, H. Dense Human Body Correspondences Using Convolutional Networks. In Proceedings of the 2016 European Conference on Computer Vision, Amsterdam, The Netherlands, 11–14 October 2016; pp. 1544–1553.
127. Bruna, J.; Zaremba, W.; Szlam, A.; LeCun, Y. Spectral Networks and Locally Connected Networks on Graphs. In Proceedings of the International Conference on Learning Representations (ICLR2014), CBLS, Banff, AB, Canada, 14–16 April 2014.

128. Masci, J.; Boscaini, D.; Bronstein, M.M.; Vandergheynst, P. Geodesic Convolutional Neural Networks on Riemannian Manifolds. In Proceedings of the 2015 IEEE Workshop on 3D Representation and Recognition, Santiago, Chile, 17 December 2015; pp. 832–840.
129. Boscaini, D.; Masci, J.; Melzi, S.; Bronstein, M.M.; Castellani, U.; Vandergheynst, P. Learning Class-Specific Descriptors for Deformable Shapes Using Localized Spectral Convolutional Networks. *Comput. Graph. Forum* **2015**, *34*, 13–23. [\[CrossRef\]](#)
130. Shuman, D.I.; Ricaud, B.; Vandergheynst, P. Vertex-Frequency Analysis on Graphs. *Appl. Comput. Harmon. Anal.* **2016**, *40*, 260–291. [\[CrossRef\]](#)
131. Andreux, M.; Rodolà, E.; Aubry, M.; Cremers, D. Anisotropic Laplace-Beltrami Operators for Shape Analysis. In Proceedings of the Sixth Workshop on Non-Rigid Shape Analysis and Deformable Image Alignment (NORDIA), Zurich, Switzerland, 12 September 2014.
132. Monti, F.; Boscaini, D.; Masci, J.; Rodolà, E.; Svoboda, J.; Bronstein, M. Geometric deep learning on graphs and manifolds using mixture model CNNs. In Proceedings of the 2017 Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017.
133. Liu, C.; Liu, J.; He, Z.; Zhai, Y.; Hu, Q.; Huang, Y. Convolutional Neural Random Fields for Action Recognition. *Pattern Recognit.* **2016**, *59*, 213–224. [\[CrossRef\]](#)
134. Du, L.; Chen, H.; Mei, S.; Wang, Q. Real-time human action recognition using individual body part locations and local joints structure. In Proceedings of the ACM SIGGRAPH Conference on Virtual-Reality Continuum and Its Applications in Industry, Zhuhai, China, 3–4 December 2016; pp. 293–298.
135. Vinagre, M.; Aranda, J.; Casals, A. A New Relational Geometric Feature for Human Action Recognition. In *Informatics in Control, Automation and Robotics*; Springer: Berlin/Heidelberg, Germany, 2015; pp. 263–278.
136. Li, C.; Wang, P.; Wang, S.; Hou, Y.; Li, W. Skeleton-based Action Recognition Using LSTM and CNN. In Proceedings of the 2017 IEEE International Conference on Multimedia and Expo Workshops, Hong Kong, China, 10–14 July 2017; pp. 585–590.
137. Wu, C.; Zhang, J.; Savarese, S.; Saxena, A. Watch-n-patch: Unsupervised understanding of actions and relations. In Proceedings of the 2015 Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 4362–4370.
138. Leonardos, S.; Zhou, X.; Daniilidis, K. Articulated Motion Estimation from a Monocular Image Sequence Using Spherical Tangent Bundles. In Proceedings of the IEEE International Conference on Robotics and Automation, Stockholm, Sweden, 16–21 May 2016; pp. 587–593.
139. Lea, C.; Reiter, A.; Vidal, R.; Hager, G.D. Segmental Spatiotemporal CNNs For Fine-Grained Action Segmentation. In Proceedings of the 2016 European Conference on Computer Vision, Amsterdam, The Netherlands, 11–14 October 2016; pp. 36–52.
140. Ke, Q.; An, S.; Bennamoun, M.; Sohel, F.; Boussaid, F. SkeletonNet: Mining Deep Part Features for 3D Action Recognition. *IEEE Signal Process. Lett.* **2017**. [\[CrossRef\]](#)
141. Veeriah, V.; Zhuang, N.; Qi, G. Differential Recurrent Neural Networks for Action Recognition. In Proceedings of the 2015 IEEE International Conference on Computer Vision, Región Metropolitana, Chile, 11–18 December 2015; pp. 4041–4049.
142. Liu, J.; Shahroudy, A.; Xu, D.; Wang, G. Spatio-Temporal LSTM with Trust Gates for 3D Human Action Recognition. In Proceedings of the 2016 European Conference on Computer Vision, Amsterdam, The Netherlands, 11–14 October 2016; pp. 816–833.
143. Liu, M.; Chen, C.; Liu, H. Learning informative pairwise joints with energy-based temporal pyramid for 3D action recognition. In Proceedings of the 2017 IEEE International Conference on Multimedia and Expo, Hong Kong, China, 10–14 July 2017; pp. 901–906.
144. Chang, A.X.; Funkhouser, T.; Guibas, L.; Hanrahan, P.; Huang, Q.; Li, Z.; Savarese, S.; Savva, M.; Song, S.; Su, H.; et al. ShapeNet: An Information-Rich 3D Model Repository. *arXiv* **2015**, arXiv:1512.03012.
145. Savva, M.; Chang, A.X.; Hanrahan, P. Semantically-Enriched 3D Models for Common-sense Knowledge. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops, Boston, MA, USA, 7–12 June 2015.
146. Bronstein, A.M.; Bronstein, M.M.; Kimmel, R. *Numerical Geometry of Non-Rigid Shapes*; Springer: New York, NY, USA, 2008; ISBN: 978-0-387-73300-5.

147. Ionescu, C.; Papava, D.; Olaru, V.; Sminchisescu, C. Human3.6M: Large Scale Datasets and Predictive Methods for 3D Human Sensing in Natural Environments. *IEEE Trans. Pattern Anal. Mach. Intell.* **2014**, *36*, 1325–1339. [[CrossRef](#)]
148. Ionescu, C.; Li, F.; Sminchisescu, C. Latent Structured Models for Human Pose Estimation. *Int. Conf. Comput. Vis.* **2011**, *58*, 2220–2227.
149. Bourdev, L.; Malik, J. Poselets: Body Part Detectors Trained Using 3D Human Pose Annotations. In Proceedings of the IEEE International Conference on Computer Vision, Kyoto, Japan, 29 September–2 October 2009; pp. 1365–1372.
150. Spinello, L.; Arras, K.O. People Detection in RGB-D Data. In Proceedings of the International Conference on Intelligent Robots and Systems, San Francisco, CA, USA, 25–30 September 2011; pp. 3838–3843.
151. Luber, M.; Spinello, L.; Arras, K.O. People Tracking in RGB-D Data with On-line Boosted Target Models. In Proceedings of the International Conference on Intelligent Robots and Systems, San Francisco, CA, USA, 25–30 September 2011; pp. 3844–3849.
152. Song, S.; Xiao, J. Tracking Revisited Using RGBD Camera: Unified Benchmark and Baselines. In Proceedings of the 2013 IEEE International Conference on Computer Vision, Sydney, Australia, 1–8 December 2013; pp. 233–240.
153. Stein, S.; McKenna, S.J. Combining Embedded Accelerometers with Computer Vision for Recognizing Food Preparation Activities. In Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing, Zurich, Switzerland, 8–12 September 2013; pp. 729–738.
154. Kwolek, B.; Kepski, M. Human Fall Detection on Embedded Platform Using Depth Maps and Wireless Accelerometer. *Comput. Methods Programs Biomed.* **2014**, *117*, 489–501. [[CrossRef](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).