The Metric Approach to Shape Matching

Alfonso Ros
Informatics - Technische Universität München

Abstract

The problem of shape matching can be modeled as a metric space where the objects are shapes grouped by their isometric transformations, and the distance between them is defined by a similarity metric. The Gromov-Hausdorff distance defines such a metric for the space of isometric classes of shapes. This is because of its property of isometric invariance. Nevertheless, due to the abstraction of its original definition, a relaxation is presented with the goal of reaching a practical use. Also very important properties from the distance are pointed such as the consistence to sampling. Finally the application of this approach is shown in the work of Facundo Memoli, demonstrating the practical use in real live applications.

1 Introduction

One of the fundamental problems in computer vision is that of finding the correspondence between shapes. The applications range from robotics to medical imaging. There exist multiple approaches to find reasonable solutions to this problem. One of them is the Metric approach.

When thinking about the similarity between two shapes, it is very natural to express it as a distance on some abstract shape space. In other words, in this space, shapes that are very different should be far from each other and shapes that look similar should be closer. Moreover, shapes that are intrinsically the same object, should be grouped as one point in this space. For example, when comparing two shapes of a hand with different poses, we say that intrinsically they are the same object, but extrinsically they are different.

Formally, these intrinsic properties are preserved by isometric transformations. It is said that two shapes are equivalent if they are isometric. These isometries define an equivalence relation on the space of shapes.

This new space constructed by the equivalent classes of shapes under isometric transformation is called the quotient space. In the figure 1 you can find an intuitive representation of the quotient space.

![Figure 1: Illustration of the quotient space M* constructed from the equivalence relation given by the isometries between shape](image)

It turns out that this space exist and there is a metric distance that defines it. This distance is insensitive to non-rigid deformations, or in other words, it is isometry invariant. This distance is known as the Gromov-Hausdorff distance and it is a metric over the space of isometric classes.
2 Metric Spaces and Hausdorff distance

In this approach we treat the shapes and the similarity between shapes as a metric space. Given the definition of a metric space, the set will contain the equivalent classes of shapes defined by isometric transformations, and will be the Gromov-Hausdorff distance.

**Definition 1.** (Metric space) A metric space is a ordered pair \((M,d)\) where \(M\) is a set and \(d\) is a metric on \(M\), i.e. a function \(d : M \times M \to \mathbb{R}\) such that for every \(x,y,z \in M\), the following holds:

1. \(d(x,y) \geq 0\) (non-negative)
2. \(d(x,y) = 0 \iff x = y\) (identity of indistinguishables)
3. \(d(x,y) = d(y,z)\) (symmetry)
4. \(d(x,z) \leq d(x,y) + d(y,z)\) (triangle inequality)

The Hausdorff distance presented in the definition can be used to measure how far are two subsets of a metric space. Informally, this two sets are close if every point in one set is close to a point in the other set.

**Definition 2.** Let \((Z,d_Z)\) be a metric space and \(X,Y \subset Z\). The Hausdorff distance between \(X\) and \(Y\) is defined as

\[
d_Z^H(X,Y) := \max \left( \sup_{x \in X} \inf_{y \in Y} d_Z(x,y), \sup_{y \in Y} \inf_{x \in X} d_Z(x,y) \right)
\]

In the figure a diagram is shown to illustrate how the Hausdorff distance look. It is easy to see that if this distance is equal to zero, then the two subsets have to be exactly the same. This basically means the Hausdorff distance is perfect to compare shapes by their extrinsic properties. For this reason, it is one of the best performing cost function in the well known ICP algorithm.

Nevertheless, the Hausdorff distance is not invariant to isometries. This means, if two shapes are isometric, meaning that there exist an isometric map (definition 3) between the two, the Hausdorff distance will put them far in the metric space, even if they share the same intrinsic properties.

**Definition 3.** (Isometry) Let \((X,d_X)\) and \((Y,d_Y)\) be two metric spaces. A map \(f : X \to Y\) is called isometry if it is distance preserving \((d_X(x,x') = d_Y(f(x),f(x')))\) and bijective.

It is important to note that an isometry could be define between different metric spaces but the Hausdorff distance is defined only inside one metric space. This is important because sometimes different shapes are going to be represented in different metric spaces with different intrinsic distances and in order to compare them, they need to be in the same metric space.
3 Gromov-Hausdorff distance

We finished the last section mentioning that the Hausdorff distance is not isometric invariant and is not defined between different metric spaces. This is what motivates the main idea of the Gromov-Hausdorff distance. That is, to embed the two shapes represented by two metric spaces into a common ambient space where the two shapes can be compared with the Hausdorff distance. The way this embedding is done, is through isometric maps, what will give the isometric invariant property.

The Gromov-Hausdorff distance was introduced in 1981 by the Russian-born mathematician Mikhail Gromov and first applied to the field of pattern recognition by Facundo Mémoli and Guillermo Sapiro in 2004. Originally, the Gromov-Hausdorff distance was defined as

**Definition 4.** (Gromov-Hausdorff distance) Let \((X, d_X)\) and \((Y, d_Y)\) be two compact metric spaces, \(f : X \to Z\) and \(g : Y \to Z\) two isometric embedding mappings into the ambient metric space \((Z, d_Z)\). The Gromov-Hausdorff distance is defined as

\[
d_{GH}(X, Y) := \inf_{f : X \to Z, g : Y \to Z} d_Z(f(X), g(Y))
\]

The Gromov-Hausdorff distance is equal to zero when there exist an isometry between the two metric spaces, which means it is isometric invariant. This is a really notable result and in fact the Gromov-hausdorff distance define a metric on the space of isometry classes of compact metric spaces.

The notion of the Gromov-Hausdorff distance seems very abstract and in practical terms it is not very helpful. For this reason, an alternative definition is needed. In order to do this, in [4] it is first introduced the concept of correspondence.

**Definition 5.** (Correspondence). For sets \(A\) and \(B\), a subset \(R \subset A \times B\) is a correspondence (between \(A\) and \(B\)) if and only if

1. \(\forall a \in A\), there exist \(b \in B\) s.t. \((a, b) \in R\)
2. \(\forall b \in B\), there exist \(a \in A\) s.t. \((a, b) \in R\)

Let \(\mathcal{R}(A, B)\) denote the set of all possible correspondences between sets \(A\) and \(B\).

In particular, it is consider a correspondence between the elements of the metric spaces representing the shapes. Now, it is possible to quantify the quality of the correspondences by the difference of the distances between two pairs of corresponding points in each of the shapes. This is called the distortion of the correspondence. More precisely, suppose two points \((x, x'), (y, y') \in R\), the distortion is given by the function

\[
\Gamma(x, y; x', y') = |d_X(x, x') - d_Y(y, y')|
\]

In the figure 3 is a simple diagram trying to illustrate the notion of distortion between the two pairs of corresponding points. If the correspondence is good, the difference between the distances on each space (red dashed line) should be low, meaning a low distortion.

![Figure 3: Distance in each Metric space between two correspondent points](image-url)
In general, the distortion of the correspondence \( R \) is defined as the \( L^\infty \) norm of the distortion between each pair of corresponding points. More precisely, the distortion of \( R \) is defined as

\[
\| \Gamma \|_{L^\infty(X \times Y)} = \sup_{x_1, x_2 \in X \atop y_1, y_2 \in Y} \Gamma(x_1, x_2; y_1, y_2)
\]  \hspace{1cm} (4)

It is important to note that when the distortion is equal to zero, the correspondence \( R \) is associated with an isometry. This means that the problem now is about finding the correspondence \( R \) such has zero distortion. In other words, this brings the possibility to alternatively write the concept of the Gromov-Hausdorff distance as

\[
d_{GH}(X, Y) = \frac{1}{2} \inf_{R \in R} \| \Gamma \|_{L^\infty(X \times Y)}
\]  \hspace{1cm} (5)

Note that the \( L^\infty \) norm of the distortion has replaced the Hausdorff distance and more importantly, the infimum is now only over all the possible correspondences between the points in the two metric spaces. In terms of practice and implementation, this formulation is much more friendly than the original presented in equation 2. Nevertheless, this formulation is sensitive to errors and outliers and for this reason, a relaxation of the \( L^\infty \) norm is needed.

### 4 Relaxation of the Gromov-Hausdorff

In the previous section, we obtained an equivalent definition for the Gromov-Hausdorff distance as the infimum \( L^\infty \) norm distortion of a correspondence. The problem with this formulation is the sensitivity of the \( L^\infty \) norm to noise error and outliers. It is easy to see why because the \( L^\infty \) norm gives us the largest distortion produced by a particular combination of corresponding points. In this sense, having just one mismatching correspondence, even if the rest are perfect corresponding pairs, the general distortion will be given by this particular mismatching pair of points.

The natural solution would be to relax the \( L^\infty \) norm to the \( L^p \) norm to make the distance robust against outliers. But, we can not simply substitute the norms inside the formula, because the \( L^p \) norm is not consistent to non-uniform sampling. This comes from the fact that the \( L^p \) norm introduces an integration over the distortions between corresponding points.

In a practical sense, the integration is performed over a discrete set of points on each shape. In order to make the \( L^p \) norm consistent, this set of points should be uniform in the sense of the areas they represent. In case that the area is not the same, in order to preserve the consistency the distortion should be weighted by the area the points represent.

With this idea in mind, we proceed first by thinking about a correspondence between two shapes as a generalization of a function. In particular, in \([5]\) the correspondence is formulated alternatively as a binary function \( \mu^R : X \times Y \to [0, 1] \) satisfying for every \( x \in X \) and \( y \in Y \)

\[
\max_{y \in Y} \mu^R(x, y) = \max_{x \in X} \mu^R(x, y) = 1
\]  \hspace{1cm} (6)

Meaning, \( \mu^R(x, y) = 1 \) if \( (x, y) \in R \) and zero otherwise. This basically means that there exist only one correspondent point in the set \( Y \) for every point in \( X \) and vice versa. Now, restricting ourselves to the case of a finite metric space, we can rewrite the Gromov-Hausdorff distance defined in the previous section with the binary function \( \mu \) replacing the infimum and supremum with the minimum and maximum as

\[
d_{GH}(X, Y) = \frac{1}{2} \min_{R \in R} \max_{x_1, x_2 \in X} \max_{y_1, y_2 \in Y} \Gamma(x_1, x_2; y_1, y_2) \mu^R(x_1, y_1) \mu^R(x_2, y_2)
\]  \hspace{1cm} (7)

The function \( \mu \) is the key to express the area represented by each pair of corresponding point. In order to relax the \( L^\infty \) norm, in \([5]\) the binary function \( \mu \) is first relaxed to the continuous interval \([0, 1]\). With this relaxation, the constraint \([6]\) is also relaxed in the continuous version as follows

\[
\int_A \int_Y \mu(x, y) dy dx = \int_A dx; \int_B \int_X \mu(x, y) dx dy = \int_B dy
\]  \hspace{1cm} (8)
Figure 4: Relaxed condition for the function associated to the correspondence between two metric spaces. The area of $A$ should be equal to the integration of the elements in the metric space $Y$ weighted by the function $\mu$, and the same with the subset $B$ and the set $X$. Where $A \subset X$ and $B \subset Y$ are any measurable subsets. This condition is illustrated on the figure. Basically, the condition is constraining the corresponding areas to be equal. Maybe an easier way to understand the condition, is thinking about the set $A$ as a single element of the metric space $X$, then the equation would look like

$$\int_Y \mu dy = 1$$

And it is easy to see that the function $\mu$ is nothing but a probability distribution over the elements of the set $Y$. In other words, the function $\mu$ is just preserving the area being mapped from $X$ to the weighted area of $Y$. This is exactly the condition needed to make the $L^p$ norm consistent to sampling. Now we can introduce this relaxation to the last definition of the Gromov-Hausdorff distance as follows

$$d_{GH}(X, Y) = \frac{1}{2} \inf_{\Gamma} \|\Gamma\|_{L^p(R \times R)}$$

Where $1 \leq p \leq \infty$, and

$$\|\Gamma\|_{L^p(R \times R)} = \left(\int_{(R \times R)^2} \Gamma^p(x, x'; y, y') \mu^R(x, y) \mu^R(x', y') dx dy dx' dy'\right)^{1/p}$$

This family of equations are the so called Mass Transportation distances or in the Math community as Wasserstein-Kantorovich-Rubinstein distances or Earth Mover’s distance. The intuition behind the name is that the quantity $\mu(x, y) dx$ can be thought of as the infinitesimal amount of mass transported from the point $x$ in $X$ to the point $y$ in $Y$ and $\Gamma^p$ is the cost associated.

The most remarkable of this formulation is that not only the metric spaces $(X, d_X)$ and $(Y, d_Y)$ are considered, but now this is a more restrictive setting where the probability measurement $\mu^R$ is added into the formula. These together form what is formally called metric measure spaces or mm-space for short. These mm-spaces are denoted as triplets $(X, d_X, \mu_X)$ and $(Y, d_Y, \mu_Y)$ where the functions $\mu_X$ and $\mu_Y$ are called measure. In this more restrictive setting, only the isometries that preserve the area on the surface are solutions (isomorphism). If such a isometry exist, the two mm-spaces are called isomorphic.

5 Discretization

In the last section we saw a continuous formulation of the relaxed Gromov-Hausdorff distance using the $L^p$ norm. Unfortunately, it is not possible to work in a continuous setting with a computer. For this reason, we are constraint to work in a discrete setting, what would transform the previous formulation of the distance as follows

$$d_{GH}(X, Y) = \frac{1}{2} \min_{\Gamma} \left(\sum_{i,j,k,l} \sum_{m,n} C_{ijkl} M_{i,j}^R \mu_{kl}^R\right)^{1/p}$$
Notice that the integrals have been replaced by sums and the infimum by a minimum. This expression can be rewrite in matrix notation as

\[
d_{GH}(X,Y) = \frac{1}{2} \min_{R \in [0,1]^{n \times n}} \text{vec}(R)^T C \text{vec}(R)
\]  

s.t. \( R^T 1 = 1, R 1 = 1 \)  

(12) (13)

Where \( R \in \mathbb{R}^{n \times n} \) is a square matrix representing the correspondence between each possible pair of points between the two shapes. Formally, due the two constraints, the matrix \( R \) is doubly-stochastic. In other words, all of its rows have to sum to one and all of its columns have to sum to one. The matrix \( C \in \mathbb{R}^{n^2 \times n^2} \) is a huge cost matrix containing the cost of all possible pairs of corresponding points.

An interesting problem arises in practice when the number of points in the two shapes is not equal. If this is the case, the matrix \( R \) is no longer a square matrix and can not fulfill both of the constraints that makes it doubly-stochastic. For this reason, in practice, one of the constraints must be dropped, making the resulting matrix substochastic. This means that either the rows of the matrix should sum up to one or the columns of the matrix should sum up to one.

One solution to this problem is to add an extra point to the matrix in such a way that the points with no correspondences in the matching shape are assigned to this extra point. For example, lets suppose we have \( n \) and \( m \) points in the two shape respectively. Without lost of generality suppose that \( n > m \), we add a point \( m^* \) which would be represented as an extra column in the matrix \( R \). We drop the column-stochastic constraint and proceed to find the solution.

Notice that this strategy is useful to partial matches, where one of the shapes may have missing parts and some points will not have any correspondence in the other shape. This points without a correspondence, should be matched to the \( m^* \) point. An illustration of this particular case is presented in the figure. It is important to mention that finding the correct cost in assigning points to the point \( m^* \) is a really difficult problem and there is no an optimal solution.

Figure 5: Partial matching of shapes.

At this point we have to stablish the connection between this formulation and the definition of the Quadratic assignment problem which is defined as

\[
L_p(\delta) := \min_{\delta \in \Delta} \sum_{ij} \sum_{kl} t_{ij}^{(p)} \delta_{ij} \delta_{kl}
\]

(14)

Where \( \Delta \) is the set of matrices constrained to

1. \( \delta_{ij} \in \{0, 1\} \) for all \( i, j \)
2. \( \sum_i \delta_{ij} = 1 \) for all \( j \)
3. \( \sum_j \delta_{ij} = 1 \) for all \( i \)

Is almost evident to see the direct connection with the problem of calculating the Gromov-Hausdorff distance. This connection is important because it is know that the QAP is an NP-hard problem, which directly gives us the order of complexity of calculating the Gromov-Hausdorff distance.
6 Sampling

In the previous section was shown a discreet formulation of the Gromov-Hausdorff distance. Now, in order to compute the distance between two shapes we need to pick the discreet points from each of the shapes. This subset of points is called sampling. Optimally this sampling should be as dense as possible in order to have a good representation of the underlying surface, but at the same time should be possible to handle in terms of computational complexity and storage.

Obviously, there are infinitely many ways to obtain a sampling from a continuous surface, so the criterion to define a good sampling is needed. In general, what is desired is to select the minimum subset of points such that they are distributed over the surface in a way that every other point in the surface is not farther than a fixed distance. This idea is capture essentially by two concepts, \( r \)-covering and \( r \)-separated.

An \( r \)-covering is a subset of points such that every other point in the surface is contained in a neighborhood of radius \( r \) centered in one of the points inside the covering. Formally, the definition of an \( r \)-covering is given by

**Definition 6.** \((r\text{-covering})\) \( X' \) is an \( r \)-covering of \( X \) if

\[
\bigcup_{x' \in X'} B_r(x') = X
\]

Where \( B_r(x) := \{ x' \in X \mid d_X(x, x') \leq r \} \)

An \( r \)-covering does not leave out regions in diameter larger than \( r \) in the sense of the metric \( d_X \). Note that an \( r \)-covering could have an arbitrary large of points in the sense that if we add one more point from \( X \) to the covering, it will continue being an \( r \)-covering. Therefore, as we are interested in keeping just the minimum number of points such that they form an \( r \)-covering of \( X \), we need the concept of \( r \)-separation.

An \( r \)-separated set of points is just a set of point where the distance from each individual point to all the others is larger than \( r \), or more formally, \( d_X(x, x') \geq r \) for all \( x, x' \in X' \). A visual example of how the sampling would look like is presented on figure ??

![Figure 6: This picture shows an \( r \)-covering \( X' \) of the surface \( X \). The points inside \( X \) are the sampled points and the blue spheres have radius \( r \).](image)

We present now the Farthest Point Sampling algorithm, which is used to obtain a sampling which is both, \( r \)-covering and \( r \)-separated by construction.

6.0.1 Farthest point sampling

The Farthest point sampling algorithm is a greedy way to sample a surface. The algorithm is shown in the figure ?? The idea is to pick a starting point \( x_1 \) in some surface \( X \) and then sample the next point in such a way that it is the farthest to the initial point. Then simply proceed in the same way, sampling points with the largest distance to all the already sampled points. In the figure ?? is presented an example of sampling from a shape and a illustration of the FPS algorithm starting from some point \( x_1 \).
**Procedure 1** Farthest point sampling

**Input:** Metric space \((X, d_x)\), some initial point \(x_1 \in X\) and the desire sampling radius \(r_0\).

**Output:** a sampling \(X' = \{x_1, \ldots, x_N\}\).

\[X' \leftarrow \{x_1\}\]

\[d(x) \leftarrow d_X(x, x_1)\]

while \(r > r_0\) do

Find the farthest point from \(X', x' = \arg\max_{x \in X} d(x)\)

Update the set of selected samples: \(X' \leftarrow X' \cup \{x'\}\)

Update the distance function \(d(x)\),

\[d(x) \leftarrow \min_{x \in X}\{d(x), d_X(x, x')\}\]

\[r \leftarrow \max_{x \in X} d(x)\]

end

(a) Example of a sampling obtained from a shape using the Farthest Point Algorithm.

(b) Illustration of the farthest sampling point algorithm iterations starting from the point \(x_1\).

Obviously, there are many possible starting points to use in the FPS algorithm. A common solution of this problem is to start taking the two points that are far the most in the shape as initial points.

The resulting sampling will be by construction an \(r\)-covering and it will be \(r\)-separated. It is then evident that for a sampling \(\{x_1, \ldots, x_N\} \subset X\) then \(d_{GH}(X, \{x_1, \ldots, x_N\}) \leq r\), which means the Gromov-Hausdorff distance is consistent to sampling. Basically the more sampling points we get, the closer the sampling gets to the underlying surface. If we have two shapes \(X\) and \(Y\), let \(X = \{x_i\}_{i=1}^m\) be an \(r\)-covering of \(X\) and \(Y = \{y_j\}_{j=1}^{m'}\) be an \(r'\)-covering of \(Y\). Then

\[
|d_{GH}(X, Y) - d_{GH}(X, \mathcal{X})| \leq d_{GH}(X, \mathcal{X}) + d_{GH}(Y, \mathcal{Y})
\]

This means that the true distance between two continuous shape is approximated by the distance between their coverings up to the sum of the radius of their coverings. The more points the samplings have, smaller the radius of their coverings and closer to the true distance between the shapes.

### 7 Applications

Finally, in the application of the metric aproach, some results are showed on [4]. These results are presented on the figure [7]. A Database with seven types of shapes was used for comparison and
Figure 7: Results showed by F. Memoli on his paper named *On the use of Gromov-Hausdorff distances for shape comparison*

each of the them with ten different poses. From each shape 50 points where sampled using the FPS algorithm and for the measure was used the density of points in their Voronoi partitions.

The negative aspects of this approach is first of all, the slow convergence of the algorithm and the problems associated to the gradient descent algorithm. This problems are for example choosing the correct step size. Also In this particular case, as the minimization is no convex, there are local minima where the algorithm can get stuck. And finally, the solution of this formulation is a soft assignation of each pair of points which has to be binarized to a final correspondence. This final binarization is not clear how it should be done to give the optimal solution. On the other hand, the implementation is easy and the local optima are usually good enough in practice.

In general, The Gromov-Hausdorff distance have been proven to be a really elegant solution to the matching problem. Nevertheless, there is still a lot of research over the metric approach for shape matching.

References


