

# Rotation Invariant Local Shape Descriptors for Classification of Archaeological 3D Models

Edgar Roman-Rangel<sup>1</sup>(✉), Diego Jimenez-Badillo<sup>2</sup>,  
and Stephane Marchand-Maillet<sup>1</sup>

<sup>1</sup> CVMLab, University of Geneva, Geneva, Switzerland

{edgar.romanrangel, stephane.marchand-maillet}@unige.ch

<sup>2</sup> National Institute of Anthropology and History of Mexico (INAH),  
Mexico City, Mexico

diego-jimenez@inah.gob.mx

**Abstract.** We introduce a method for estimation of rotation invariant local shape descriptors for 3D models. This method follows a successful idea commonly used to obtain rotation invariant descriptors in 2D images, and improves it by tackling the difficulty of the 3 degrees of freedom that exists in 3D models. Our method is simple, yet it achieves high levels of invariance after rotation transformations, and it produces short descriptors that can be efficiently used in several tasks. Such is the case of automatic classification of 3D surfaces with archaeological value, in which the proposed method attains state-of-the-art results in shorter times when compared with previous methods.

**Keywords:** 3D models · Shape descriptor · Rotation invariant

## 1 Introduction

Achieving invariance to rotation is among the principal challenges when designing local shape descriptors [1, 2]. This problem becomes specially difficult when dealing with 3D data, where an extra degree of freedom comes into play, such that it results unclear how to define an oriented reference frame [3].

Many previous works have proposed approaches of different nature to address the problem of rotation invariance for local 3D shape descriptors. From approximating Gaussian convolutions of point cloud vertices [4], which roughly resembles the traditional Gaussian scale space used for 2D images [5]; to computing local descriptors several times, each one from a different viewpoint, and then selecting one of these descriptors according to the specific needs of the problem of interest [6]; to implementing Singular Value Decomposition (SVD) or Eigenvalue Decomposition (EVD) for the definition of oriented reference frames [3, 7]. However, these methods either fail in providing unambiguous reference orientations for the local descriptors, or result in very time-consuming approaches.

We introduce a method that is of fast computation and that produces rotation invariant local shape descriptors for 3D models. We combined our method for

rotation invariance with a method for local description that roughly follows the well known Shape Context used in 2D images [8]. This is, it describes a point of interest using the spatial distribution of its nearby points. Namely, our method aligns each point of interest using spherical coordinates that are computed with respect to the centroid of the cloud. This allows to produce local descriptors that are consistent across rotations changes, even without knowing the actual orientation of the 3D model itself, i.e., free positioning of the model in the 3D space. Note that in general, our method is independent from the local descriptor, as it can be plugged to different descriptive functions.

We evaluated the proposed method on a set of classification experiments, using 3D surfaces that represent archaeological potsherds from the Teotihuacan culture. Classification of potsherds is a very challenging problem because: (1) of the large amount of them that are found in a common excavation site, e.g., tens of thousands of fragments; and (2) the fact that often potsherds from different ceramic types are discovered together and scrambled. Also, it is of special interest in Archaeology, as classifying newly discovered potsherds brings most of the insights regarding the daily life of ancient civilizations. Our results show that the proposed method is both robust to rotations transformations, as it is fast.

The remaining of this papers is organized as follows. Section 2 introduces previous works that have addressed the problem of rotation invariance for 3D shape descriptors. Section 3 presents the framework we propose in this paper to achieve rotation invariant local shape descriptors of 3D models. Section 4 details the experimental setup we followed to evaluate the proposed method. Section 5 discusses our results. And Sect. 6 presents our conclusions.

## 2 Related Work

Several method for local description of 3D models have been developed [1,2] focusing on different types of information: such as geometry or depth.

Spin Images [9] are one of the earliest methods to show promise for description of 3D models. It consists in counting the number of neighboring points that fall inside the regions of a volumetric space, which is centered at a point of interest, and that is orientated according to the normal of the tangent plane at such point of interest. In turn, the number of neighbor points in each region is assigned to bins in a histogram. This method was latter improved by using a 3D approximation of a Gaussian scale space that enables invariance to scale transformations: Scale Invariant Spin Images (SISI) [4]. Although this approach works well for oriented 3D models, it fails in the presence of rotated instances due to the lack of definition of an oriented reference frame.

The 3D Shape Context (3DSC) [6] is an extension of the Shape Context descriptor [8]. This method defines a spherical space around a point of interest, and divides it into regions. Then, the number of nearby points, with respect to the point of interest, that are in each of the  $d$ -th regions is assigned as the value of the  $d$ -th dimension of a descriptive vector. Given the lack of an oriented reference frame, 3DSC aligns the north pole of the sphere with the normal of the 3D surface

estimated at the point of interest, then it computes  $L$  discrete rotations and a local descriptor for each of them. Finally, during a given comparison task, e.g., classification, a permutation step is conducted to compare the  $L$  versions of the descriptor, thus impacting its computational cost.

As improvement to 3DSC, the Unique Shape Context for 3D data descriptor (USC) [7] implements a SVD or EVD approach to estimate an oriented frame. Namely, it approximates a Total Least Square estimation with the three principal eigenvectors of a weighted covariance matrix, followed by a sign disambiguation that defines the orientation of two axis in the 3D space.

Similar to USC, the Unique Signature of Histograms for Local Surface Description (SHOT) [3] also implements SVD or EVD. However, it computes a local reference rather than a global one, as USC does. Namely, SHOT has two main features that differ from USC: (1) it only considers points within the local sphere, and (2) it performs reorientation of local eigenvectors such that their signs are coherent with the majority of vectors that each of them represents.

Besides the similarity in construction that all these methods share, i.e., they rely on the counting of nearby points within spherical regions, they behave differently under rotation variations, i.e., not all of them are robust enough. Also, 3DSC, USC, and SHOT might be very time consuming.

### 3 Local Descriptor

As previously mentioned, the base approach we use for local descriptions is a Shape Context-like method [8], which builds an histogram of the relative position of the nearby points for a given point of interest. More precisely, we define a spherical local context around a point of interest. Such a sphere is subdivided into 2 distance intervals, 8 azimuth intervals, and 4 zenith intervals, which are equally spaced across their corresponding axis, thus forming 64 spherical regions  $R_{d=1,\dots,64}$ , each of which corresponds to one of the dimensions  $d$  of the local descriptor  $V(d)$ .

Mathematically, the value of the  $d$ -th dimension of the descriptor is give by,

$$V_i(d) = |\{p_j \neq p_i : (p_i - p_j) \in R_d\}|, \quad (1)$$

where,  $|\cdot|$  indicates cardinality operator,  $\{\cdot\}$  denotes a set,  $p_i$  is the point of interest,  $p_j$  is a neighbor point of  $p_i$ , and the term  $(p_i - p_j)$  in Eq. (1) denotes vector difference.

In practice, we are interested in knowing which of the neighboring points  $p_j$  are located within the  $d$ -th spherical region defined for the point of interest  $p_i$ . This can be easily estimated by knowing the spherical relative coordinates of all points  $p_j$  with respect to  $p_i$ , i.e., redefining  $p_j$  as  $p_j = (\rho_j^i, \theta_j^i, \phi_j^i)$ , which can be computed by,

$$\begin{aligned}
\rho_j^i &= \sqrt{d_x^2 + d_y^2 + d_z^2}, \quad 0 \leq \rho \leq \infty, \\
\theta_j^i &= \tan^{-1} \left( \frac{d_y}{d_x} \right), \quad 0 \leq \theta \leq 2\pi, \\
\phi_j^i &= \cos^{-1} \left( \frac{d_z}{\rho_j^i} \right), \quad 0 \leq \phi \leq \pi,
\end{aligned} \tag{2}$$

where,  $(d_x, d_y, d_z)$  is the component-wise distance between points  $p_i$  and  $p_j$ . Note that the boundaries in Eq. (2) indicate that the azimuth intervals cover a whole round along the sphere, whereas the zenith intervals cover half a round along it.

### 3.1 Rotation Invariance

To account for rotation invariance, we can realign the point of interest with respect to a reference point, which must be the same for all points of interest. In the case of 2D images, this alignment is computed by subtracting the value of the local orientation of the point of interest from itself and from the local orientation of all its neighboring points, such that  $p_i$  would be at  $0^\circ$  and all its nearby points would be shifted accordingly [8].

This procedure is rather straight forward in the 2D case, where the local orientation of a point is defined as the counter-clock wise inclination with respect to the horizontal axis. In the case of 3D models, this definition is less clear given the three degrees of freedom that exist. Therefore, we propose the use of the centroid of the point cloud as a reference point to account for rotation invariance. More specifically, we realign the point of interest (and all its nearby points) by subtracting from it its spherical orientations computed with respect to the centroid  $c$  of the 3D model, such that  $p_i$  would rest at  $0^\circ$  both in the azimuth and the zenith axis, and its neighboring points would be shifted accordingly. Mathematically, this is computed as,

$$\begin{aligned}
\bar{\theta}_j^i &= (\theta_j^i - \theta_i^c) \bmod 2\pi, \\
\bar{\phi}_j^i &= (\phi_j^i - \phi_i^c) \bmod \pi.
\end{aligned} \tag{3}$$

where,  $\theta_i^c$  and  $\phi_i^c$  indicate, respectively, the azimuth and zenith orientations of  $p_i$  computed with respect to the centroid  $c$ , which in turn, is estimated as the mean of the  $(x, y, z)$  coordinates of the points in the 3D model,

$$x_c = \frac{1}{N} \cdot \sum_{i=1}^N x_i, \quad y_c = \frac{1}{N} \cdot \sum_{i=1}^N y_i, \quad z_c = \frac{1}{N} \cdot \sum_{i=1}^N z_i, \tag{4}$$

note the modulus normalization in Eq. (3), which is required to comply with the boundaries defined in Eq. (2).

By using  $(\rho_j^i, \bar{\theta}_j^i, \bar{\phi}_j^i)$ , from Eqs. (2) and (3), as the arguments to identify the spherical region for each neighboring point  $p_j$  (Eq. (1)), it is possible to construct a rotation invariant descriptor for  $p_i$ .

Finally, the visual similarity between two points can be estimated as the distance between their respective local descriptors. To account for sets of different size (3D models with different amount of points), we normalize the descriptors such that summing up their dimensions equals 1,

$$H_i(d) = \frac{V_i(d)}{\sum_{d=1}^{64} V_i(d)}, \quad (5)$$

where  $V_i$  is defined as in Eq. (1).

## 4 Experimental Setup

In this section we first provide a description of the data used to evaluate the proposed method. Then we present two types of evaluation that we conducted: one for measuring the level of invariance against rotation, and one to evaluate the impact of the method in the task of classification of 3D data.

### 4.1 Data

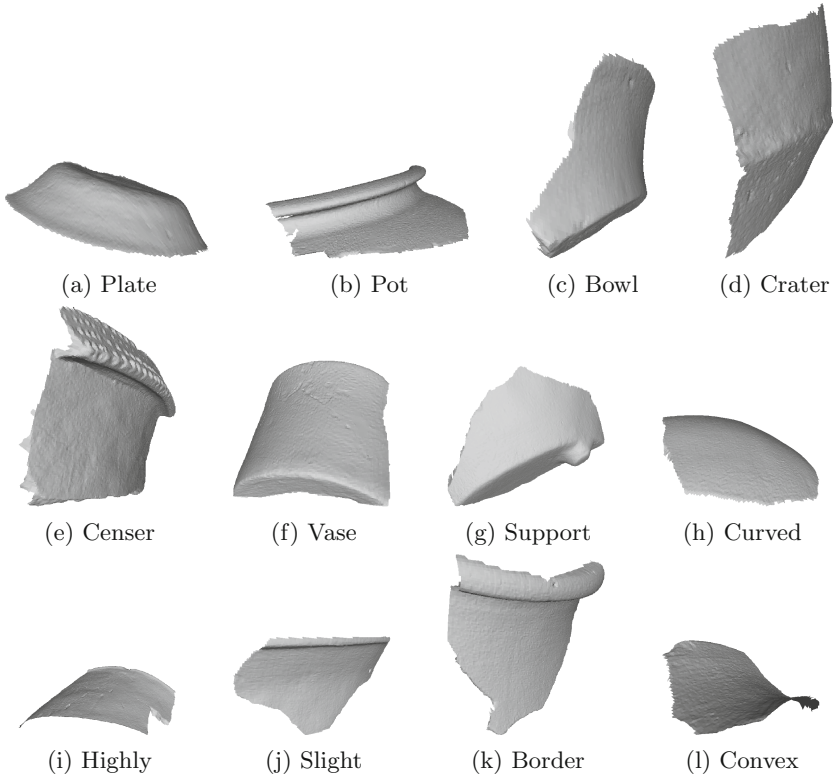
As previously mentioned, we used a set of 3D surfaces, which correspond to potsherds from the Teotihuacan culture that developed in ancient Mesoamerica. An important fact about this specific collection is that it has been already cataloged manually by archaeologist, thus it was possible to annotate all its instances with the name of the ceramic type of their provenience, e.g., plate, pot, crater, bowl, vase, amphora, etc. Note however, that these labels are recognizable only for the so-called diagnostic potsherds, this is, potsherds containing specific parts of the ceramic type, e.g., the neck of a pot, the border of a jar, the supports of a base, etc.

Besides those diagnostic potsherds, there also exist potsherds that are very difficult to recognize. Most of these examples come from sections of the main body of the ceramics, and they often correspond to simple curved or flat sections with no much of discriminative visual information. Figure 1 shows examples of both diagnostic and non-diagnostic potsherds.

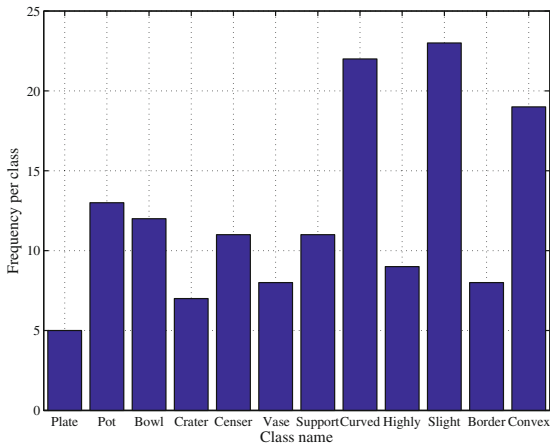
As shown in Fig. 1, the annotations for the diagnostic potsherds consist in the name of their respective ceramic type: plate, pot, bowl, crater, censer, vase, and vase with support. Regarding the non-diagnostic potsherds (or regular potsherds), their annotations are as follows:

- Curved: potsherds that have moderately curved shapes.
- Highly (curved): potsherds with highly visible curved shapes.
- Slight (border): sections towards the ceramics border, that are barely visible.
- Border: potsherds with clearly visible border.
- Convex: different from the previous concave curved potsherds, this class consists of potsherds whose curvature is convex.

Overall, this dataset is composed of 148 surfaces and 12 visual classes. Figure 2 shows the distribution of instances over them. Note that although the classes are not well balanced, they remain within the same order of magnitude.



**Fig. 1.** Visual examples of potsherds. Examples 1a to 1g correspond to diagnostic potsherds, and examples 1h to 1l are of non-diagnostic potsherds.



**Fig. 2.** Frequency of potsherds in each class.

## 4.2 Consistency and Efficiency

We first compared the time that different methods require for computing local descriptors, including Scale-Invariant Spin Images (SISI) [4], 3D Shape Context (3DSC) [6] with  $L = 8$ , Unique Shape Context for 3D data (USC) [7], and Unique Signatures of Histograms for Local Surface Description (SHOT) [3].

To check for consistency of the proposed method, we also evaluated the level of degradation that a local descriptor suffers after a rotation change. To this end, we randomly selected 1000 points from all the dataset (different 3D surfaces randomly selected), and computed their local descriptors before and after rotating the 3D surface. Then, we computed the Euclidean distance between the two instances of the same descriptor (before and after rotation).

## 4.3 Classification Performance

Finally, we compared the classification performance of each method using the dataset of potsherds. To this end, we relied on bag-of-words representations. Namely, we repeated the following protocol independently for each method:

1. Compute a vocabulary of local descriptors using a subset of 20,000 of them randomly selected. In practice, we computed vocabularies of different sizes using the k-means clustering algorithm [10].
2. Represent each 3D surface with a bag representation using the vocabulary previously estimated.
3. Use each 3D surface as query to be classified using a k-NN approach (k=1). This is, a leave-one-out full-cross validation.
4. Compute the average classification accuracy.

# 5 Results

This section presents the results of our evaluations. First the evaluation of efficiency and consistency, and later the evaluation on classification performance.

## 5.1 Consistency

Table 1 shows the average time that each method takes to compute a local descriptor alongside its dimensionality. As one can see, both SISI and the proposed method, which we refer to as RI, are the fastest methods for computation of local shape descriptors, as they are also the shortest. Note that the use of the covariance matrix and computation of principal eigenvectors (USC and SHOT) requires more time than the proposed subtraction of local orientations (RI). Special attention is worth to 3DSC, which do not relies on the computation of principal eigenvectors, however it does repeat the description process 8 times.

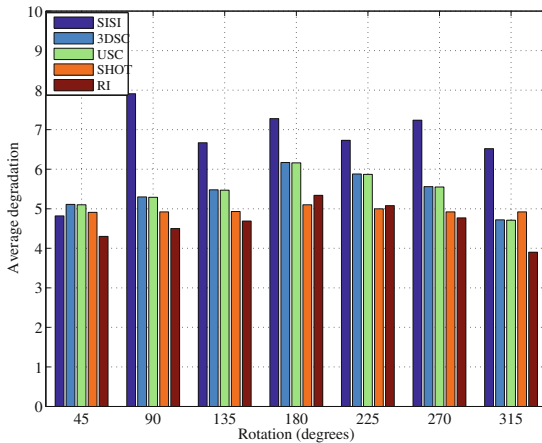
We compared the degradation induced by our proposed method with respect to that induced by the different previous methods. Figure 3 shows these results. As shown in Fig. 3, almost all descriptors induce comparable levels of distortion

**Table 1.** Vector size and computational time in seconds for each method.

Method	Vector size	Time(s)
SISI [4]	64	$0.91 \pm 1.29$
3DSC [6]	160 (x8)	$33.06 \pm 6.29$
USC [7]	160	$6.03 \pm 8.08$
SHOT [3]	160	$3.12 \pm 4.28$
RI (ours)	64	$0.93 \pm 1.20$

after rotation transformations, with SISI been the method with highest degradation, as it lacks of a proper definition of orientation frame.

Although the use of eigenvectors induces relative low degradation (USC, SHOT), it is not as low as that induced by the proposed RI. This suggests that the computation of principal eigenvectors is not as strong approach as expected for addressing rotation variations. Also, the results of Fig. 3 shows that the proposed subtraction technique works well in practice.



**Fig. 3.** Euclidean distance computed between two descriptors corresponding to the same point localized at rotated instances of the same 3D model. Rotation in steps of 45 degrees

## 5.2 Classification Performance

We considered two cases for the classifications experiments. The first one, when the complete dataset is used, and the second case when only the subset of diagnostic potsherds is considered. Table 2 shows the average classification accuracy obtained by the different methods for local description.



**Table 2.** Average classification accuracy achieved by different 3D local descriptor. These results are for two cases: (1) using the complete dataset, and (2) using only diagnostic potsherds. Best results are highlighted in bold.

Method	Vocabulary size					
	100	250	500	1000	2500	5000
Complete dataset						
SISI [4]	0.295	0.285	0.266	0.261	0.217	0.193
3DSC [6]	0.301	0.304	0.278	0.255	0.246	0.244
USC [7]	<b>0.481</b>	0.480	0.458	0.426	0.406	0.367
SHOT [3]	0.466	0.469	0.465	0.460	0.425	0.405
RI (ours)	0.470	0.472	0.447	0.385	0.386	0.374
Only diagnostic potsherds						
SISI [4]	0.362	0.404	0.362	0.319	0.255	0.255
3DSC [6]	0.448	0.492	0.480	0.441	0.416	0.412
USC [7]	0.647	0.652	0.641	0.606	0.588	0.581
SHOT [3]	0.658	<b>0.668</b>	0.615	0.639	0.592	0.577
RI (ours)	0.647	0.653	0.622	0.591	0.598	0.579

From Table 2, one can see that higher classification accuracy is achieved by using only diagnostic potsherds. This is expected as it corresponds to a more controlled scenario with respect to using both diagnostic and generic potsherds. Furthermore, although the ceramic pieces from which generic potsherds come are virtually unknown, in practice they will correspond to fragments coming from the same ceramic pieces than diagnostic potsherds. Therefore, they might have sections that visually resemble to one another, which in turn, might lead to confusion during classification. We can also see that short vocabularies suffice for accurate description. In particular, 250 visual words are enough.

Regarding the comparison of methods, SISI and 3DSC achieved the lowest performance. Whereas USC and RI, which are very similar descriptors by construction, performed alike each other. Although USC and RI are partially similar to 3DSC by construction, 3DSC requires the computation of 8 several variants of the local descriptor, and it performs poorer than USC and RI. Note that SHOT achieved a slightly better performance than RI. However, it produces a larger vector, and it takes as much as three times to compute it compared with RI.

## 6 Conclusions

We presented a new methodology for achieving rotation invariance on 3D local shape descriptor. Besides its simplicity, the proposed method has high levels of invariance against rotation transformations. Also, it produces a short vector that achieves state-of-the-art performance in classification results, with respect to several previous methods for local shape description of 3D models.

One main feature of the proposed model is that it is independent from the actual function used as descriptor, such that it can be plugged to many local descriptors to boost their robustness against rotation transformations.

We evaluated our method on a set of 3D surfaces that represent archaeological potsherds, and obtained high performance in the classification task. Namely, classifying potsherds is of great importance to archaeologist, and one of the many tasks where pattern analysis could provide with new tools. In particular, this is the case of our research effort, which seeks to develop a machinery for presenting classification suggestions for newly discovered potsherds, such that it could assist archaeologists in deciding classes for potsherds.

Currently, we continue working on the collection of more data for further testing of our method, and the eventual implementation of such system that could handle not only potsherds, but different archaeological artifacts, e.g., masks, ceramics, jewelry, etc.

**Acknowledgments.** This work was supported by Swiss National Science Foundation through the Tepalcatl project P2ELP2.152166.

## References

1. Li, B., Lu, Y., Li, C., Godil, A., Schreck, T., Aono, M., Burtscher, M., Chen, Q., Chowdhury, N.K., Fang, B., Fu, H., Furuya, T., Li, H., Liu, J., Johan, H., Kosaka, R., Koyanagi, H., Ohbuchi, R., Tatsuma, A., Wan, Y., Zhang, C., Zou, C.: A comparison of 3D shape retrieval methods based on a large-scale benchmark supporting multimodal queries. *Comp. Vis. Image Underst.* **131**, 1–27 (2015)
2. Li, B., Lu, Y., Li, C., Godil, A., Schreck, T., Aono, M., Chen, Q., Chowdhury, N.K., Fang, B., Furuya, T., Johan, H., Kosaka, R., Koyanagi, H., Ohbuchi, R., Tatsuma, A.: SHREC'14 track: large scale comprehensive 3D shape retrieval. In: *Eurographics Workshop on 3D Object Retrieval 2014 (3DOR)* (2014)
3. Tombari, F., Salti, S., Di Stefano, L.: Unique signatures of histograms for local surface description. In: Maragos, P., Paragios, N., Daniilidis, K. (eds.) *ECCV 2010, Part III*. LNCS, vol. 6313, pp. 356–369. Springer, Heidelberg (2010)
4. Darom, T., Keller, Y.: Scale-invariant features for 3-D mesh models. *IEEE Trans. Image Proc.* **21**(5), 2758–2769 (2012)
5. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. *Int. J. Comp. Vis.* **60**(2), 91–110 (2004)
6. Frome, A., Huber, D., Kolluri, R., Bülow, T., Malik, J.: Recognizing objects in range data using regional point descriptors. In: Pajdla, T., Matas, J.G. (eds.) *ECCV 2004*. LNCS, vol. 3023, pp. 224–237. Springer, Heidelberg (2004)
7. Tombari, F., Salti, S., Di Stefano, L.: Unique shape context for 3D data description. In: *Proceedings of the ACM Workshop on 3D Object Retrieval* (2010)
8. Belongie, S., Malik, J., Puzicha, J.: Shape matching and object recognition using shape contexts. *IEEE Trans. Pattern Anal. Mach. Intell.* **24**(4), 509–522 (2002)
9. Johnson, A.E., Hebert, M.: Using spin images for efficient object recognition in cluttered 3D scenes. *IEEE Trans. Pattern Anal. Mach. Intell.* **21**(5), 433–449 (1999)
10. Lloyd, S.: Least squares quantization in PCM. *IEEE Trans. Inf. Theory* **28**(2), 129–137 (1982)