

Gait-Based Carried Object Detection Using Persistent Homology

Javier Lamar-Leon¹, Raul Alonso Baryolo¹,
Edel Garcia-Reyes¹, and Rocio Gonzalez-Diaz²

¹ Patterns Recognition Department, Advanced Technologies Application Center,
7a # 21406 e/ 214 y 216, Rpto. Siboney, Playa, C.P. 12200, La Habana, Cuba

² Applied Math Dept., School of Computer Engineering,
Campus Reina Mercedes, University of Seville, Seville, Spain
{jllamar,rbaryolo,egarcia}@cenatav.co.cu, rogodia@us.es

Abstract. There are surveillance scenarios where it is important to emit an alarm when a person carrying an object is detected. In order to detect when a person is carrying an object, we build models of naturally-walking and object-carrying persons using topological features. First, a stack of human silhouettes, extracted by background subtraction and thresholding, are glued through their gravity centers, forming a 3D digital image I . Second, different *filters* (i.e. orderings of the cells) are applied on $\partial K(I)$ (cubical complex obtained from I) which capture relations among the parts of the human body when walking. Finally, a *topological signature* is extracted from the persistence diagrams according to each filter. We build some clusters of persons walking naturally, without carrying object and some clusters of persons carrying bags. We obtain vector prototypes for each cluster. Simple distances to the means are calculated for detecting the presence of carrying object. The measure cosine is used to give a similarity value between topological signatures. The accuracies obtained are 95.7% and 95.9% for naturally-walking and object-carrying respectively.

Keywords: gait-based recognition, topology, persistent homology, carrying object detection.

1 Introduction

The functionality of detecting a person carrying an object, combined with the capabilities of tracking and person re-identification, can be used to discover the owner of an abandoned luggage or to detect the interchange of objects between two persons. This problem is very difficult to be tackled due to the great variety of bags, suitcases, backpacks, handbags, and so on. There are variations in size, shape, weight and different ways in which they can be carried. There are two approaches reported in the literature to face the carried object detection. One aims to localize the object and its possible classification, the other one aims to determine whether a person is carrying an object. Both may be combined in order to increase the efficacy and provide more information to the operator.

The first intent to tackle this problem took into account the symmetry of the human body that is modified when a protuberance appears in the human silhouette due to a carried object [1]. Other cue used by authors is the periodicity in the moving of different body parts and their relationship through the time [2–4]. Approaches based on gait are an alternative to the localization approach when there are self-occlusions that prevent the object visualization or when the difference between the object’s color and the clothes’ color is very small to obtain correct segmentation [5]. Some papers describe a methodology based on segmenting the human image using different cues and classifying the candidate regions in order to detect the presence of a carried object [6]. Neural Network and Support Vector Machine are the classifiers reported for carried object detection.

Homology is a topological invariant frequently used in practice [7, 8]. The ranks of the homology groups, also called Betti numbers, coincide, respectively, with the number of connected components, tunnels and cavities of the object. In particular, the homology could be a robust representation because the shape of connected components and holes may change under geometric transformations, but the number of components and holes will be more stable.

In this paper, we follow the gait based approach without localizing the carried object. We build models of naturally-walking and object-carrying using persistent homology groups. These features have been previously used for gait identification and gender classification [9, 10]. Our methodology is evaluated using the CASIA-B gait database ¹, which consists of 738 video sequences of persons walking naturally, without carrying object and 246 video sequences of persons walking with different kinds of bags. Given that there are no precedents in the usage of that database for the task of carried object detection and the absence of an established experimental protocol, we don’t show comparisons with other works.

The rest of the paper is organized as follows. Section 2 is devoted to recall topological signature for gait description. Section 3 shows the method proposed. Experimental results are reported in Section 4. We conclude this paper and discuss some future work in Section 5.

2 Recall of Topological Signature for Gait Description

In this section, we explain how to obtain the topological signature of a person walking. According to a view angle defined on a cubical complex $\partial K(I)$, an ordering (filter) of the cells is obtained. Persistent homology obtained from each filter is represented in persistence diagrams format [11]. Finally, a topological signature is extracted from those diagrams.

2.1 The Cubical Complex $\partial K(I)$

In order to increase the efficiency of our algorithms, in this paper we introduce the usage of cubical complexes. This new approach allows to decrease the number

¹ <http://www.cbsr.ia.ac.cn/english/GaitDatabases.asp1>

of maximal cells to the half respect to the usually applied simplicial complexes [9, 10], achieving a much better efficiency without sacrificing the efficacy.

To start, the foreground (person) is segmented from the background by applying background modeling and thresholding. The sequence of resulting silhouettes is analyzed to extract one *subsequence of representation*, which include at least a gait cycle [12].

The 3D binary digital picture $I = (\mathbb{Z}^3, B)$ (where $B \subset \mathbb{Z}^3$ is the foreground), is built by stacking silhouettes of a subsequence of representation, aligned by their gravity centers (gc)[9, 10]. See Fig. 1.a. The 3D cubical complex $Q(I)$ associated to I contains the unit cubes with vertices $V = \{(i, j, k), (i + 1, j, k), (i, j + 1, k), (i, j, k + 1), (i + 1, j + 1, k), (i + 1, j, k + 1), (i, j + 1, k + 1), (i + 1, j + 1, k + 1)\}$ and all its faces (vertices, edges and squares) iff $V \subseteq B$.

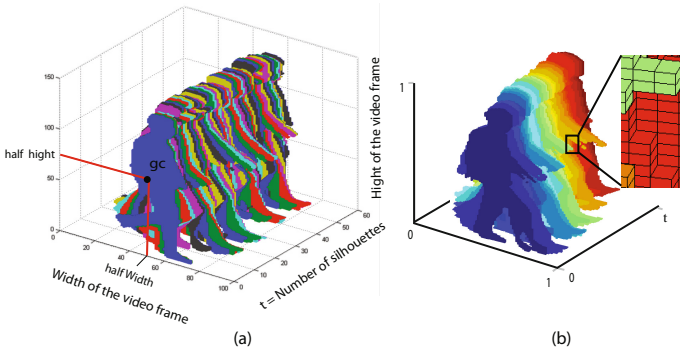


Fig. 1. (a) Silhouettes aligned by their gc forming $I = (\mathbb{Z}^3, B)$. (c) The border cubical complex $\partial K(I)$.

The squares of $Q(I)$ that are faces of exactly one cube in $Q(I)$ and their faces (vertices and edges) make up the *border cubical complex* $\partial K(I)$ (see Fig. 1.c). Finally, coordinates of the vertices of $\partial K(I)$ are *normalized* to coordinates (x, y, t) , where $0 \leq x, y \leq 1$ and t is the number of silhouette of the subsequence of representation.

2.2 Filters for $\partial K(I)$

In this subsection we present how to define the filters on $\partial K(I)$. These filters capture relations among the parts of the human body when walking.

When a view direction d is chosen, two filters for $\partial K(I)$ are obtained as follows. All the cells belonging to $\partial K(I)$ are associated with two *filtering functions* f_+ and f_- . For each vertex $v \in \partial K(I)$, $f_+(v)$ is the distance between v and the plane normal to d passing through the origin of the reference frame, while $f_-(v) = -f_+(v)$. Edges and squares are associated to the largest value that f_+ (resp. f_-) assumes on their vertices. Being the cells of $\partial K(I)$ finite in number, we can determine a minimum value for f_+ , say f_{\min} , and a maximum one,

f_{\max} . It is now possible to induce two filters on $\partial K(I)$ by ordering its cells according to increasing values of f_+ and f_- , respectively. Denote these filters by $K_{[f_{\min}, f_{\max}]} = \{\sigma_1, \dots, \sigma_k\}$ and $K_{[-f_{\max}, -f_{\min}]} = \{\sigma'_1, \dots, \sigma'_k\}$.

2.3 Persistence Diagrams and Topological Signatures

Given a cubical complex K and a filter $(f(\sigma)) \{\sigma_1, \dots, \sigma_k\}$ for K , if σ_i completes a p -cycle (p is the dimension of σ_i) when σ_i is added to $K_{i-1} = \{\sigma_1, \dots, \sigma_{i-1}\}$, then a p -homology class γ is born at time i ; otherwise, a $(p-1)$ -homology class dies at time i . The difference between the birth and death time of a homology class is called its persistence, which quantifies the significance of a topological attribute. If γ never dies, we set its persistence to infinity. Drawing a point (i, j) for an homology class that is born at time i and dies at time j , we get the p -persistence diagram of the filtration, denoted as $Dgm(f)$. It represents a p -homology class by a point whose vertical distance to the diagonal is the persistence. Since always $i < j$, all points lie above the diagonal (see [11]).

In this paper, persistence diagrams are first computed for $K_{[f_{\min}, f_{\max}]}$ and $K_{[-f_{\max}, -f_{\min}]}$. Then, the diagrams are explored according to a uniform sampling. More precisely, given a positive integer n , compute the integer $h = \lfloor \frac{k}{n} \rfloor$ representing the width of the ‘‘window’’ we use to analyze the persistence diagram. Indeed, for $i = 0, \dots, (n-1)$, the p -persistence diagram of $K_{[f_{\min}, f_{\max}]}$ (resp. $K_{[-f_{\max}, -f_{\min}]}$) show:

- (a) Number of homology classes that are born after $i \cdot h$ and before $(i+1) \cdot h$.
- (b) Number of homology classes that are born before $i \cdot h$ and die after $i \cdot h$.

A vector of $2n$ entries is then formed containing (a) in entry $2i$ and (b) in $2i+1$; this way we obtain for a given dimension p , a vector for $K_{[f_{\min}, f_{\max}]}$ (resp. $K_{[-f_{\max}, -f_{\min}]}$).

The topological signature for a gait subsequence considering a fixed direction of view consists in four $2n$ -dimensional vectors: $V = \{V_1, \dots, V_4\}$ constructed as explained above. Consider that we take into account two dimensions ($p = 0$ and $p = 1$) and two filters ($K_{[f_{\min}, f_{\max}]}$ and $K_{[-f_{\max}, -f_{\min}]}$).

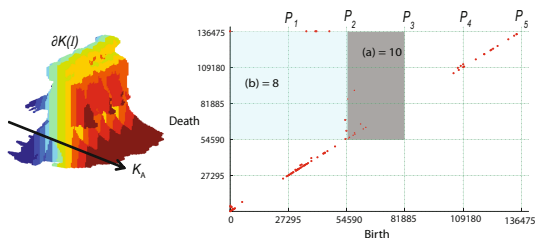


Fig. 2. An example of computation of the first element of a topological signature

For example, consider K_A given in Fig. 2 which consists in $k = 136475$ cells. We perform $n = 5$ uniform cuts on the 0-persistence diagram, then $h = 27295$.

According to $i = 3$, the number of the life-lengths for the 0-homology classes that are born after $(i - 1) \cdot h = 54590$ and before $i \cdot h = 81885$ is $(a) = 10$, and those that are born before $(i - 1) \cdot h$ and die after $(i - 1) \cdot h$ is $(b) = 8$.

2.4 Comparing Topological Signatures

The topological signatures for two gait subsequences associated with a fixed view direction, say $V = \{V_1, \dots, V_4\}$ and $W = \{W_1, \dots, W_4\}$, can be compared according to the following procedure: for every $i = \{1, \dots, 4\}$ compute:

$$S_i = \frac{V_i \cdot W_i}{\|V_i\| \cdot \|W_i\|}. \quad (1)$$

Which is the cosine of the angle between the vectors V_i and W_i . Observe that $0 \leq S_i \leq 1$ since the entries of both vectors are always non-negative. Then, the total similarity value for two gait subsequences, O_1 and O_2 , considering a fixed view direction, is the sum of the 4 similarity measures computed before:

$$S(O_1, O_2) = S_1 + S_2 + S_3 + S_4. \quad (2)$$

3 Carried Object Detection Using Topological Signature

We cluster the persons according to the topological signature explained above. Each cluster will be associated to a common gait pattern shared by a set of persons. There is a wide range of objects that can be carried by a person and different ways in which they can be carried. This modifies the dynamic and shape patterns in the way of walk. Even for natural walking without carrying object there is intra-class variation. The set of samples should be subdivided in order to capture the multimodality of the probabilistic distribution, for both classes of naturally-walking and carrying-object. We propose to use $k - means$ to obtain a structural partition on a set of topological signatures. The cosine measure is used as distance in the clustering algorithm. The number of clusters depends on the scenario and the kind of objects expected. We obtain a prototype vector for each cluster (the mean). The classification step uses simple distances to the means for detecting the presence of carried object.

3.1 Filtration

We use 4 directions of view to filter the cubical complex: The first one is horizontal(axis X), the second one is vertical(axis Y), the third one forms 45 degrees with the X and Y and is orthogonal to the axis T . The fourth direction of view is orthogonal to the axis T and to the third direction of view. See Fig. 3. In our experiment, the number of cuts in each direction of view is $n = 23$.

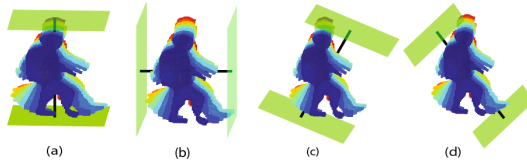


Fig. 3. Directions of view for filtering the cubical complex

3.2 Clustering

According to the surveillance scenario (airports, train-station, metro, etc.) we can estimate the kind of objects carried by people. For instance, in airports is common to see suitcases. We visually examined the CASIA-B database to determine the kinds of carried object. See Figure 4.

Also, it is important to take into account the different ways in which the objects are carried. We assume 15 clusters to obtain a multimodal representation for carrying-walk model. On the other hand, it is difficult to estimate a number of clustering in the set of topological signatures from natural walk. Therefore, we use 15 clusters too. There are different ways to select the prototypes for each model, in this case we selected the mean vectors.



Fig. 4. Some kinds of carried object in CASIA-B database

4 Experimental Results

Currently, researchers are not following any established experimental protocol to compare the results. In this paper we aim to introduce the CASIA-B database for this kind of evaluation. This database contains 124 subjects, 93 men and 31 women. In our experiments the person 005 (a man) was removed due to a bad segmentation. There are 6 video sequences for each naturally-walking person and 2 video sequences for each carrying-walking person. Furthermore, each sequence is captured from 11 view angles. In this paper we use the lateral view. We analyze 984 video sequences (738 naturally-walking and 246 carrying-walking).

4.1 Experiment 1

In this experiment the CASIA-B database is divided in training and testing sets. To train we selected 50 persons and to test 73 persons. We used 300 video sequences of natural walk ($50 * 6 = 300$) in the clustering process, obtaining a model with 15 means. On the other hand, we used 100 video sequences of carrying walk ($50 * 2 = 100$), obtaining another model with 15 means.

Table 1. Confusion matrix : 592 video sequences in the test set from CASIA-B

	Test Sequences	Naturally-Walking	Carrying-Walking	Avg
Naturally-Walking	438	419	19	95.7
Carrying-Walking	146	6	140	95.9
Global accuracy	584			95.7



Fig. 5. Persons bad identified carrying baggage

Table 2. Confusion matrix : 984 video sequences from the CASIA-B database. Leave one out mode

	Test Sequences	Naturally-Walking	Carrying-Walking	Avg
Naturally-Walking	738	712	26	96.5
Carrying-Walking	246	18	228	92.7
Global accuracy	984			95.5

For testing, we used 438 video sequences of natural walk ($73 * 6 = 438$) and 146 video sequences of carrying walk ($73 * 2 = 146$). The topological feature for each test video sequence was evaluated against 30 mean vectors. The results obtained are showed in Table 1. In this experiment the false alarm rate was 4.28%.

The six false negatives (persons 055 sample 1 and 2, 064, 074, 104 and 114) for the class carrying-walking are showed in Figure 5. Persons 055 and 064 carry backpacks that almost don't create protuberance that changes the shape pattern and they don't affect the arms moving. In the case of persons 074, 104 and 114 the handbags weights don't affect their normal-walking pattern and the bags are almost occluded by the body.

4.2 Experiment 2

In this experiment we used the topological features of the 123 persons in the CASIA-B database in a cross validation mode (leave one out). In the same way that in the previous experiment, we used 15 means for naturally-walking and 15 means for carrying-walking. Only one of these 123 sets was used to test. The remaining 122 sets were used for training. The average results for the 123 iteration is showed in Table2. In this experiment the false alarm rate is 4.5%.

5 Conclusion and Future Work

In this paper, a representation based on topological features, previously used for gait based human identification at a distance, is used for carried object detection task, with encouraging results of 95.9% in the first experiment and 92.7% in the second one. Our future work consists in trying to improve our results for cross view angles problems. On the other hand, the combination between this non-localization approach and the possibility of localizing the carried object is an interesting research line.

References

1. Haritaoglu, I., Cutler, R., Harwood, D., Davis, L.S.: Backpack: Detection of people carrying objects using silhouettes. In: *The Proceedings of the Seventh IEEE International Conference on Computer Vision*, vol. 1, pp. 102–107. IEEE (1999)
2. Tavanai, A., Sridhar, M., Gu, F., Cohn, A.G., Hogg, D.C.: Carried object detection and tracking using geometric shape models and spatio-temporal consistency. In: Chen, M., Leibe, B., Neumann, B. (eds.) *ICVS 2013. LNCS*, vol. 7963, pp. 223–233. Springer, Heidelberg (2013)
3. Senst, T., Kuhn, A., Theisel, H., Sikora, T.: Detecting people carrying objects utilizing lagrangian dynamics. In: *2012 IEEE Ninth International Conference on Advanced Video and Signal-Based Surveillance (AVSS)*, pp. 398–403. IEEE (2012)
4. Senst, T., Evangelio, R.H., Eiselein, V., Pätzold, M., Sikora, T.: Towards detecting people carrying objects—a periodicity dependency pattern approach. In: *VISAPP (2)*, vol. 10, pp. 524–529 (2010)
5. Tao, D., Li, X., Maybank, S.J., Wu, X.: Human carrying status in visual surveillance. In: *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2, pp. 1670–1677. IEEE (2006)
6. Dondera, R., Morariu, V., Davis, L.: Learning to detect carried objects with minimal supervision. In: *2013 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 759–766. IEEE (2013)
7. Zomorodian, A., Carlsson, G.: Localized homology. *Comput. Geom.* 41(3), 126–148 (2008)
8. Zomorodian, A.: Computational topology. In: Atallah, M., Blanton, M. (eds.) *Algorithms and Theory of Computation Handbook*, 2nd edn., vol. 2, Chapman & Hall/CRC Press, Boca Raton (2010)
9. Nixon, M.S., Carter, J.N.: Automatic recognition by gait. *Proc. of IEEE* 94(11), 2013–2024 (2006)
10. Leon, J.L., Cerri, A., Reyes, E.G., Diaz, R.G.: Gait-based gender classification using persistent homology. In: Ruiz-Shulcloper, J., Sanniti di Baja, G. (eds.) *CIARP 2013, Part II. LNCS*, vol. 8259, pp. 366–373. Springer, Heidelberg (2013)
11. Edelsbrunner, H., Harer, J.: *Computational Topology - an Introduction*. American Mathematical Society (2010)
12. Nixon, M.S., Carter, J.N.: Automatic recognition by gait. *Proc. of IEEE* 94(11), 2013–2024 (2006)