Pattern Recognition 48 (2015) 1074–1085

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

Salient feature based graph matching for person re-identification

Sara Iodice [1], Alfredo Petrosino *,[1]

Department of Science and Technology, University of Naples Parthenope, Centro Direzionale Is C4, 80143 Naples, Italy

ARTICLE INFO

Article history:
Received 2 January 2014
Received in revised form
12 September 2014
Accepted 15 September 2014
Available online 4 November 2014

Keywords:
Computational symmetry
Salient features
Graph representation and matching
People re-identification

ABSTRACT

We propose a person re-identification non-learning based approach that uses symmetry principles, as well as structural relations among salient features. The idea comes from the consideration that local symmetries, at different scales, also enforced by texture features, are potentially more invariant to large appearance changes than lower-level features such as SIFT, ASIFT. Finally, we formulate the re-identification problem as a graph matching problem, where each person is represented by a graph aimed not only at rejecting erroneous matches but also at selecting additional useful ones.

Experimental results on public dataset i-LIDS provide good performance compared to state-of-the-art results.

1. Introduction

Symmetry detection is highly relevant in pattern recognition. Indeed, the description of a figure may be different when it is embedded in a context with horizontal or vertical symmetry [19]. Besides, in tasks requiring the completion of partially occluded visual stimuli, subjects tend to produce systematically symmetrical [14]. The concept of symmetry is not univocal: various kinds of properties of an image are defined as symmetry [30,28]. As instance, a figure has *rotational symmetry* when it can be rotated less than 360° around its central point, or axis, and still matches the original figure.

This cue is peculiar in person re-identification where the problem consists in recognizing people in different poses from images coming from distinct cameras. This is an important task in the video surveillance, where large and structured environments must be supervised (such as airport, metro, station or shopping centres) and it becomes more critical when the cardinality of gallery set increases.

Symmetry is adopted in [6], a person re-identification method that weighs appearance information, extracted from different body parts, in accordance with their distance from symmetry axes computed on the whole figure. Therefore, symmetry appears to have been used as a global property on this previous work, not as a local one. Conversely, symmetry is adopted as both global and local property in our approach. In particular, the global symmetry axis is exploited to select salient locations representing each pedestrian, while local symmetry is adopted like local feature in order to describe each salient location.

Researchers in computer vision have made significant progress in representing and detecting symmetries in images and other types of data [15].

However, there has been relatively little work on using local symmetries as explicit features for matching tasks [2], and no work about matching the same pedestrian with different poses, like in re-identification. The main idea is to use a variety of symmetries, rather than repetitions, together with texture as cues, in order to define a complex feature detector, based on the consideration that, at different scales, the symmetry is potentially more invariant to large appearance changes than lower-level features such as SIFT, and, when combined with a texton-based feature, is highly discriminative [11].

In this context, the main contribution resides in the image features and the adopted feature organization for matching. The discriminative power could be more appreciated if the features are organized in an appropriate manner. Most of the researchers focus their attention on the features extracted in many ways from the scene. The Harris and SIFT [16] features are important to identify what is distinctive and discriminative for the purpose of a correct recognition of the scene. The bag of words algorithm has been applied to SIFT descriptors, to identify discriminative combinations of descriptors [4]. In [18] the application of a clustering to descriptors leads to results which are less distinctive in a large cluster than those in a small cluster. For example in indoor navigation, window corners are common, so they are not good features to uniquely identify scenes, while corners found on posters or signs are much better. In [7] an effective approach based on real-time loop detection has proved to be efficient using a hand-held camera, through SIFT features and intensity and hue

* Corresponding author.
E-mail addresses: iodice.uniparthenope@gmail.com (S. Iodice), petrosino@uniparthenope.it (A. Petrosino).
[1] Tel./fax: +39 0815476656.

http://dx.doi.org/10.1016/j.patcog.2014.09.011

histograms combined using a bag of words approach. Up to now none of the existing approaches tackled the relations among features in terms of similarity and spatial homogeneity.

Our main contribution consists in introducing an approach based on graph-based representation, according to which regions with their corresponding feature vector and the geometric relationship between these regions are encoded in the form of a graph. According to the idea that an image can be described at the higher level in terms of a nested hierarchy of local symmetries, we present a novel graph matching approach to the problem aimed at evolving an initial set of correspondences computed with the local features, as a kind of compromise between the constraints imposed by both the local features and the structural relations.

The vertices are the image key points detected by the SIFT (specifically we adopt ASIFT [27]), enriched by features based on color, texture, and local symmetry. The cost of an edge joining two vertices represents a measure of their dissimilarity. Therefore, the problem of person re-identification is formulated as a graph matching problem. Our approach is an appearance-based method which differs from the state-of-the-art:

(i) unlike [3,21] we give great weight to local features;
(ii) we do not adopt spatio-temporal information such as [17,13];
(iii) our method is non-learning based one, unlike [32,1,22].

The paper is organized as follows. Section 2 deals with pedestrian representation. Section 3 provides details about how pedestrian representations are compared each other in order to solve re-identification task. Section 4 reports testing results on public dataset ILIDS and finally in Section 5 some conclusions are drawn.

2. Pedestrian representation

The main steps to represent a pedestrian for recognition purposes using visual features can be listed as below:

1. Candidate keypoints extraction
2. Keypoints filtering by symmetry axis
3. Local feature extraction
4. Graph representation

First of all, candidate keypoints are extracted from each pedestrian image, filtered according to their distance from the symmetry axis and then local features are computed on each selected location. Finally, each pedestrian image is represented by a graph, in order to consider structural relations.

2.1. Keypoints filtering by symmetry axis

ASIFT locations [27], fully invariant with respect to zoom, rotation, translation, and the angles defining the camera and axis orientation, are extracted from each frame of the dataset as candidate keypoints.

The property of ASIFT features to be invariant to affine transformations is very attractive in person re-identification problem, where pedestrian figures, subject to viewpoints and pose variations, need to be represented for recognition aim.

On the other hand, the property of two pixels, belonging to a certain region, to be symmetric, with respect to a particular axis, remains unchanged when an affine transformation is applied to the same region. Thus, reflective symmetry should be considered invariant to affine transformations. We adopt our multi-scale symmetry detection algorithm [10] to extract the symmetry axis corresponding to the region and the direction with maximal correlation. Therefore, under the hypothesis that there are no high pose variations, the symmetry axis detected by algorithm still remains unchanged.

As instance, as shown in Fig. 1, if two pixels p_1 and p_2 belonging to the pedestrian in the frame A are symmetric with respect to the reflective symmetric axis RFS_a and are located at a distance d from RFS_a, they continue to be exactly symmetric also in the frame B, where the pedestrian changes his pose and is located at the distance $d' \approx d$.

For this reason, filtering keypoints ASIFT located at a distance $x \leq d$, the probability to keep matched keypoints clearly increases and the number of keypoints, candidate for the graph representation, decreases. Another main aspect is that candidate keypoints, with respect to the distance d from the symmetry axis, are roughly located within the foreground and, consequently, background subtraction for figure/background separation could be avoided.

2.2. Keypoints distribution around symmetry axis

To extract information about the density distribution of candidate keypoints around the symmetry axis, the distances d_j of the generic keypoint k_j from the reflection symmetry axis RFS_i of the region (head, trunk and legs) to which k_j belongs, are computed for each frame $i = 1, \ldots, n$.

For the generic distance d we estimate

- The probability density $p(d)$ that describes the probability of ASIFT keypoints to be located exactly at a distance d from the RFS reflection symmetry axis.
- The cumulative distribution $P(d)$ that describes the probability of ASIFT keypoints to be located at a distance $x \leq d$ from RFS.

The plots shown in Fig. 2 are built considering all the frame of ILDS MCTS public dataset. Moreover, the cumulative distribution accounts for 58.29% for $d = 12$, underlining that a wide set of ASIFT keypoints is located closer to the reflection symmetry axis. On the other hand, from the plot in Fig. 2a, it can be observed that the probability density decreases for values of $d \geq 3$. Therefore, ASIFT keypoints, located at a greater distance from the reflection symmetry axis, are clearly less informative and should be

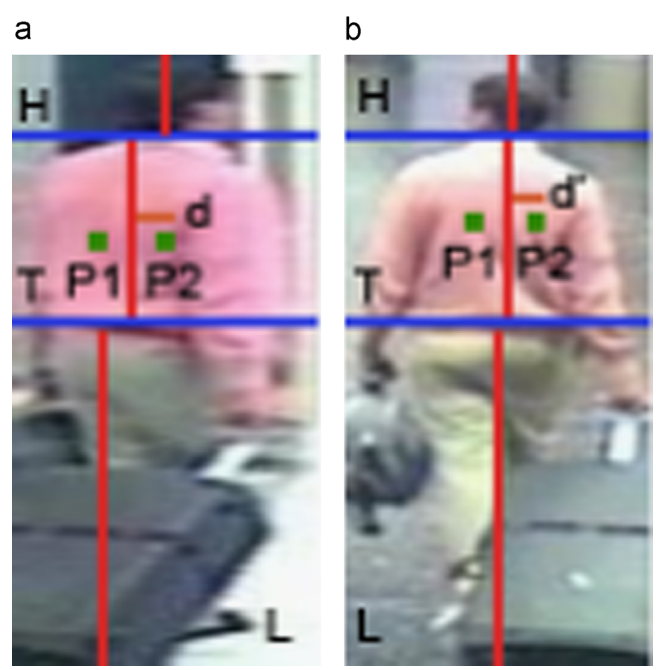

Fig. 1. (a) Frame A, (b) Frame B.