Pattern Recognition 44 (2011) 454–470

Contents lists available at ScienceDirect

Pattern Recognition

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

Enhanced Local Subspace Affinity for feature-based motion segmentation

L. Zappella [a,*], X. Lladó [a], E. Provenzi [b], J. Salvi [a]

[a] Institut d'Informàtica i Aplicacions, Universitat de Girona, Girona, Spain
[b] Departamento de Tecnologías de la Información y las Comunicaciones, Universitat Pompeu Fabra, Barcelona, Spain

ARTICLE INFO

Article history:
Received 9 July 2009
Received in revised form
22 July 2010
Accepted 10 August 2010

Keywords:
Motion segmentation
Manifold clustering
Model selection
Cluster number estimation

ABSTRACT

We present a new motion segmentation algorithm: the Enhanced Local Subspace Affinity (ELSA). Unlike Local Subspace Affinity, ELSA is robust in a variety of conditions even without manual tuning of its parameters. This result is achieved thanks to two improvements. The first is a new model selection technique for the estimation of the trajectory matrix rank. The second is an estimation of the number of motions based on the analysis of the eigenvalue spectrum of the Symmetric Normalized Laplacian matrix. Results using the Hopkins155 database and synthetic sequences are presented and compared with state of the art techniques.

1. Introduction

Motion segmentation aims to identify moving objects in a video sequence. It is a key step for many computer vision tasks such as robotics, inspection, metrology, video surveillance, video indexing, traffic monitoring, structure from motion, and many other applications. The importance of motion segmentation is evident from reviewing its vast literature. However, the fact that it is still considered a "hot" topic also testifies that there are many problems that have not yet been solved.

Based on their main underlying technique, motion segmentation strategies could be classified into the following groups: image difference, *statistical*, optical flow, wavelets, layers, and manifold clustering.

Image difference: image difference techniques are some of the simplest and most used for detecting changes. They consist in thresholding the pixel-wise intensity difference of two consecutive frames [1,2]. Despite their simplicity they provide good results being able to deal with occlusions, multiple objects, independent motions, non-rigid, and articulated objects. The main problems of these techniques are the high sensitivity to noise and to light changes, and the difficulty to deal with moving cameras and temporary stopping, which is the ability to deal with objects that may stop temporarily and hence be mistaken as background.

Statistical: statistical theory is widely used in motion segmentation. Common statistical frameworks applied to motion segmentation are Maximum A Posteriori Probability [3,4], Particle Filter [5] and Expectation Maximization [6]. Statistical approaches use mainly dense-based representations; this means that each pixel is classified, in contrast to feature-based representation techniques that classify only some selected features. This group of techniques works well with multiple motions and can deal with occlusions and temporary stopping. In general they are robust, as long as the model reflects the actual situation, but they degrade quickly as the model fails to represent reality. Moreover, most of the statistical approaches require some kind of a priori knowledge.

Wavelets: these methods exploit the ability of wavelets to analyse the different frequency components of the images, and then study each component with a resolution matched to its scale [7,8]. Wavelet solutions seem to provide overall good results but are limited to simple cases (such as translations in front of the camera).

Optical flow (*OF*): OF can be defined as the apparent motion of brightness patterns in an image sequence. Like image difference, OF is an old concept greatly exploited in computer vision and used also for motion segmentation [9–11]. OF, theoretically, can provide useful information to segment the motions. However, OF alone cannot deal with occlusions or temporary stopping. Moreover, in its simple version it is very sensitive to noise and light changes.

Layers: the key idea of layer based techniques is to divide the image into layers with uniform motion. Furthermore, each layer is associated with a depth level and a "transparency" level that determines the behaviour of the layers in the event of overlaps. Recently, new interest has arisen for this technique [12,13]. Layers are probably the most natural solution for occlusions. The main drawback is the level of complexity of these algorithms and the typically large number of parameters that have to be tuned.

* Corresponding author.
E-mail address: zappella@eia.udg.edu (L. Zappella).

0031-3203/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.patcog.2010.08.015

Manifold clustering: these techniques aim at defining a low-dimensional embedding of the data points (trajectories in motion segmentation) that preserves some properties of the high-dimensional data set, such as geodesic distance or local relationships. This class of solutions, usually based on feature points, can easily tackle temporary stopping and provides overall good results. A common drawback to all these techniques is that they perform very well when the assumptions of rigidity and independence of the motions are respected, but problems arise when one of these assumptions fails. The intense work done on manifold clustering for motion segmentation led to satisfactory performances, which make these solutions appealing. However, more work has to be done in order to have a motion segmentation algorithm that is completely automatic and independent from a priori knowledge.

In this paper we present the Enhanced Local Subspace Affinity (ELSA), a motion segmentation algorithm based on manifold clustering. ELSA is inspired by the Local Subspace Affinity (LSA) [14,15] technique introduced by Yan and Pollefeys. In contrast to LSA, ELSA is able to automatically tune its most sensitive parameter and it does not require previous knowledge of the number of motions. Such a result is achieved thanks to two improvements. The first is a new model selection technique called Enhanced Model Selection (EMS). EMS is able to adjust automatically to different noise conditions and different number of motions. A preliminary version of EMS was first presented in [16]. The second improvement introduced in this paper is an estimation of the number of motions based on finding, dynamically, a threshold for the eigenvalue spectrum of the Symmetric Normalized Laplacian matrix. By doing so, the final segmentation can be achieved by any spectral clustering algorithm without requiring any a priori knowledge about the number of motions. For all the other parameters we propose a fixed value that we use in all our experiments, showing that even without tuning them, they lead to good results in most of the cases. If one wants, all the parameters could be manually tuned in order to achieve even better performance but we were not interested in obtaining "the best result" but rather in having a good behaviour in the majority of cases without requiring manual tuning. A full source code implementation of ELSA is available at http://eia.udg.edu/~zappella.

The rest of the paper is structured as follows. In Section 2 we review the state of the art focusing on manifold clustering techniques. In particular, in Section 3, LSA [14,15] is analysed in detail. Our new proposed algorithm ELSA is presented in Section 4. The experimental results, shown in Section 5, are computed on the Hopkins155[1] database [17], which is a reference database for motion segmentation. We use also noise perturbed versions of the Hopkins155 database in order to test the behaviour of our algorithm with different noise levels. Moreover, to test the behaviour with more than 3 motions we use a synthetic database with 4 and 5 motions and controlled noise conditions. The results of ELSA are compared with LSA in order to test the new EMS. Furthermore, ELSA is compared with the recently proposed Agglomerative Lossy Compression (ALC) algorithm [18] which is, to the best of our knowledge, the best performing manifold clustering algorithm without a priori knowledge. In Section 6 conclusions are drawn, and future work is discussed.

2. Manifold clustering state of the art

This section provides a complete review on manifold clustering algorithms applied to motion segmentation. A comprehensive review on different motion segmentation techniques can be found in [19].

In general, manifold clustering solutions consist of clustering data that has common properties by, for example, fitting a set of hyperplanes to the data. Frequently, when the ambient space is very big they project the original data set into a smaller space. Most solutions assume an affine camera model, however, it is possible to extend them to the projective case by an iterative process as shown in [20].

Manifold clustering comprises a large number of different techniques, and a further classification can help in giving some order. Manifold clustering can be divided into: iterative solutions, *statistical* solutions, Agglomerate Lossy Compression (ALC), factorization solutions, and subspace estimation solutions. The techniques revised here are summarised in Table 1, which offers a compact at-a-glance overview of the manifold clustering category.

An *iterative* solution is presented in [21] where the RANdom SAmple Consensus (RANSAC) algorithm is used. RANSAC tries to fit a model to the data by randomly sampling n points, computing the residual of each point to the model and counting the number of inliers, which are those points whose residual is below a threshold. The procedure is repeated until the number of inliers is above a threshold, or enough samples have been drawn. Another iterative algorithm called "K-Subspaces Clustering" is presented in [22] for face clustering, however, the same idea could be adopted to solve the motion segmentation problem. K-Subspaces can be seen as a variant of K-means. K-Subspaces iteratively assigns points to the nearest subspace, updating each subspace by computing the new basis that minimises the sum of the square distances to all the points of that cluster. The algorithm ends after a predefined number of iterations. With a different strategy, the authors of [23] propose a subspace segmentation algorithm based on a Grassmannian minimisation approach. This technique consists in estimating the subspace with the maximum consensus (MCS), defined as the maximum number of data points that are inside the subspace. Then, the algorithm is recursively applied to the data inside the subspace in order to look for smaller subspaces included within it. The MCS is efficiently built by a Grassmannian minimisation problem.

Iterative solutions are in general robust to noise and outliers, and they provide good solutions if the number of clusters and the dimensions of the subspaces are known. This a priori knowledge can be clearly seen as their limitation as this information is not always available. Moreover, they require an initial estimation and are not robust against bad initializations and hence, are not guaranteed to converge.

The authors of [24] use a *statistical* framework for detecting degeneracies of a geometric model. They use the geometric Akaikes information criterion (AIC) defined in [25] in order to evaluate whether two clouds of points should be merged or not. Another statistic based technique is presented in [26]. This work analyses the geometric structure of the degeneracy of the motion model, and suggests a multi-stage unsupervised learning scheme, first using the degenerate motion model and then using the general 3D motion model. The authors of [27] extend the Expectation Maximization algorithm proposed in [28] for the single object case, to multiple motions and missing data. In [29] the same authors further extend the method incorporating non-motion cues (such as spatial coherence) into the M-step of the algorithm.

Statistical solutions have more or less the same strength and weaknesses of iterative techniques. They can be robust against noise whenever the statistic model is built taking the noise explicitly into account. However, when noise is not considered, or is not properly modeled, their performances rapidly degenerate. As previously mentioned statistical approaches are robust as long as the model reflects the actual situation.

A completely different idea is the basis of [18], which uses the *Agglomerative Lossy Compression* (ALC) algorithm [30].

[1] Available at http://www.vision.jhu.edu.