Simultaneous Compaction and Factorization of Sparse Image Motion Matrices

Susanna Ricco and Carlo Tomasi

Department of Computer Science
Duke University
Durham, NC USA
{sricco,tomasi}@cs.duke.edu

Abstract. Matrices that collect the image coordinates of point features tracked through video – one column per feature – have often low rank, either exactly or approximately. This observation has led to many matrix factorization methods for 3D reconstruction, motion segmentation, or regularization of feature trajectories. However, temporary occlusions, image noise, and variations in lighting, pose, or object geometry often confound trackers. A feature that reappears after a temporary tracking failure – whatever the cause – is regarded as a new feature by typical tracking systems, resulting in very sparse matrices with many columns and rendering factorization problematic. We propose a method to simultaneously factor and compact such a matrix by merging groups of columns that correspond to the same feature into single columns. This combination of compaction and factorization makes trackers more resilient to changes in appearance and short occlusions. Preliminary experiments show that imputation of missing matrix entries – and therefore matrix factorization – becomes significantly more reliable as a result. Clean column merging also required us to develop a history-sensitive feature reinitialization method we call feature snapping that aligns merged feature trajectory segments precisely to each other.

1 Introduction

Many problems in computer vision require tracking point features through video sequences. Classical examples include three-dimensional geometric reconstruction, motion segmentation, and motion compression. Features are distinguished from each other by the appearance of small image windows centered around the points of interest, and are tracked by correlating descriptors of appearance across consecutive frames. The resulting image coordinates are often arranged in a measurement matrix $M$ with one column per point. If the world is simple – perhaps characterized by one or a few rigid motions, or by an articulated body – the matrix $M$ has been shown to be low rank, and its factorization $\arg\min_{L,R} \|M - LR^T\|$ in some norm yields important information about the scene and its motions.
However, tracking point features is brittle, because the appearance within a feature window can change as a result of many factors including image noise, variations in lighting or viewpoint, or object deformations. In addition, the window being tracked may become occluded (hidden from view), perhaps to reappear several frames later. Although many tracking failures are short-lived, a reappearing feature is typically viewed as an entirely new feature, and a new column is added for it to the matrix $M$. This causes $M$ to be extremely sparse, with only a small fraction of known entries in each column, and makes finding a good factorization a challenge.

In this paper, we propose a method to simultaneously factor and compact the measurement matrix $M$ into a smaller and denser matrix $\hat{M}$ by merging groups of columns of $M$ that correspond to the same feature into single columns of $\hat{M}$. Specifically, we formulate a mixed integer program that factors $M$ while simultaneously merging together columns that are temporally, photometrically and geometrically consistent with each other. Two columns are temporally consistent with each other if their known entries come from disjoint time intervals. They are photometrically consistent if the corresponding feature windows look similar to each other. The two columns are geometrically consistent if the trajectories that they represent align well in the image. The resulting integer program is intractable, and we solve it approximately by alternating between factorization and compaction until convergence. In so doing, we lose guarantees of convergence to a global optimum, but our experiments show satisfactory results nonetheless.

Two (or more) columns can be merged when they correspond exactly to the same point. However, standard trackers initialize new features independently of past history, so it is unlikely that a new feature is found to coincide perfectly with an old one. To address this difficulty, we propose a history-sensitive feature reinitialization method, in which the image coordinates of newly defined features are snapped whenever possible to image positions that maximize the photometric and geometric similarity with a previously seen feature.

The proposed method yields a compacted matrix $\hat{M}$ that is denser than its original counterpart $M$. Our preliminary experiments show that this greater density in turn leads to better generalization, in the sense that the missing entries that the compact factorization $L\hat{R}^T$ imputes in $\hat{M}$ are more accurate than those that the sparse factorization $LR^T$ imputes in $M$.

In summary, our main contributions are (1) a method for coupling matrix compaction and factorization using chronological, geometric and photometric evidence, and (2) a history-sensitive feature reinitialization method we call feature snapping that precisely aligns merged feature trajectories to each other.

The rest of the paper is organized as follows. After a review of related literature, Section 3 presents our formulation of the combined factorization and matrix compaction problem. Section 4 discusses our technique for solving the resulting optimization problem. Section 5 describes feature snapping, and the experimental results in Section 6 show the benefits of our method. Section 7 concludes with summary, a discussion of limitations, and future work.