In Proc. IEEE CVPR 2007.

Two-View Motion Segmentation from Linear Programming Relaxation

Hongdong LI
RSISE, Australian National University
VISTA, National ICT Australia

Abstract

This paper studies the problem of multibody motion segmentation, which is an important, but challenging problem due to its well-known chicken-and-egg-type recursive character.

We propose a new Mixture-of-Fundamental-matrices *model to describe the multibody motions from two views. Based on the maximum likelihood estimation, in conjunction with a random sampling scheme, we show that the problem can be naturally formulated as a* Linear Programming *(LP) problem. Consequently, the motion segmentation problem can be solved efficiently by* linear program relaxation*. Experiments demonstrate that: without assuming the actual number of motions our method produces accurate segmentation result. This LP formulation has also other advantages, such as easy to handle outliers and easy to enforce prior knowledge etc.*

1. Introduction

Dynamic multibody scenes are very common in reality: e.g., traffic surveillance camera monitoring a busy intersection, sport camera tracking a group of soccer players, or a hand-held camera following a flying bird, etc. In the last case, the camera and the bird each contributes to an independent motion.

To enable a computer understanding a multibody dynamic scene, an efficient *multibody motion segmentation* algorithm is desirable. This problem, also known as the *multibody structure-and-motion* problem ([21][14][9][11]), consists of the following tasks: segmenting image points according to their motions, estimating individual motion's parameters, and recovering 3D structure of the points.

We propose a new algorithm in this paper, based on a principled framework of Linear Programming Relaxation. We assume that the camera model is fully projective (as oppose to linear affine camera models) so as to address close range applications where large perspective distortions appear in the images.

1.1. A brief review: existing approaches

Multibody motion segmentation is a challenging problem. This is mainly due to its well known chicken-and-egg-type character: in order to estimate multiple motions' parameters, one has better segment these motions first (i.e., find each point's membership); Reversely, to segment the multiple motions the information of each individual motion is much helpful.

EM algorithm is commonly adopted for solving such a recursive problem. It performs by alternating between parameter estimation and membership segmentation. Based on the EM algorithm, moderately successful applications have been reported [5][20]. However, the EM is only guaranteed to converge to a local minimum. Practices often show that an EM algorithm gets trapped into a local minimum thus produces erroneous segmentation.

Subspace separation has been suggested for motion segmentation. The most known algorithm is the multibody factorization due to Costeria and Kanade [9]. There are much other incremental work, some of them have substantially improved this technique (c.f., [3] [26][25] [5][16]). Unfortunately, without nontrivial modification these methods have to confine themselves only to linear camera model.

Model selection is another adopted method for the problem [19][10]. Torr [19] proposes an algorithm based on removing single motion (inliers) sequentially according to a residual analysis. Schindler et al's work [15] represents some recent development along this direction. A drawback of this method is that many parameters (e.g., threshold) need to be turned simultaneously.

GPCA is an interesting algebraic method, proposed by Vidal et al [22][8]. It is elegant and has broad applications. However, when used to segment multiple of general motions (say, m motions, $m > 5$) it requires a large number of feature points, say, in the order of $O(m^4)$, which is impractical in many circumstances. Furthermore, being algebraic in nature, the method is vulnerable to outliers.

In the present paper, we propose a novel algorithm for two-view multibody motion estimation. The algorithm is based on the Linear Programming framework, thus has a

guaranteed global optimality. It needs not an initial estimate of the the number of independent motions, and can handle a large number of motions. Moreover, it deals with outliers naturally and under the same framework with little (computational) overhead.

2. Optimal Motion Segmentation

2.1. Optimal criterion

In this paper we formulate the multi-body motion segmentation problem as a global optimization. To achieve a truly *globally-optimal* motion segmentation two issues must be addressed. For one thing, the recovered motion models must fit well with the feature points; for the other, the estimated number of motions must reflect the physical fact. To this end, we need to design a proper optimal criterion (i.e., an objective function). Akaike Information Criterion (AIC) and the Maximum Likelihood principle are adopted in this paper. The choice of the AIC is only for its simplicity but not essential. The user may replace it with any other criterion, such as the BIC [19][10], when necessary.

Specifically, we are seeking a *global optimum* which minimizes the following AIC criterion:

$$J = -2\log L + 2C \tag{1}$$

where L is a *likelihood* term (i.e. *data term*) measuring how well the motion models explain the data, and C is a *complexity* term measuring how complex the models are.

In the present work, C is proportional to the number of motions, i.e., $C \propto m$. It remains to derive the likelihood term. Under a perspective camera model, a pair of matching points from two images, $\mathbf{x}_i$ and $\mathbf{x}'_i$, are related by the epipolar equation [6]: $\mathbf{x}_i'^{\mathrm{T}}\mathtt{F}\mathbf{x}_i = 0$, where $\mathtt{F}$ is the fundamental matrix (FM). For any FM, we have $\det(\mathtt{F}) = 0, \mathtt{F} \in \mathbb{R}^{3\times 3}$.

As noise is unavoidable, the right-hand-side is nonzero, and can be used as a metric of the estimation—so-called *algebraic distance*. It is well known that the algebraic distance is problematic in practice. Another more popular, better metric, is the (squared) Sampson's distance [19][6], given by

$$d_s(\mathbf{x}_i, \mathbf{x}'_i, \mathtt{F}_k) = \frac{(\mathbf{x}_i'^{\mathrm{T}}\mathtt{F}_k\mathbf{x}_i)^2}{(\mathtt{F}_k\mathbf{x}_i)_1^2 + (\mathtt{F}_k\mathbf{x}_i)_2^2 + (\mathtt{F}_k^{\mathrm{T}}\mathbf{x}'_i)_1^2 + (\mathtt{F}_k^{\mathrm{T}}\mathbf{x}'_i)_2^2}. \tag{2}$$

In this formula, we explicitly denote the FM by $\mathtt{F}_k$ to indicate that $\{\mathbf{x}_i, \mathbf{x}'_i\}$ belong to motion-k. In reality, before the motions are segmented we have no knowledge of which point belongs to which motion and how many motions are involved. To circumvent this, Vidal et al's GPCA (see also Wolf and Shashua [24]) proposes a concept of multibody-fundamental-matrix (mFM) by constructing the product of all possible motion membership assignments: $\prod_{k=1}^{m}(\mathbf{x}_i'^{\mathrm{T}}\mathtt{F}_k\mathbf{x}_i)$=0. This mFM equation always holds regardless of from which motion the image points actually arise. But the number of terms of the product grows dramatically as the number of matches (n) or the number of motions (m) increases.

2.2. Mixture of fundamental matrices (MoF)

In reality, we do not know how to actually write eq.(2) of $d_s(\mathbf{x}_i, \mathbf{x}'_i, \mathtt{F}_k)$, because whether a matched pair $(\mathbf{x}_i, \mathbf{x}'_i)$ does arise from the motion $\mathtt{F}_k$ is not known *a-priori*.

To circumvent the obstacle, we propose a new model—called mixture-of-fundamental-matrix model (MoF). This model is inspired by the EM algorithm.

Define binary *membership variables* z_{ik} with $z_{ik} = 1$ if the image point-i is from motion-k, and z_{ik}=0 otherwise.

Under the condition of $\sum_{k=1}^{m} z_{ik} = 1$, we can replace the conventional FM equation $\mathbf{x}'\mathtt{F}\mathbf{x} = 0$ with the following mixture-of-fundamental-matrices (MoF) form,

$$\mathbf{x}_i'^{\mathrm{T}}(\sum_{k=1}^{m} z_{ik}\mathtt{F}_k)\mathbf{x}_i = 0. \tag{3}$$

This equation is equivalent to $\sum_{k=1}^{m}(\mathbf{x}_i'^{T} z_{ik}\mathtt{F}_k\mathbf{x}_i)^2 = 0$, and it always holds regardless of the actual image point memberships.

Analogously, we can replace the algebraic distance with the Sampson's distance, and obtain the following mixture-of-Sampson's-distance (MoS) formula,

$$d_{mos}(\mathbf{x}_i, \mathbf{x}'_i) = \sum_{k=1}^{m} z_{ik} d_s(\mathbf{x}_i, \mathbf{x}'_i, \mathtt{F}_k). \tag{4}$$

Using *finite mixture model* is not a new concept in computer vision, even in the context of motion segmentation. Torr [19] and Weiss [5] had exploited this idea before. Nevertheless, the idea of mixing a set of independent fundamental matrices (each of which corresponds to a distinct motion) has not been reported previously. More remarkably, we will show later that such an MoF (or MoS) model, when incorporated into the maximum likelihood estimation framework, naturally leads to a simple Linear Program formulation.

2.3. Maximum Likelihood Estimation

With the aid of the MoF model, now we are allowed to formally write the two terms of the AIC criterion function, as described below.

Under the Gaussian noise assumption, a matched image pair $(\mathbf{x}_i, \mathbf{x}'_i)$, given motion models $\mathtt{F}_k$ and memberships z_{ik},

will contribute to a *likelihood* term p as:

$$\begin{aligned} & p(\mathbf{x}_i, \mathbf{x}_i' | \mathtt{F}_k, z_{ik}, k = 1, \cdots, m) \\ &= \frac{1}{\sqrt{2\pi}\sigma} \exp(-\frac{d_{mos}(\mathbf{x}_i, \mathbf{x}_i')}{\sigma^2}) \\ &= \frac{1}{\sqrt{2\pi}\sigma} \exp(-\frac{\sum_{k=1}^{m} z_{ik} d_s(\mathbf{x}_i, \mathbf{x}_i', \mathtt{F}_k)}{\sigma^2}) \end{aligned} \quad (5)$$

At this stage we temporarily assume that the number of motions m is known.

Consider all n image point matches. The complete likelihood function L is thus (assuming statistical independency):

$$\begin{aligned} & L(\mathbf{x}_i, \mathbf{x}_i', i = 1..n | \mathtt{F}_k, z_{ik}, i = 1..n, k = 1..m) \\ &= \prod_{i=1}^{n} p(\mathbf{x}_i, \mathbf{x}_i' | \mathtt{F}_k, z_{ik}, k = 1..m). \end{aligned} \quad (6)$$

Substituting this into eq.(1) and after some algebraic simplifications, we reach the following **minimization** problem:

$$\begin{aligned} & \min_{\mathbf{z}, \mathtt{F}, m} J(z_{ik}, \mathtt{F}_k, m | i = 1..n, k = 1..m) \\ &= \min_{\mathbf{z}, \mathtt{F}, m} \sum_{i=1}^{n} \sum_{k=1}^{m} z_{ik} d_s(\mathbf{x}_i, \mathbf{x}_i', \mathtt{F}_k) + \beta m, \end{aligned} \quad (7)$$

such that,

$$m \geq 1, \quad (8)$$

$$z_{ik} \in \{0, 1\}, \sum_{k=1}^{m} z_{ik} = 1, \text{for } i = 1..n, k = 1..m, \quad (9)$$

$$\det(\mathtt{F}_k) = 0, \text{for } k = 1..m. \quad (10)$$

Recall that the last term of the AIC criterion $\mathbf{C}$ represents the model's complexity. Here we simply let $\log(\mathbf{C}) = \beta m$, where m is the (unknown) number of motions and β a trade-off factor. This amounts to say that: more motions are considered more complex.

Now that the multibody motion segmentation problem is converted to a typical ML optimization problem:

- *Find the best $\mathtt{F}_k$, z_{ik} and m, so that the above cost function is minimized.*

2.4. Convert to Linear Programming Problem

However, solving the above optimization problem is very hard. This is because: the cost function (7) itself is highly nonlinear and non-convex in the unknowns $\mathtt{F}_k$; the det constraint of (10) is cubic and non-convex; the variables z_{ik} are binary integers. Besides, even the number of motions m is unknown.

The readers who are familiar with the EM algorithm might recognize that the form of (7) is similar to the *complete-data log-likelihood* in EM, hence may wonder whether a *marginalization* (as the EM does) over the unknown z_{ik} would be of any help. We think this is a promising direction for future work.

In this work we explore a different direction based on Linear Programming Relaxation idea, which is non-conventional and very effective.

Examine eq.(7) again. Suppose that somehow we already have a list of M *candidate* motion models, denoted by $\Phi = \{\mathtt{F}_1, \mathtt{F}_2, ..., \mathtt{F}_M\}$, $|\Phi| = M$. We assume that the candidate list does contain the m true motions.

Define auxiliary binary *indicating variables* y_k with y_k=1 if the motion-$k \in [1, \cdots, M]$ is indeed one of the m true motions and y_k=0 otherwise. Clearly we have $\max_{i=1}^{n}\{z_{ik}\} = y_k$ and $\sum_{k=1}^{M} y_k = m$. Using these notations we re-write eq.(7) as

$$\begin{aligned} & \min_{\mathbf{z}, \mathtt{F}, \mathbf{y}} J(z_{ik}, \mathtt{F}_k, y_k | i = 1..n, k = 1..M) \\ &= \min_{\mathbf{z}, \mathbf{y}} \sum_{i=1}^{n} \sum_{k=1}^{M} z_{ik} d_s(\mathbf{x}_i, \mathbf{x}_i', \mathtt{F}_k) + \beta \sum_{k=1}^{M} y_k \\ &= \min_{\mathbf{z}, \mathbf{y}} \sum_{i=1}^{n} \sum_{k=1}^{M} z_{ik} d_{sik} + \beta \sum_{k=1}^{M} y_k. \end{aligned} \quad (11)$$

In this formulation, if we assume that the M candidate motions have already been found somehow, and they do contain the m true motions, then the terms of $d_s(\mathbf{x}_i, \mathbf{x}_i', \mathtt{F}_k)$ can be pre-computed and considered as known coefficients (of $z_i k$). Now the problem becomes linear in the unknown binary variables z_{ik} and y_k, $k = 1, \cdots, M$. In fact, it is a standard 0-1 Integer-Linear-Programming problem.

3. Facility Location Problem and Relaxation

So far we have successfully reduced the problem to an integer linear programming (under the assumption that all the candidate motions $\mathtt{F}_k$ are known beforehand). Now a question arises: how to solve the problem?

To *exactly* solve an integer linear programming is extremely hard (it is NP-hard in general). By *exactly*, we mean that all the unknowns (to be solved) are well constrained to be integers.

Before proceeding to explain how we actually solve the NP-hard problem, we make a detour and introduce the *facility location problem* (FLP)—more precisely the uncapacitated FLP—a well-known subject of Operations Research ([1][17]). The reason for such a detour will become clear shortly.

Imagine a big retailer company plans to open some local shops (i.e. *facilities*) to serve n customers. The locations of all customers are known beforehand. The locations of shops are not known but to be chosen from a set of candidate sites, denoted by $\mathcal{F}$. The number of shops to be opened (denoted by m) is initially unknown. Suppose there is a