

Estimation of Deformable Structure and Motion

Henrik Aanæs

Department of Mathematical Modelling
Technical University of Denmark
DK-2800 Kgs. Lyngby, Denmark
haa@imm.dtu.dk

Fredrik Kahl

Centre for Mathematical Sciences
Lund Institute of Technology
Box 118, SE-221 00 Lund, Sweden
fredrik@maths.lth.se

1 Introduction

The estimation of structure and motion from image sequences is one of the most studied problems within computer vision. However, almost all the efforts in this area have dealt with rigid objects. Our surrounding environment is generally not a rigid place, with swaying trees, moving people and rolling waves. Hence, to have structure and motion systems work effectively in our world, in general, the assumption of rigid objects has to be relaxed.

We present an approach for estimating the structure and motion of deforming or non-rigid objects. This is done by employing the Principal Component Analysis (PCA) framework, whereby the object is supposed to deform according to a linear model. That is, the structure of frame i ,

$$\mathbf{S}_i = [Q_{i1} \quad \cdots \quad Q_{in}],$$

where Q_{ij} denote the coordinates of a 3D point, can be expressed as

$$\mathbf{S}_i = S_\mu + \sum_{k=1}^r \beta_{ik} S_k, \quad (1)$$

where β_{ik} is a scalar, S_k is a 3D mode of variation and S_μ is the mean shape. These types of models have proven to be highly effective in expressing many types of deforming objects, e.g. [4]. Thus, the model is fairly general making it applicable in many different scenarios and does not require a strong domain specific prior.

The estimation of non-rigid structure and motion is done by generalizing the 'standard' rigid structure and motion approach, which is based on the following to two steps:

1. Make an initial estimate via factorization, e.g. [8].

2. Perform bundle adjustment to minimize the ML-error.

In this work, we have focused on the geometry of the problem, and hence we have not dealt with feature tracking or surface reconstruction. The main challenges in this extension are determining the complexity of the model, i.e. the value of r in (1), and dealing with the additional ambiguities accompanying the extended modeling framework.

There is of course previous work on non-rigid structure and motion estimation. Most notably [2, 3, 9] all successfully employ a similar framework for solving the same problem. They have developed a more or less heuristic factorization based approach, which is also extended to perform automatic feature tracking using optical flow. As the tracking is automatic, they are able to deal with longer image sequences and more feature points, resulting in very visually appealing reconstructions. Their experiments show the great potential of the approach. It is straightforward to extend our work in the same direction, but this will not be pursued here. Hence, as far as we have been able to establish, the *novel contributions* of this work are:

- the model selection approach.
- the identification of the additional ambiguities in the non-rigid case.
- the bundle adjustment with regularization to account for the inherent ambiguities¹, using the full perspective camera model.

Apart from this, we also propose a new heuristic for making an initial estimate of the non-rigid structure and motion.

¹In [2], a regularization prior is also used to stabilize the solution, but the prior has no direct geometric or statistical meaning.

2 Factorization Heuristic

We propose a rather simple heuristic for obtaining an initial estimate of the structure and motion. According to (1), the mean of the 3D structure is S_μ . We use the assumption that the mean is in some sense dominant. More precisely, translated into the language of factorization, *if the factorization method of Tomasi-Kanade [8] is used, then the resulting structure is similar to S_μ* . This yields a simple algorithm for producing an estimate, and it is only needed as an initial estimate. Our experiments indicate that this estimate is accurate enough for many scenarios. However, it is unclear exactly when the assumption is not valid.

Along with the estimate of S_μ , a valid motion estimate is also produced by the Tomasi-Kanade factorization algorithm. Based on this motion estimate, the non-rigid structure and its covariance can be estimated.

In order to make the initial estimate, one may instead use the methods of Brand [2] or Bregler et al. [3].

2.1 Varying Structure Estimation

Given the motion, a method for estimating the mean structure and most importantly its variance is described below. For further details, see [1].

It is well-known from statistics that efficient estimates of mean and variance are obtained by computing the mean of the observations and the squared residuals (w.r.t. the mean), respectively. However, an image of the 3D non-rigid structure is only a 2D projection. Thus, an image can be viewed as having a 3D observation with high uncertainty along the viewing direction. Given this, the mean and the variance of the non-rigid structure can be estimated by weighting according to the directional certainty of the observation. The exact formulas can be found in [1].

Assuming that the number of variation modes, r , is a priori known, then a non-rigid estimate of the structure could easily be constructed by forming the S_k of (1) from the r largest principal components of the estimated variance. Here upon the β_{ik} can be linearly solved for, giving a complete model of the non-rigid structure.

3 Model Selection

In general, the number of modes of variation, r , is not known. So, r is among the entities that needs to be estimated. This is a model selection problem and it

is by no means negligible. If too few modes are selected, the model cannot fully express the non-rigid structure. This does *not* imply that an over parameterization is preferred, since that would lead to modeling noise, again decreasing the quality of the structure estimation. In the problem at hand, there is also a worry that excess modes of variation could be 'used' to model the perspective distortion of the 3D structure. This in effect means that the motion is being 'modeled' by these excess modes. So getting a good estimate of r is a vital part of the non-rigid structure and motion estimation.

Given that we would rather under parameterize than over parameterize in order to avoid 'modeling' the motion, we propose to use Bayes information criteria (BIC) for model selection. BIC is less likely to over parameterize than other popular information criteria e.g. Akaike information criteria. We use the BIC formulation of [6], with a small alteration due to the fact that the variance is likely to be underdetermined. See [1] for further details.

In order for these information criteria to work, a reliable method for estimating the variance of the structure is needed. Thus, this section and Section 2.1 both form the basis for the model selection process. Our experiments indicate the proposed approach give good estimates of r . The whole process could be improved by incorporating the bundle adjustment algorithm in an iterative loop.

4 Perspective Solution

The final estimate of the non-rigid structure and motion is refined using the full perspective camera model.

4.1 Bundle Adjustment

Similar to traditional bundle adjustment [7], we propose to use a non-linear optimization algorithm on the observation model to get a "gold standard solution" [5]. A Levenberg-Marquardt approach is applied in order to minimize the reprojection errors in the images. The approximate solution obtained in Section 2 is used as an initial guess.

We will assume that the cameras are calibrated, but the same framework and approach would work in the uncalibrated case as well. Each camera is parameterized with a rotation matrix and the coordinates of the camera centre. The additional part in the problem at hand, is the different parameterization of the structure.

In the optimization, the collection of object points is parameterized by (1).

4.2 Ambiguities

The problem of estimating deformable structure and motion as defined has some inherent ambiguities which are not present in rigid structure and motion estimation. Also, the parameterization itself induces some additional ambiguities into the solution.

The ambiguity of rigid structure and motion concerning world coordinate system and global scale naturally extends to the non-rigid case. Thus, the recovery of structure and motion can at best be determined up to an unknown (scaled) Euclidean transformation [5].

In addition, each mode in the linear model (1) introduces four extra degrees of freedom in the reconstruction. The ambiguity corresponds to an indeterminacy of relative translation and scale between the camera centres and the object. For example, suppose the object is modeled with one mode and let $\mathbf{P}_i = \mathbf{R}_i[I - C_i]$ denote a 3×4 camera matrix with camera centres C_i . Then, a point $Q_i = X_\mu + \beta_i X_1$ is projected to

$$\begin{aligned} \lambda \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} &= \mathbf{R}_i[X_\mu + \beta_i X_1 - C_i] \\ &= \mathbf{R}_i[X_\mu + \beta_i(X_1 + T) - (C_i + \beta_i T)], \end{aligned}$$

where T is an arbitrary 3-vector. This shows that there is a translational ambiguity between the column vectors of the mode S_1 and the camera centres C_i . Similarly, one can show that there is a scale indeterminacy.

The indeterminacy means that the different time instances of the object cannot be aligned properly. A natural way to restrict the ambiguity is to impose a cost for two consecutive object instances to differ, which can be seen as a regularizing prior:

$$\delta \sum_{i=1}^m \sum_{j=2}^n \|Q_{ij} - Q_{i-1,j}\|_2^2 \quad (2)$$

where δ is a small number, e.g. $\delta = 10^{-4}$. This prior states, that if there is an ambiguity of how the 3D structure should move relative to the world coordinate system and scale, then stay stationary². The value δ should be chosen such, that the size of (2) is minute compared

²By changing global scale, the cost in (2) can be lowered. Thus, the global must be fixed and this is done by keeping the standard deviation of all 3D points in the first frame to 1.

to the residual from the data, but large enough that it is clearly distinguishable from numerical noise.

Similar to the ambiguity between the structure and the motion due to the world coordinate system, there is an ambiguity between (i) the mean S_μ and the modes S_k and (ii) the weights β_{ik} . In total, this introduces $r(r+1)$ extra degrees of freedom for r modes. These extra degrees of freedom could be removed by a different parameterization (or by fixing some parameters), but it does not change the solution.

5 Experimental Results

To validate the proposed approach, we applied it to simulated and real data. See [1] for a more extensive presentation. In the simulated experiments we focused on the factorization heuristic, in which our proposed approach proved a viable candidate compared to [3].

The proposed approach was applied to two sets of real data, where the features were tracked by hand. The first data set consisted of a group of flags flying in the wind, see [1].

The second data set consisted of an image sequence of a moving toy skeleton. One image is depicted in Figure 1. In particular, the arms and the legs in of the doll were moved in a dancing manner. The algorithm detected three modes of variation, $\{S_1, S_2, S_3\}$, on top of the mean mode, S_μ . The mean mode and the first mode of variation is depicted in Figure 2. As can be seen in the figure, the modes detected describe the deformations of the object well. The Root Mean Square (RMS) errors between the measured and reprojected points were 3.0 and 1.9 pixels after the affine factorization and after the bundle adjustment, respectively. Considering that the features were tracked by hand in images of size 960×1280 pixels, the resulting errors are indeed plausible.

References

- [1] H. Aanæs and F. Kahl. Estimation of deformable structure and motion. Technical report, Centre for Mathematical Sciences, Lund University, January 2002.
- [2] M. Brand. Morphable 3d models from video. *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, 2:456–463, 2001.
- [3] C. Bregler, A. Hertzmann, and H. Biermann. Recovering non-rigid 3d shape from image streams. *Proceed-*

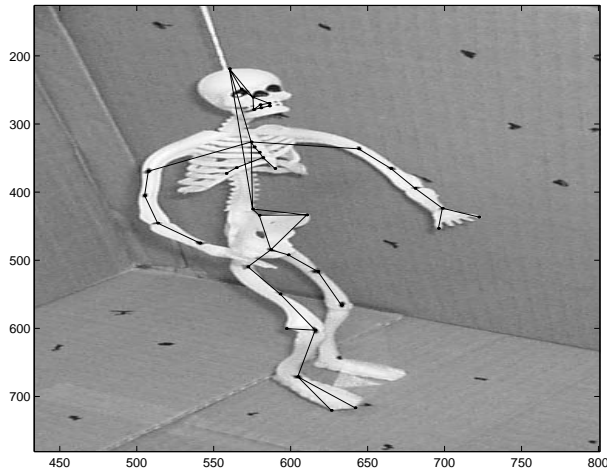


Figure 1: A frame of the skeleton sequence, with denoted features. The lines connecting the features are added for comprehensible 3D illustrations.

Mode 1

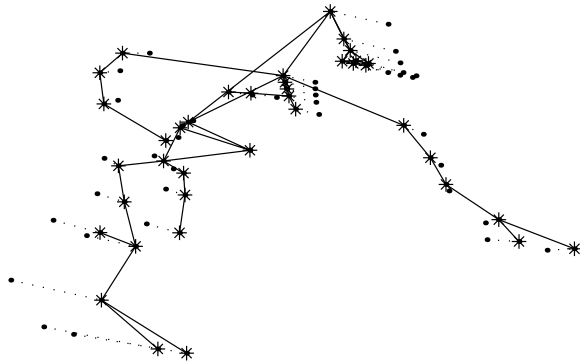


Figure 2: The mean mode, S_{μ} , and the first mode, S_1 , of the skeleton sequence. The solid lines denote the mean mode and the dotted lines illustrates the deformations of the first mode. For correspondance to the image data see Figure 1.

ings *IEEE Conference on Computer Vision and Pattern Recognition. CVPR 2000 (Cat. No. PR00662)*, pages 690–6 vol.2, 2000.

- [4] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. Active shape models – their training and application. *Computer Vision, Graphics and Image Processing*, 61(1):38–59, January 1995.
- [5] R. Hartley and A. Zisserman. *Multiple View Geometry*. Cambridge University Press, The Edinburgh Building, Cambridge CB2 2RU, UK, 2000.
- [6] Thomas P. Minka. Automatic choice of dimensionality for PCA. In *NIPS*, pages 598–604, 2000.
- [7] C.C. Slama, editor. *Manual of Photogrammetry*. American Society of Photogrammetry, Falls Church, VA, 4:th edition, 1984.
- [8] C. Tomasi and T. Kanade. Shape and motion from image streams under orthography: a factorization method. *Int'l J. Computer Vision'92*, 9(2):137–154, November 1992.
- [9] L. Torresani, D.B. Yang, E.J. Alexander, and C. Bregler. Tracking and modeling non-rigid objects with rank constraints. *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, 1:493–500, 2001.