FROM LOCAL TO GLOBAL PARAMETER ESTIMATION IN PANORAMIC PHOTOGRAPHIC RECONSTRUCTION

D. Hasler, L. Sbaiz, S. Ayer and M. Vetterli

Laboratory of Audio-visual communication
Swiss Federal Institute of Technology (EPFL)
Lausanne, Switzerland

ABSTRACT

This paper addresses a key issue in the problem of reconstructing a panoramic view out of several pictures taken with a hand held camera, namely the estimation of some ill-posed parameters using an external constraint. For many practical reasons, a panoramic reconstruction has to be performed in several independent steps, resulting in a set of different measurements of the same reality. For example, the focal length can be estimated with each pair of overlapping images. The idea is to introduce some a priori knowledge about the world by means of a constraint on the parameter set. In the former example, the constraint would impose equality on all the focal lengths estimates. This paper describes the appropriate correction that needs to be applied to the parameters in order to obtain a coherent result. It also suggests a way to evaluate if a constraint is plausible given a set of initial estimates. The basic idea behind the method is to modify the parameters without significantly changing the overlapping part of the images.

The method is evaluated using two different experimental setups. The first aims at improving the quality of a full panoramic image. The second measures independently the positions in space of two planes using two pictures. The latter experiment shows that the two computed motions can be considered as a single one with two different planes in space.

1. INTRODUCTION

To build a high quality reconstruction from several images, precise estimates of the extrinsic and intrinsic camera parameters, like the rotations angles and the focal length, are necessary. The motion estimation can be ill-posed, enabling several different sets of parameters to register the image correctly. To improve the precision, our idea is to impose some constraints on the motion, like for example the 360 degrees constraint on a full panoramic image [1][2]. This paper addresses the issue of how to apply such a constraint.

Among the existing unconstrained estimation methods, we can distinguish two families: the feature-based methods [3] and parametric or optical flow methods [4][5]. The feature-based methods look for specific patterns in the image, and try to find these same patterns in the adjacent image. The parametric methods consider the image as a set of pixels, and perform an optimisation on the pixel-wise difference of the overlapping pictures. The latter - which are considered here - often use a coarse to fine approach to converge to a solution. In general, the more parameters involved, the better the initial estimate should be to get convergence. For that reason, some 3D algorithms begin by performing some 2D estimations on the data which require less parameters and are more likely to converge [6]. This procedure ends up with a set of parameters that are inconsistent with each other. For example, one can estimate the focal length of the camera with each overlapping image pair, and will have as many different measurements as there are images in the panorama. The question is then: What is the best value for these parameters? A possible answer is to introduce some a priori knowledge about the scene by using a constraint on the parameters.

To apply a contraint, there are basically two alternatives: The first is to perform an optimisation by adding a cost which accounts for the the divergence from the constraint [2][6]. The second alternative is to use the constraint to reduce the parameter set size. The first alternative does not guarantee that the constraint is satisfied and needs to carefully control the cost function in order to end-up "close enough" to the constraint. The second is better from a precision point of view, but increases the complexity and needs a starting parameter set to find iteratively a better solution.

This paper presents a technique that implements the second alternative to the application of the constraint. The algorithm has two parts: the first part finds a set of parameters that meet the constraint using a linear approximation, and the second part refines the result iteratively. The key idea behind these techniques is to use a criterion that measures the change of the pixels position in the overlapping images. The algorithm finds a set of parameters that minimises this change with respect to the initial unconstrained estimates. In other words, if one looks at the overlapping parts of the images before and after imposing the constraint, they should look the same. It is worth pointing out that the criterion can be used to determine if a constraint can be enforced given an acceptable pixel shift on the image overlaps.

The first application improves a full panoramic image. The ill-posed parameter is the focal length that will be estimated by imposing a 360 degrees constraint on the recontruction. The second example measures the position of two planes in space. The ill-posedness comes from an inherent ambiguity between translation and rotation [7]. The constraint imposes a single camera motion between the two images.

0-7803-5467-2/99/ $ 10.00 © 1999 IEEE

2. PROBLEM FORMULATION

2.1. Notations

Two frameworks are used throughout the paper. The first uses independent and unconstrained measurements and will be denoted by the letter ξ. The parameter set under the constraint will be denoted by the letter θ or r_i. It's important to notice that θ_i and ξ_i will represent the same parameters expressed in two different spaces.

2.2. Linear optimisation on a simple constraint

Let's consider a motion model described by a two-dimensional function U_ξ which defines a mapping between two images:

$$U_\xi : p \rightarrow p'_\xi \qquad (1)$$

where p and p'_ξ denote respectively the pixel position on the warped[1] image and on the initial image. ξ stands for the motion parameters. The pixel value at position p in the warped image is the same as the one at position p'_ξ in the initial image. Let's suppose there are 2 parameter sets ξ_1 and ξ_2 on which a constraint has to be imposed. The simplest constraint is the one that enforces equality of some parameters of ξ, implying that they measure the same physical entity. For example, in a scene composed by two planes, one can measure the camera displacement between two images and the plane position in space, using an algorithm designed to work with a single plane [8][9]. Applying the algorithm twice: once for plane 1, and once for plane 2, will result in parameters ξ_1 and ξ_2. The parameter sets can be rewritten in two parts: a common part ξ^c and an autonomous part ξ^a. In the planar motion case, the ξ^c will denote the camera motion (rotation, translation and focal length), ξ_1^a the position of plane 1 and ξ_2^a the position of plane 2. Rewriting the expression in a new framework gives:

$$\begin{aligned} \theta_1 &= [\xi_1^c, \xi_1^a, 0]^T \\ \theta_2 &= [\xi_2^c, 0, \xi_2^a]^T, \\ \theta_i &= \theta + w_i \end{aligned}$$

where w_i represents the measurement - or estimation - noise and θ the "true" parameter set. Note that both θ_1 and θ_2 are expressed in what we will call the *constraint parameter space*, and are considered as being two noisy measurements of the same physical entity. In theory $\xi_1^c = \xi_2^c$. We arbitrarilly put 0 where the parameter is independent from the measurement. Now, given a set of overlapping positions p_i^1 and p_i^2 we propose to take the average position displacement in image space as a measure of the distance between θ_1 and θ_2. The goal is then to find the parameter set θ that minimises this distance

$$\begin{aligned} D &= D(\theta_1, \theta) + D(\theta_2, \theta) && (2) \\ D(\theta_i, \theta) &= \frac{1}{n_i} \sum_{j=1}^{n_i} \|U_i(p_j, \theta_i) - U_i(p_j, \theta)\|^2. && (3) \end{aligned}$$

$U_i(\cdot, \theta_i)$ is the function that performs the warping according to parameter set i, n_i is the number of pixels in the overlap area[2]. Using a first order approximation, (3) can be rewritten as

$$D(\theta_i, \theta) \approx \frac{1}{n} \sum_{j=1}^{n} \|\frac{\partial U_i}{\partial \theta}|_{(p_j, \theta_i)} (\theta_i - \theta)\|^2. \qquad (4)$$

Thus, the distance between the parameter sets is

$$\begin{aligned} D &= \sum_{i=1}^{2} \frac{1}{n_i} (\theta_i - \theta)^T E_i^T E_i (\theta_i - \theta) && (5) \\ E_i &= \frac{\partial U_i}{\partial \theta} && (6) \end{aligned}$$

where E_i has $2n_i$ rows, containing the derivatives along the x and y image axis for each pixel. To find the value of θ that minimises the distance in (5) we set the derivative with respect to θ to zero and we get

$$(E_1^T E_1 + E_2^T E_2)\theta = E_1^T E_1 \theta_1 + E_2^T E_2 \theta_2 \qquad (7)$$

where θ is obtained solving the linear system of equation (7).

2.3. Linear optimisation on a panoramic constraint

For full panoramic images, the combination of all rotations along the panorama reduces to identity:

$$R^{(1)}(\theta) \cdot \ldots \cdot R^{(m-1)}(\theta) \cdot R^{(m)}(\theta) = I \qquad (8)$$

where $R^{(i)}(\cdot)$ is the rotation matrix representing the motion between image i and image $i+1$. As previously, the individual estimations ξ_i can be written in the *constraint parameter space*, that is

$$\theta_i = [\xi_i^c, 0, .., \xi_i^a, 0, ...]^T \qquad (9)$$

where ξ^c represents the focal length of the camera, and $\xi^a$the rotation angles of each camera motion. θ only contains the parameters of the $m-1$ first rotations. In order to meet the constraint, the rotation between the last and the first image ($R^{(m)}(\theta_m)$ or $R(\xi_m)$), has to be expressed as a combination of all the other rotations:

$$\theta_m = [\xi_m^c, \upsilon_1^a, \upsilon_2^a, ..., \upsilon_{m-1}^a]^T \qquad (10)$$

such that

$$R(\xi_m) = R^{T(m-1)}(\theta_m) \cdot \ldots \cdot R^{T(1)}(\theta_m). \qquad (11)$$

In other words, the rotation $R^{(m)}(\theta_m)$ should be equal to the one found in the unconstrained estimation. There are several ways to find θ_m that will specifically be discussed in Section 3. Now the goal is to find θ that minimises the distance sum of equation (4) with

$$D(\theta_m, \theta) \approx \frac{1}{n_m} \sum_{j=1}^{n_m} \|\frac{\partial U}{\partial \xi_m} \frac{\partial \xi_m}{\partial \theta}|_{(p_j, \theta_m)} (\theta_m - \theta)\|^2. \qquad (12)$$

[1] A *warped* image is one that gets transformed by the mapping.

[2] $U_1(p_j, \theta)$ makes the correpondances between the two images of the pixels of plane 1 in the former example. This function extracts the parameters of plane 1 from θ.

Then setting

$$E_m = \frac{\partial U}{\partial \xi_m} \frac{\partial \xi_m}{\partial \theta} \tag{13}$$

the solution is found like in (7) solving the system

$$(\sum_j^m E_j^T E_j)\theta = \sum_j^m E_j^T E_j \theta_j. \tag{14}$$

The computation of $\frac{\partial \xi_m}{\partial \theta}$ is equivalent to $\frac{\partial r_m}{\partial \theta}$ which is discussed in the next section.

2.4. Non-linear optimisation

So far we have based our global optimisation task on the first order approximation expressed in equation (4). The purpose was to have a good parameter set that satisfies a constraint. Now the next step is to optimise globally the picture alignment. At this point, we have a set of points correspondences (p, p'_ξ) resulting from the first individual and non-coherent motion estimation. From the new coherent parameter set θ, one can recompute a new set of correspondences (p, p'_θ).

An unconstrained optimisation produces a good alignment of the image overlaps. Therefore, we want to minimise the distance between the points correspondences of the unconstrained method and the points correspondences of the coherent parameter set, by solving

$$\min \sum_j^m \rho(p'_{\xi_j} - p'_\theta) \tag{15}$$

where in general $\rho(\cdot)$ is the squared norm or any other function used in robust estimation [4][10]. The optimisation can be performed using a standard descent algorithm, like the Gauss-Newton algorithm [11]. These algorithms make use of the derivative of the mapping function with respect to the motion parameters $\frac{\partial U_\theta(p)}{\partial \theta}$. The computation of this function is the same as for the unconstrained case and can be found in [4] except for the last rotation in a panoramic image. The remaining part of the section explains how to compute the derivative of the function for the last rotation.

We denote by U_{r_m} the function that describes the motion between the last and the first image of the panorama, subject to the 360 degrees constraint. r_m denotes the rotation angles expressed as a function of θ. We obtain

$$\frac{\partial U_{r_m}(p)}{\partial \theta} = \frac{\partial U_{r_m}(p)}{\partial r_m} \frac{\partial r_m}{\partial \theta} \tag{16}$$

where $\frac{\partial U_{r_m}(p)}{\partial r_m}$ is the usual derivative as in the unconstrained case. To compute the derivative of the rotation angles between the last and the first image with respect to all the other rotation angles $(\frac{\partial r_m}{\partial \theta})$, we have to make a change of variables

$$\frac{\partial r_m}{\partial \theta} = \frac{\partial r_m}{\partial R_{ij}^{(m)}} \frac{\partial R_{ij}^{(m)}}{\partial \theta} \tag{17}$$

where $R_{ij}^{(m)}$ are the components of the 3×3 rotation matrix associated to the rotation angles r_m and can be found by solving equation (8). Then, we have to make use of the implicit function theorem:

$$F = \left[R_{ij}^{(m)}\right]_{9\times1} - [M(r_m)]_{9\times1} = [0]_{9\times1} \tag{18}$$

$$0 = \frac{\partial F}{\partial \left[R_{ij}^{(m)}\right]_{9\times1}} - \left[\frac{\partial [M(r_m)]_{9\times1}}{\partial r_m}\right]_{9x3} \left[\frac{\partial r_m}{\partial \left[R_{ij}^{(m)}\right]_{9\times1}}\right]_{3\times9} \tag{19}$$

where $M(r_m)$ is the rotation matrix associated to the three parameters r_m. The $[\cdot]_{lxc}$ sign denotes a matrix of l lines and c columns. Then,

$$\frac{\partial r_m}{\partial R_{ij}^{(m)}} = \left[\frac{\partial [M(r_m)]_{9\times1}}{\partial r_m}\right]^{-1} \frac{\partial F}{\partial \left[R_{ij}^{(m)}\right]_{9\times1}} \tag{20}$$

where $[\cdot]^{-1}$ denotes the generalised inversion. The inverse exists if $\frac{\partial [M(r_m)]_{9\times1}}{\partial r_m}$ has rank 3, which is untrue only when the tilting angle is equal to [3] $\pm\frac{\pi}{2}$. The remaining term to calculate is $\frac{\partial R_{ij}^{(m)}}{\partial \theta}$. To simplify the notation (e.g. to avoid the use of tensors), we will consider the derivative of the matrix with respect to one of the components of θ. We express θ as

$$\theta = [f, r_1, r_2, ..., r_{m-1}]^T \tag{21}$$

and consider component θ_α of θ. Suppose that $\theta_\alpha \in r_n$, then from (8), we get

$$\frac{\partial \left[R_{ij}^{(m)}\right]_{3\times3}}{\partial \theta_\alpha} = R^{T(m-1)}(\theta) \cdot \ldots \cdot R^{T(n+1)}(\theta) \cdot \frac{\partial \left[R_{ij}^{(n)}\right]^T}{\partial \theta_\alpha} R^{T(n-1)}(\theta) \cdot \ldots \cdot R^{T(1)}(\theta). \tag{22}$$

3. PANORAMIC IMAGES

For constructing a panoramic view of several images, the camera motion can be estimated on each pair of overlapping images. We estimated it by using a parametric model, and performing a gradient descent on the error obtained by comparing pixel-wise the overlapping parts of the images [11] [5] (it doesn't make use of feature points). We can see in Figure 1 that an accumulation of little estimation errors caused the last picture to be out of alignment with the first one. The images have been projected onto a cylinder for printing . The gap represents the misalignment with respect to the 360 degrees. By solving (14), we obtained the image in Figure 2. We can see that the error has been spread out onto each rotation, and the focal length has been adjusted to keep the overlaps between the images. The next optimisation step using criterion (15) did not produce any significant change to the result. This suggests that the approximation used in equation (4) is accurate enough for this particular example.

As mentionned in Section 2, the computation of θ_m in equation (10) is not unique. Indeed, we are trying to express

[3] In that case, the panning and rolling produce the same effect.

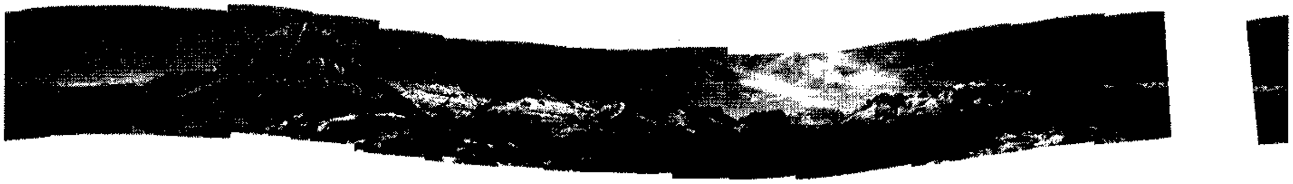


Figure 1: Full panoramic image. The estimation has been performed using only pairs of images. The little misregistration on each image results in a large error when comparing both ends of the view.

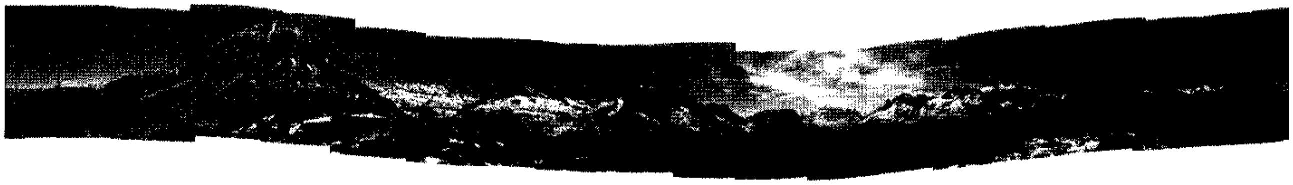


Figure 2: Full panoramic image. The result has been found by solving a linear equation to meet the 360 degrees constraint. The gap in the panorama has disappeared. Note that the panorama goes up and down a little bit. This is not due to any estimation error, but rather to the way the images are projected onto the cylinder used to print the image.

a rotation (ξ_m) as a combination of $m-1$ rotations (remember that θ_m and ξ_m represent the same rotation expressed in a different way). The way it has been performed here is by alternatively computing each v_i^a and setting $v_k^a = \xi_k^a, k \neq i$ in (10) in order to meet the constraint of equation (11). This will produce $m-1$ different results for θ_m so that we minimised the distance

$$\sum_{i=1}^{m-1} D(\theta_i, \theta) + \frac{1}{m-1} \sum_{k=0}^{m-1} D(\theta_m^{(k)}, \theta) \qquad (23)$$

instead of distance (2), where $\theta_m^{(k)}$ are the different alternatives for θ_m. This is equivalent to applying the results of Section 2 by weighting the contribution of each $\theta_m^{(k)}$ by $\frac{1}{m-1}$. The weighting step avoids the last rotation to have more weight in the optimisation process than any other rotation of the panorama.

It is worth pointing out that the constraint expressed in equation (8) is not exactly a 360 degrees constraint but rather a 0 modulus 360 degrees constraint. The side effect is that the initial estimate should be close enough to 360 in order to produce a good result. It may happen that the solution of equation (14) will lead to a 720 degrees panorama or to an identity panorama, which superimposes every image onto each other. The singularity of the method is easily detectable, but is not easy to correct in the case of an identity result.

4. EXPERIMENTS WITH TWO PLANES

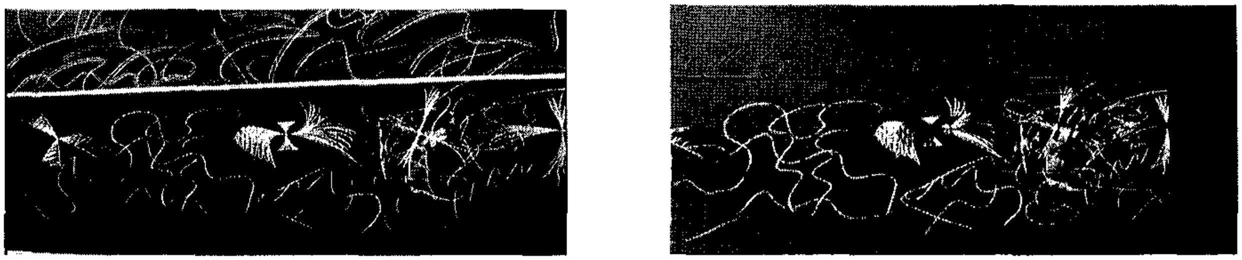
We dispose of two photographs of two scribbled blackboards taken in a classroom. The blackboards are both parallel to the wall of the room, and can slide one over the other. The top blackboard is the one that is the furthest away from the camera. The pictures have been taken with a 52mm optics and scanned at a resolution of 1860 × 1222 dots. There is a distance of 3 meters from the lower blackboard to the camera, and the camera moved 2m to the left parallely to the blackboard. The planar surface motion model is used to model the warping of the blackboards on the image and is defined by

$$\begin{aligned} X' &= RX + T && (24) \\ N^T X &= 1 && (25) \end{aligned}$$

where X are the space coordinates, R is the rotation matrix, T is the translation direction ($\|T\| = 1$) and N is the perpendicular to the plane; $\frac{1}{\|N\|}$ being the distance from the camera to the plane. Here, the image coordinate correspondences $p = (u, v) \leftrightarrow (u', v') = p'$ are known, along with the focal lengths f_1 and f_2 of the two images. The motion is computed in the following way:

$$\begin{aligned} P &= [u, v, f_1]^T && (26) \\ P' &= (R + TN^T)P && (27) \\ p' &= f_2[\frac{P_1'}{P_3'}, \frac{P_2'}{P_3'}]^T. && (28) \end{aligned}$$

By making the correspondences of some points in the image, we calculated the motion parameters using some standard techniques [8] [9]. The segmentation is supposed to be known. Here we present the most relevant results from the computation, namely a measure of the distance between the 2 blackboards. This measure depends on the camera translation which is indeed one of the ill-posed parameter of the problem [7]. Another reason for choosing this parameter is that it can easily be measured with a precision of 1 mm. Three computations have been done: The first using



(a)

(b)

Figure 3: Image of the two blackboards (separated by the white line): The warped image has been subtracted from the original one. Perfect registration is represented by grey colour. (a) The original image that gets warped in image (b). (b) Warping using the parameters for the upper blackboard. The upper blackboard is almost cancelled (e.g. uniformly grey).

two independent planes (two independent use of equations (24)...(26)); The second by calculating one single camera motion using (7); and the last performing a joint optimisation for one motion and two planes by solving equation (15) and using the result of the second computation as a starting parameter set for an iterative algorithm. The results are summarized in the following table:

type of computation	distance	error
Independent planes	19,7 cm	8.1 cm
Coherent motion	8.7 cm	2.9 cm
Joint optimisation	11.2 cm	0.4 cm
Measured	11.6 cm	-

The result shows that there is a stepwise improvement in precision. The distance computed using the 2 independent estimation (that gave 2 different camera motion) is mesured along the optical axis from the first camera position. By imposing a single camera motion, using first the linear approximation of equation (4), and the non-linear optimisation, lead to an improved result. The distance between parameter set 1 and 2 (or θ_1 and θ_2) gave 4.6 "pixels" using (4) and 1.3 "pixels" using the more precise (3), e.g. the correspondences p'_ξ have to be moved by 1.3 pixels in average in order to get a coherent parameter set. This makes the hypothesis of a single camera motion quite reasonable from the data point of view.

5. CONCLUSIONS

In this paper, we presented a technique for evaluating and combining a set of incoherent parameters. These parameters resulted from several measurements of noisy data performed on overlapping images. The basic idea behind the technique was to consider the influence of the parameter changes in image space, where the measurements are actually made. The principle is quite general and can be applied to any situation where measurements are made on image overlaps with some ill-posed parameters. The method has been tested to improve the motion estimation on a scene with two planes. Also we described a way to apply the technique to the alignment of a full panoramic mosaic. The techniques presented in this article can be viewed as an intermediate step between a local parameter estimation and a comprehensive global estimation problem.

6. REFERENCES

[1] Sing Bing Kang and R. Weiss. Characterization of errors in compositing panoramic images. In *CVPR97*, 1997.

[2] Harpreet S. Sawhney, Steve Hsu, and R. Kumar. Robust video Mosaicing through Topology Inference and Local to Global Alignement. In *ECCV98*, 1998.

[3] Olivier Faugeras. *Three-dimensional computer vision: a geometric viewpoint.* the MIT Press, 1993.

[4] Serge Ayer. *Sequential and competitive Methods for Extimation of Multiple Motions.* PhD thesis, Swiss Federal Institute of Technology (EPFL), CH-1015 Lausanne, Switzerland, 1995.

[5] S. Mann and R. W. Picard. 'Video orbits': characterizing the coordinate transformation between two images using the projective group. In *ICCV95*, 1995.

[6] H.-Y. Shum and R. Szeliski. Panoramic image mosaicing. Technical Report MSR-TR-97-23, Microsoft Research, Sep 1997.

[7] K. Daniilidis and H.-H. Nagel. Coupling of rotation and translation in motion estimation of planar surfaces. In *CVPR93*, 1993.

[8] H.C. Longuet-Higgins. The reconstruction of a plane surface from two perspective projections. *Proc. R. Soc. Lond.*, B 227:399–410, 1986.

[9] Juyang Weng, Narendra Ahuja, and Thomas S. Huang. Motion and Structure Form Point Correspondences with Error Estimation: Planar Surfaces. *IEEE transactions on Signal Processing*, 39:2691–2717, Dec 1991.

[10] Pascal Fua. Modeling heads from uncalibrated video sequences. *Videometrics*, IV, 1999.

[11] G.A.F. Seber and C.J. Wild. *Nonlinear Regression.* John Wiley and Sons, 1989.