An Application of Differential Geometric Techniques to a Problem in Optical Navigation

Sven Herzberg[1], Andrew Shaw[2], and Karlheinz Spindler[1]

Abstract – Differential geometric techniques in estimation theory are applied to an image processing problem arising in optical spacecraft navigation. Data from local images of high resolution and from global images of low resolution are optimally merged and integrated into a homogeneous information context. Two variations of an algorithm used to match the various data are presented which both satisfactorily solve the problem at hand.

Keywords – Image processing, optical navigation, pattern matching, differential geometric techniques, spacecraft applications.

I. Introduction

We consider the following situation: One is given a global map of a planetary surface resulting from a surveying campaign mapping a large region of the planet with rather low resolution. On the other hand, one is given a sequence of images from a low-flying spacecraft mapping small regions of the planet with much higher resolution. The images obtained from the spacecraft are of much better quality than the global map, but many of the images will contain no single landmark which is also contained in the global map. As long as this is the case, the only way to proceed is to optimally match image coordinates of landmarks which can be seen on subsequent pictures (like the ones shown in Fig. 1 and Fig. 2).

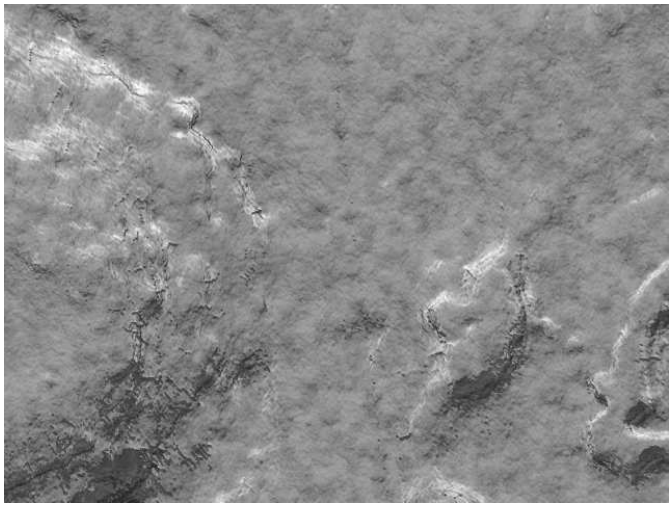

Fig. 1: Local image of a planetary surface.

[1] Hochschule RheinMain, Kurt-Schumacher-Ring 18, D - 65197 Wiesbaden, Germany. [2] SciSys Ltd., Clothier Road, Bristol, BS4 5SS, United Kingdom. This work was partially supported by the German Federal Ministry of Education and Research (BMBF) under grant no. 1763X07.

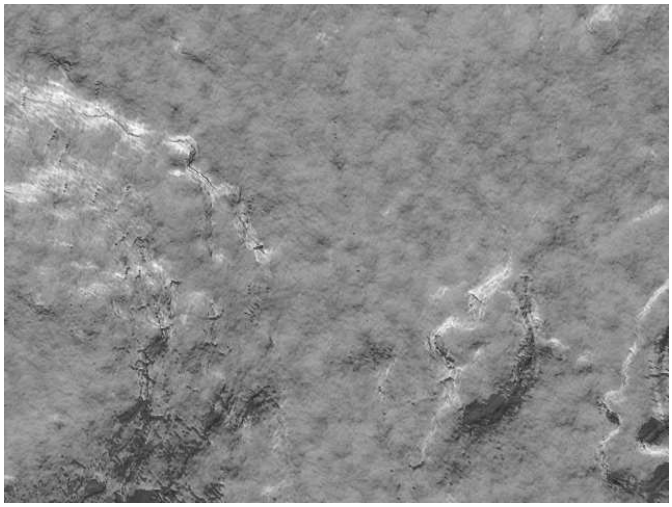

Fig. 2: Same scene under different lighting conditions.

This, however, typically leads to a sequence of estimated image frames drifting away from their true locations. (Examples of this phenomenon will be shown in a subsequent section.) Thus when at some point a landmark can be identified which is also contained in the global map rather large deviations are observed. To see an example of the kind of problems which can occur, we show simulated images of the surface shown in Fig. 3.

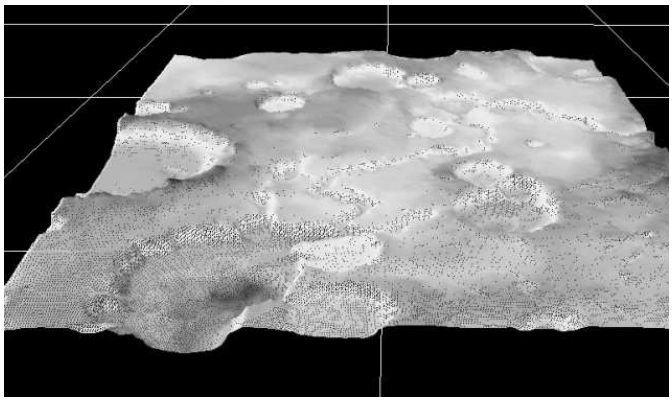

Fig. 3: Three-dimensional model landscape used for simulations.

Fig. 4 shows landmarks which can be seen from a spacecraft flying along a loop. Always trying to match subsequent image frames by optimally filtering out the noise, one encounters the problem that the overlapping regions of the first image and the last image taken cannot be properly matched any more; see Fig. 5 which shows a reconstruction based on subsequent local matchings.

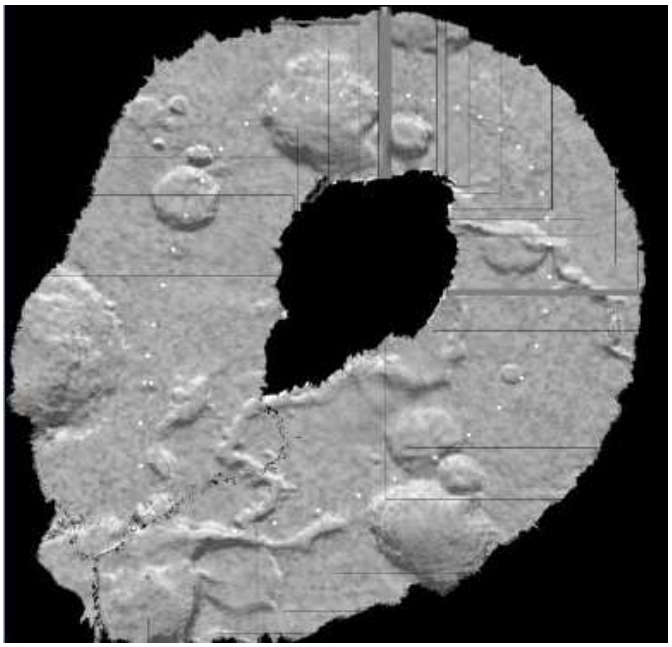

Fig. 4: Matching of unperturbed images around a closed loop.

Fig. 5: Attempt to match noisy images taken along a closed loop.

The problem which needs to be addressed is how the (relative) local information relating locally identifiable landmarks on the various satellite images can be optimally combined with the (absolute) global information expressed by the coordinates of globally identifiable (but scarce) landmarks in the absolute reference frame of the global map. An algorithm is proposed to approach this problem, and simulation results demonstrating the feasibility of this algorithm are presented.

II. Purely Local Pattern-matching

We assume that the picture sequence can be started with an image (called image 0) on which two landmarks can be seen which are available on the global map. We then take as reference frame the frame of image 0 (which amounts to defining a coordinate system for the global map). Next, we introduce the following notation: $P_1, \ldots, P_K$ are all the landmarks which can be identified on any of the images; $w_1, \ldots, w_K \in \mathbb{R}^2$ are the coordinate vectors of these landmarks with respect to the reference system; and $v_1^{(i)}, \ldots, v_K^{(i)} \in \mathbb{R}^2$ are the image coordinates of these landmarks in the picture frame of image number i. (Generally, we index landmarks by the letter k and images by the letter i.) Denoting by R_i the rotation and by τ_i the translation which carry the reference frame to the frame of the i-th image, we have

$$w_k \;=\; \tau_i + R_i v_k^{(i)} \tag{1}$$

for all k and all i. Thus if the landmark P_k can be seen in two subsequent images i and $i+1$ we have

$$\tau_{i+1} + R_{i+1} v_k^{(i+1)} = \tau_i + R_i v_k^{(i)} \,. \tag{2}$$

We now write $D_{i+1} := R_{i+1} R_i^{-1}$ and $d_{i+1} := \tau_{i+1} - \tau_i$, denote by $\mathfrak{K}_i^{i+1}$ the set of those indices k for which P_k is visible on images i and $i+1$ and also denote by $\mathfrak{G}^{i+1}$ the set of those indices k for which landmark k is visible both on the global map and on image $i+1$. Then the information to be gleaned from image $i+1$ is contained in the equations

$$DR_i v_k^{(i+1)} + d - R_i v_k^{(i)} = 0 \quad (k \in \mathfrak{K}_i^{i+1}) \tag{3}$$

and

$$DR_i v_k^{(i+1)} + d - (w_k - \tau_i) = 0 \quad (k \in \mathfrak{G}^{i+1}) \tag{4}$$

in which $D := D_{i+1}$ and $d := d_{i+1}$. (There is no need to take care of the index, because the algorithm uses one picture at a time so that in each step we have to deal with one image index $i+1$ only.) Now these equations are treated as measurement equations in which D and d are treated as estimation parameters, $v_k^{(i)}$ and w_k are treated as measurement vectors whose components are independently distributed with standard deviations σ (accuracy of local imaging) and Σ (accuracy of global map), respectively, and where the true values of R_i and τ_i are replaced by their estimates as obtained by using the previous (i.e., the i-th) picture. The sought estimates for D and d are found by minimizing the cost function

$$\begin{aligned} f(D,d) \;:=\; & \frac{1}{\sigma^2} \sum_{k \in \mathfrak{K}_i^{i+1}} \left\| DR_i v_k^{(i+1)} + d - R_i v_k^{(i)} \right\|^2 \\ & + \frac{1}{\Sigma^2} \sum_{k \in \mathfrak{G}^{i+1}} \left\| DR_i v_k^{(i+1)} + d - (w_k - \tau_i) \right\|^2 \end{aligned} \tag{5}$$

with respect to D and d. (Note that if $k \in \mathfrak{K}_i^{i+1} \cap \mathfrak{G}^{i+1}$, i.e., if P_k can be identified on both the overlapping domain of the two images and on the global map, then the index k contributes to both sums representing f.) Omitting the (fixed) index i from our notation and writing

$$U_k := R_i v_k^{(i+1)}, \quad V_k := R_i v_k^{(i)}, \quad W_k := w_k - \tau_i \tag{6}$$

the cost function can be rewritten as

$$\begin{aligned} &\frac{1}{\sigma^2} \sum_{k \in \mathfrak{K}_i^{i+1}} \left(\|U_k\|^2 + 2\langle DU_k, d - V_k \rangle + \|d - V_k\|^2 \right) \\ &+ \frac{1}{\Sigma^2} \sum_{k \in \mathfrak{G}^{i+1}} \left(\|U_k\|^2 + 2\langle DU_k, d - W_k \rangle + \|d - W_k\|^2 \right). \end{aligned} \tag{7}$$

Now let us write $U_1, \ldots, U_r$, $\overline{U}_1, \ldots, \overline{U}_s$ and $\overline{\overline{U}}_1, \ldots, \overline{\overline{U}}_t$ for those U_k for which k is in the set $\mathfrak{K}_i^{i+1} \setminus \mathfrak{G}^{i+1}$, in the set $\mathfrak{G}^{i+1} \setminus \mathfrak{K}_i^{i+1}$ and in the set $\mathfrak{K}_i^{i+1} \cap \mathfrak{G}^{i+1}$, respectively; then the cost function is the sum of the four terms

$$\begin{aligned} &\frac{1}{\sigma^2} \sum_{i=1}^{r} \left(\|U_i\|^2 + 2\langle DU_i, d - V_i \rangle + \|d - V_i\|^2 \right), \\ &\frac{1}{\sigma^2} \sum_{i=1}^{t} \left(\left\|\overline{\overline{U}}_i\right\|^2 + 2\langle D\overline{\overline{U}}_i, d - \overline{\overline{V}}_i \rangle + \left\|d - \overline{\overline{V}}_i\right\|^2 \right), \\ &\frac{1}{\Sigma^2} \sum_{i=1}^{s} \left(\left\|\overline{U}_i\right\|^2 + 2\langle D\overline{U}_i, d - \overline{W}_i, \rangle + \left\|d - \overline{W}_i\right\|^2 \right), \\ &\frac{1}{\Sigma^2} \sum_{i=1}^{t} \left(\left\|\overline{\overline{U}}_i\right\|^2 + 2\langle D\overline{\overline{U}}_i, d - \overline{\overline{W}}_i \rangle + \left\|d - \overline{\overline{W}}_i\right\|^2 \right). \end{aligned} \tag{8}$$

Using the abbreviations

$$\begin{aligned} U &:= \frac{1}{\sigma^2} \sum_{i=1}^{r} U_i, \qquad V := \frac{1}{\sigma^2} \sum_{i=1}^{r} V_i, \\ \overline{U} &:= \frac{1}{\Sigma^2} \sum_{i=1}^{s} \overline{U}_i, \qquad \overline{W} := \frac{1}{\Sigma^2} \sum_{i=1}^{s} \overline{W}_i, \end{aligned} \tag{9}$$

and

$$\begin{aligned} \overline{\overline{U}} &:= \left(\frac{1}{\sigma^2} + \frac{1}{\Sigma^2} \right) \sum_{i=1}^{t} \overline{\overline{U}}_i, \\ \overline{\overline{Z}} &:= \frac{1}{\sigma^2} \sum_{i=1}^{t} \overline{\overline{V}}_i + \frac{1}{\Sigma^2} \sum_{i=1}^{t} \overline{\overline{W}}_i = \sum_{i=1}^{t} \left(\frac{\overline{\overline{V}}_i}{\sigma^2} + \frac{\overline{\overline{W}}_i}{\Sigma^2} \right), \end{aligned} \tag{10}$$

the gradient of the cost function with respect to d is given by

$$\begin{aligned} \frac{1}{2}(\nabla_d f)(D,d) \;&= D\left(U + \overline{U} + \overline{\overline{U}} \right) \\ &\quad + \left(\frac{r+t}{\sigma^2} + \frac{s+t}{\Sigma^2} \right) \cdot d \\ &\quad - \left(V + \overline{W} + \overline{\overline{Z}} \right). \end{aligned} \tag{11}$$

A necessary condition for (D,d) to minimize the cost function is $(\nabla_d f)(D,d) = 0$, which results in

$$\begin{aligned} d \;&=\; \frac{1}{c} \left(V + \overline{W} + \overline{\overline{Z}} - D\left(U + \overline{U} + \overline{\overline{U}} \right) \right) \\ \text{where} \quad c \;&:=\; \frac{r+t}{\sigma^2} + \frac{s+t}{\Sigma^2}. \end{aligned} \tag{12}$$

To derive the optimality condition with respect to D we write

$$D \;=\; \begin{bmatrix} \cos(\varphi) & -\sin(\varphi) \\ \sin(\varphi) & \cos(\varphi) \end{bmatrix} \;=:\; D(\varphi) \tag{13}$$

and observe that

$$\frac{\mathrm{d}}{\mathrm{d}\varphi} D(\varphi) \;=\; JD(\varphi) \quad \text{where} \quad J := \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}. \tag{14}$$

Thus writing $\widehat{D} := JD$, the partial derivative of the cost function f with respect to φ is given by

$$\frac{1}{2} \frac{\partial f}{\partial \varphi} \;=\; \langle \widehat{D}(U + \overline{U} + \overline{\overline{U}}, d\rangle - (S_1 + S_2 + S_3) \tag{15}$$

where

$$\begin{aligned} S_1 \;&:=\; \frac{1}{\sigma^2} \sum_{i=1}^{r} \langle \widehat{D} U_i, V_i \rangle, \\ S_2 \;&:=\; \frac{1}{\Sigma^2} \sum_{i=1}^{s} \langle \widehat{D} \overline{U}_i, \overline{W}_i \rangle, \\ S_3 \;&:=\; \sum_{i=1}^{t} \langle \widehat{D} \overline{\overline{U}}_i, \frac{\overline{\overline{V}}_i}{\sigma^2} + \frac{\overline{\overline{W}}_i}{\Sigma^2} \rangle. \end{aligned} \tag{16}$$

Now we observe that with d as in (12) we have

$$\begin{aligned} &\langle \widehat{D}(U + \overline{U} + \overline{\overline{U}}), d \rangle = \frac{1}{c} \langle \widehat{D}(U + \overline{U} + \overline{\overline{U}}), V + \overline{W} + \overline{\overline{Z}} \rangle \\ &+ \langle \widehat{D}(U + \overline{U} + \overline{\overline{U}}), J\overline{D}(U + \overline{U} + \overline{\overline{U}}) \rangle \end{aligned} \tag{17}$$

where we used the fact that $D = -J(JD) = -J\widehat{D}$; since $\langle \xi, J\xi \rangle = 0$ for all vectors $\xi \in \mathbb{R}^2$, this reduces to

$$\langle \widehat{D}(U + \overline{U} + \overline{\overline{U}}), d \rangle \;=\; \frac{1}{c} \langle \widehat{D}(U + \overline{U} + \overline{\overline{U}}), V + \overline{W} + \overline{\overline{Z}} \rangle. \tag{18}$$

Plugging (18) into (15) and letting $\partial f / \partial \varphi = 0$ results in an equation of the form

$$\sum_{n=1}^{N} \langle \widehat{D} v_n, w_n \rangle \;=\; 0 \tag{19}$$

in which the vectors $v_n, w_n \in \mathbb{R}^2$ are known and in which $\widehat{D} \in \mathrm{SO}(2)$ is sought. Writing

$$v_n = \begin{bmatrix} a_n \\ b_n \end{bmatrix} \quad \text{and} \quad w_n = \begin{bmatrix} p_n \\ q_n \end{bmatrix}, \tag{20}$$

equation (19) becomes

$$0 = \sum_{n=1}^{N} \langle \begin{bmatrix} -\sin\varphi & \cos\varphi \\ -\cos\varphi & -\sin\varphi \end{bmatrix} \begin{bmatrix} a_n \\ b_n \end{bmatrix}, \begin{bmatrix} p_n \\ q_n \end{bmatrix} \rangle =$$

$$\sum_{n=1}^{N} (\cos(\varphi)\,(b_n p_n - a_n q_n) - \sin(\varphi)\,(a_n p_n + b_n q_n)) \quad (21)$$

$$= \cos(\varphi) \cdot \sum_{n=1}^{N} \langle v_n, J w_n \rangle - \sin(\varphi) \cdot \sum_{n=1}^{N} \langle v_n, w_n \rangle.$$

Letting $A := \sum_{n=1}^{N} \langle v_n, w_n \rangle$ and $B := \sum_{n=1}^{N} \langle v_n, J w_n \rangle$, this equation reads $A\cos\varphi - B\sin\varphi = 0$ and has the two solutions

$$\begin{bmatrix} \cos\varphi \\ \sin\varphi \end{bmatrix} = \frac{\pm 1}{\sqrt{A^2 + B^2}} \begin{bmatrix} A \\ B \end{bmatrix}, \quad (22)$$

one representing the sought minimum, the other the maximum of the cost function f. Thus we obtain an explicit formula for the optimal estimates for $D = D_{i+1}$ und $d = d_{i+1}$. We remark that this remains true for the analogous three-dimensional problem; cf. [9]. The approach described so far is pursued until the first image occurs on which landmarks show up which are identifiable on the global map. Then a readjustment is necessary which takes into account the available global information. In the following two sections we describe two approaches to perform this readjustment (both of which treat the intermediate result obtained so far as the starting point of a procedure in which the estimates obtained are iteratively improved).

III. Iterative Improvement Of Estimates

Once initial estimates for the values $\tau_i = (x_i, y_i)$ and φ_i are found using the algorithm described above, we obtain estimates for the coordinate vectors w_k simply by using equation (1) and averaging over the different values of w_k obtained for any given index k. The values τ_i, φ_i and w_k thus obtained are iteratively improved by a classical estimation scheme using the measurement equations

$$v_k^{(i)} = R_i^T (w_k - \tau_i) \quad (23)$$

which are obtained from (1). In this scheme, the coordinate vectors w_k play the roles of system parameters whereas the translations τ_i and the rotation angles φ_i play the roles of measurement parameters. To formulate the estimation scheme used, we see from (23) that the first-order effect of changes δw_k, $\delta\tau_i$ and $\delta\varphi_i$ in the estimation parameters on the measurement values are given by

$$\begin{aligned} \delta v_k^{(i)} &= (\delta R_i)^T (w_k - \tau_i) + R_i^T (\delta w_k - \delta\tau_i) \\ &= (\delta\varphi_i) R_i^T J^T (w_k - \tau_i) + R_i^T (\delta w_k - \delta\tau_i) \\ &= R_i^T \left(J^T (w_k - \tau_i) \delta\varphi_i + \delta w_k - \delta\tau_i \right) \end{aligned} \quad (24)$$

which, written in coordinates $w_k = (a_k, b_k)^T$ and $\tau_i = (x_i, y_i)^T$, takes the form

$$\begin{bmatrix} \delta\xi_k^{(i)} \\ \delta\eta_k^{(i)} \end{bmatrix} = M_k^{(i)} \begin{bmatrix} \delta x_i \\ \delta y_i \\ \delta\varphi_i \end{bmatrix} + R_i^T \begin{bmatrix} \delta a_k \\ \delta b_k \end{bmatrix} \quad (25)$$

where

$$M_k^{(i)} := \begin{bmatrix} \cos\varphi_i & \sin\varphi_i \\ -\sin\varphi_i & \cos\varphi_i \end{bmatrix} \begin{bmatrix} -1 & 0 & b_k - y_i \\ 0 & -1 & -a_k + x_i \end{bmatrix} \quad (26)$$

and introducing the parameter increment vector

$$\begin{aligned} &\delta p := (u^T, v^T) \in \mathbb{R}^{3N+2n} \quad \text{where} \\ &u := (\delta x_1,\, \delta y_1,\, \delta\varphi_1,\, \ldots,\, \delta x_N,\, \delta y_N,\, \delta\varphi_N) \quad \text{and} \\ &v := (\delta a_1,\, \delta b_1,\, \ldots,\, \delta a_n,\, \delta b_n)^T \end{aligned} \quad (27)$$

of length $3N + 2n$ (where N denotes the number of images and n the number of identified landmarks), this results in a sequence of equations

$$\delta v_k^{(i)} = \left(A_k^{(i)} \mid B_k^{(i)} \right) \delta p \quad (28)$$

where

$$A_k^{(i)} := (\underbrace{\mathbf{0}_{2\times 3}}_{1} \mid \cdots \mid \underbrace{\mathbf{0}_{2\times 3}}_{i-1} \mid \underbrace{M_k^{(i)}}_{i} \mid \underbrace{\mathbf{0}_{2\times 3}}_{i+1} \mid \cdots \mid \underbrace{\mathbf{0}_{2\times 3}}_{N}) \quad (29)$$

and

$$B_k^{(i)} := (\underbrace{\mathbf{0}_{2\times 2}}_{1} \mid \cdots \mid \underbrace{\mathbf{0}_{2\times 2}}_{k-1} \mid \underbrace{R_i^T}_{k} \mid \underbrace{\mathbf{0}_{2\times 2}}_{k+1} \mid \cdots \mid \underbrace{\mathbf{0}_{2\times 2}}_{n}) \quad (30)$$

and where, generally, $\mathbf{0}_{r\times s}$ denotes the zero matrix with r rows and s columns. In addition, if we denote by $\mathbf{1}_{r\times r}$ the $(r \times r)$-identity matrix, we get an equation

$$\delta w_k = \left(\mathbf{0}_{2\times(3N)} \mid C_k \right) \delta p \quad (31)$$

where

$$C_k := (\underbrace{\mathbf{0}_{2\times 2}}_{1} \mid \cdots \mid \underbrace{\mathbf{0}_{2\times 2}}_{k-1} \mid \underbrace{\mathbf{1}_{2\times 2}}_{k} \mid \underbrace{\mathbf{0}_{2\times 2}}_{k+1} \mid \cdots \mid \underbrace{\mathbf{0}_{2\times 2}}_{n}) \quad (32)$$

for each k for which w_k can be seen on the global map. Combining the above equations results in an equation of the form

$$\delta m = A\,\delta p \quad (33)$$

in which δm is the vector of increments of all measurements and A is a matrix with $3N + 2n$ columns and as many rows as there are measurements. We now assume that there are m_1 purely local measurements and m_2 "measurements" consisting of reading off the coordinates of the globally identifiable landmarks from the global map; taking the m_1 former measurements first and the m_2 latter measurements next and letting

$$W := \begin{bmatrix} \sigma^2 \mathbf{1}_{m_1\times m_1} & 0 \\ 0 & \Sigma^2 \mathbf{1}_{m_2\times m_2} \end{bmatrix} \quad (34)$$

we obtain the optimal increment

$$\delta p \;=\; (A^T W A)^{-1} A^T W\, \delta m. \tag{35}$$

This last equation (with δm being the vector of differences between the theoretically expected and actually obtained measurements) is used to update p via $p_{\text{new}} = p_{\text{old}} + \delta p$. This method is iterated until convergence is obtained.

IV. Gradient Algorithm

Our second approach to merge the purely local information with global data is based on a gradient algorithm; cf. [1], [3], [4]. We now describe this algorithm, assuming an arbitrary dimension d, which is no more complicated than the case $d = 2$. The notation becomes slightly more complicated than before, as we formulate the algorithm in such a way that multiple overlaps can be dealt with. As before, let N denote the number of images. Assume that there are $n_{ij} = n_{ji}$ point correspondences between the i-th and the j-th image, say between the d-dimensional points p_{ij}^k and $p_{ji}^k \in \mathbb{R}^d$ where $1 \le k \le n_{ij}$. Furthermore, let c_i^k denote the points in the i-th image which correspond to global points K_i^k where $1 \le k \le m_i$. We weight the correspondences with coefficients α_{ij} and β_i, where $1 \le i, j \le N$. (In our situation we will use $\alpha_{ij} = 1/\sigma^2$ and $\beta_i = 1/\Sigma^2$ for all i, j, but it may be helpful in practical situations to use different weighting factors in order to incorporate the quality of observation conditions.) As before, let $R_i \in \mathrm{SO}(d)$ and $T_i \in \mathbb{R}^d$ denote the rotation and translation carrying the reference frame into the i-th image frame. With the notation settled, our task is to minimize the function $f : \mathrm{SO}(d)^N \times \mathbb{R}^{dN} \to \mathbb{R}_0^+$ given by

$$\begin{aligned} &f(R_1, \dots, R_N, T_1, \dots, T_N) \;:= \\ &\sum_{i=1}^{N} \sum_{j=i+1}^{N} \sum_{k=1}^{n_{ij}} \alpha_{ij} \left\| R_i p_{ij}^k + T_i - R_j p_{ji}^k - T_j \right\|^2 \\ &+ \sum_{i=1}^{N} \sum_{k=1}^{m_i} \beta_i \left\| R_i c_i^k + T_i - K_i^k \right\|^2 . \end{aligned} \tag{36}$$

We let $\mathfrak{R} := [R_1 \dots R_N] \in \mathbb{R}^{d \times dN}$ and $\mathfrak{T} := [T_1 \dots T_N] \in \mathbb{R}^{d \times N}$ and denote by $(e_1, \dots, e_N)$ the canonical basis of $\mathbb{R}^N$. Then $T_i = \mathfrak{T} e_i$ and $R_i = \mathfrak{R}(e_i \otimes I_d)$ so that the function f takes the form

$$\begin{aligned} f(\mathfrak{R}, \mathfrak{T}) \;=\; &\operatorname{tr}(\mathfrak{R} A \mathfrak{R}^T) + 2 \operatorname{tr}(\mathfrak{R} B \mathfrak{T}^T) + \operatorname{tr}(\mathfrak{T} C \mathfrak{T}^T) \\ &- 2 \operatorname{tr}(\mathfrak{R} D) - 2 \operatorname{tr}(\mathfrak{T} E) + c \end{aligned} \tag{37}$$

where

$$\begin{aligned} A \;&:=\; \sum_{i=1}^{N} \sum_{j=i+1}^{N} \sum_{k=1}^{n_{ij}} \alpha_{ij} a_{ij}^k (a_{ij}^k)^T + \sum_{i=1}^{N} \sum_{k=1}^{m_i} \beta_i b_i^k (b_i^k)^T, \\ B \;&:=\; \sum_{i=1}^{N} \sum_{j=i+1}^{N} \sum_{k=1}^{n_{ij}} \alpha_{ij} a_{ij}^k (e_{ij})^T + \sum_{i=1}^{N} \sum_{k=1}^{m_i} \beta_i b_i^k (e_i)^T, \\ C \;&:=\; \sum_{i=1}^{N} \sum_{j=i+1}^{N} \sum_{k=1}^{n_{ij}} \alpha_{ij} e_{ij} (e_{ij})^T + \sum_{i=1}^{N} \sum_{k=1}^{m_i} \beta_i e_i (e_i)^T, \\ D \;&:=\; \sum_{i=1}^{N} \sum_{k=1}^{m_i} \beta_i b_i^k (K_i^k)^T, \\ E \;&:=\; \sum_{i=1}^{N} \sum_{k=1}^{m_i} \beta_i e_i (K_i^k)^T \end{aligned} \tag{38}$$

with $a_{ij}^k := (e_i \otimes I_d) p_{ij}^k - (e_j \otimes I_d) p_{ji}^k$, $b_i^k := (e_i \otimes I_d) c_i^k$, $e_{ij} := e_i - e_j$ and a constant c. After differentiating (37) with respect to $\mathfrak{T}$, we can decouple $\mathfrak{T}$ and $\mathfrak{R}$ and find that the optimal solution T^* satisfies

$$T^* = -(\mathfrak{R} B - E^T)(C^\dagger)^T \tag{39}$$

where $C^\dagger$ denotes the pseudoinverse of C. If we use (39) to simplify (37), we see that we have to minimize the function $f : \mathrm{SO}(d)^N \to \mathbb{R}_0^+$ given by

$$f(\mathfrak{R}) = \operatorname{tr}(\mathfrak{R} M \mathfrak{R}^T) + \operatorname{tr}(\mathfrak{R} L) + \text{const.} \tag{40}$$

where $M := A - 2B(C^\dagger)^T B^T + B(C^\dagger) C (C^\dagger)^T B^T$ and $L := 2B(C^\dagger)^T E - 2B(C^\dagger) C (C^\dagger)^T E + 2B(C^\dagger) E - 2D$; using the relations $C^\dagger C = I$, $C^T = C$ and $C^\dagger C C^\dagger = C^\dagger$ this simplifies to $M = A - BC^\dagger B^T$ and $L = 2BC^\dagger E - 2D$. The constant term in (40) can be neglected for the purpose of optimization. To formulate a gradient algorithm for the minimization of f, it is helpful to make a few remarks about the manifold $\mathrm{SO}(d)^N$. At any given point $P \in \mathrm{SO}(d)$, the tangent space of $\mathrm{SO}(d)$ at P is given by

$$T_P \mathrm{SO}(d) = \{P\xi \mid \xi^T = -\xi\}. \tag{41}$$

Now for matrices $A_1, \dots, A_s$ in $\mathbb{R}^{n \times n}$ the direct sum $A_1 \oplus \dots \oplus A_s$ denotes the $n^2 \times n^2$ block diagonal matrix with $A_1, \dots, A_s$ along the diagonal; then the tangent space of $\mathrm{SO}(d)^N$ at $\mathfrak{P} = [P_1 \dots P_N] \in \mathbb{R}^{d \times dN}$ (where $P_i \in \mathrm{SO}(d)$) is given by

$$T_{\mathfrak{P}} \mathrm{SO}(d)^N = \{\mathfrak{P}(\xi_1 \oplus \dots \oplus \xi_N) \mid \xi_i^T = -\xi_i\}. \tag{42}$$

Using the exponential map, we can parametrize the manifold $\mathrm{SO}(d)$ via its tangent space, i.e. by skew-symmetric matrices. Define $\sigma : \mathbb{R}^{d \times d} \to \mathbb{R}^{d \times d}$ by $\sigma(X) = (X - X^T)/2$, i.e., by assigning to each square matrix its skew-symmetric part. Letting

$$M_{jk} \;:=\; \mathfrak{P}(e_j \otimes I_d)(e_j \otimes I_d)^T M (e_k \otimes I_d)(e_k \otimes I_d)^T \mathfrak{P}^T \tag{43}$$

we find that

$$\nabla_{\mathfrak{P}} f \;=\; \nabla_{\mathfrak{P}}^{(1)} f \oplus \dots \oplus \nabla_{\mathfrak{P}}^{(N)} f \tag{44}$$

where

$$\nabla_P^{(j)} f \;:=\; \sum_{k=1}^{N} \left(\sigma(M_{jk}) - \sigma(M_{kj})\right) + \sigma\left(\mathfrak{P}(e_j \otimes I_d)(e_j \otimes I_d)^T L\right) \in \mathbb{R}^{d\times d} \tag{45}$$

Now a gradient algorithm step of step size $\lambda \in \mathbb{R}$ to minimize f, updating an estimate $\mathfrak{P}$ to an improved estimate $\tilde{\mathfrak{P}}$, is given by

$$\tilde{\mathfrak{P}} \;=\; \mathfrak{P} \exp\left(-\lambda \nabla_{\mathfrak{P}} f\right). \tag{46}$$

In our simulations we used, in each step, the step size $\lambda = 1$ at the beginning and then iteratively multiplied this step size by 0.5 until the cost function takes a value smaller than the value in the step before; cf. [5]. This procedure was repeated until convergence was achieved.

V. Results

At the end of this paper we present some figures showing typical results of the performance of the proposed method. First (in figures 7 through 10) a sequence of ten pictures is considered such that in the tenth picture a point identifiable on the global map occurs.

- Figure 7 shows the original data without noise; points seen in the global map are marked with a circle (o) and local points are marked with a plus sign (+).
- Figure 8 shows the calculated positions of the image frames in the global map obtained by using only local information from ten subsequent images based on data corrupted with noise ($\sigma = 0.000625$ and $\Sigma = 0.001$).
- In Figure 9 we calculated the position of the images as in case 8, but this time we also made use of the global point which can be seen in the last image and leads to better results for the position of the tenth image than in Figure 8 (at the expense of an obvious discontinuity between the ninth and the tenth image).
- Figure 10 shows the position of the images obtained by the readjustment with either of the two methods described before. The results are convincing and close to the original positions. The deviations from the true position are smallest at the beginning and the end of the series of images, where global points can be seen. The deviations between the calculated positions and the true ones get larger towards the middle of the image sequence.

Figures 11 through 14 show the same cases as Figure 7 to 10 with the only difference that this time two global points arise in the last image (rather than just one as in the case before).

- Figure 11 shows the original data without noise (and coincides with Figure 12 except for the second global point in the last image).
- Figure 12 shows the calculated positions of the image frames in the global map obtained by using purely local information from ten subsequent images, again based on data corrupted by noise ($\sigma = 0.000625$ and $\Sigma = 0.001$ as before). Since no global information was used, this coincides with the results depicted in Figure 8.
- Figure 13 was obtained by also using the global position information available for the the global points in the last image. As can be seen (and was, of course, expected from the beginning) the use of this information leads to a better estimate for the location of the last image frame (and a more pronounced discontinuity between the ninth and the tenth image) in comparision with both Figure 11 and Figure 9, because the availability of two globally identifiable points in the last image frame provides much stronger information on the location of this frame.
- Finally, Figure 14 shows the position estimates of the image frames obtained by the readjustment described before. The availability of the second global point increases the accuracy of the estimates as compared to the ones shown in Figure 10.

VI. Differential Geometric Background

Before we present our conclusions we want to reveal the differential geometric ideas behind the algorithms presented. The optimal matching of the available image data closely resembles a standard estimation problem which, in reasonable generality, can be formulated as follows; cf. [6] and [7]. A measurement vector μ depends on two kinds of parameters U and u which are distinguished because of the different roles they play in the subsequent estimation process: U is treated as a solve-for parameter whereas u is taken as a consider parameter; i.e., the value of U will be estimated whereas u is only considered in assessing the accuracy of the estimate obtained for U. If $U^\star$ and $u^\star$ are the true (but unknown) parameter values then the measurement vector $\widehat{\mu}$ obtained is

$$\widehat{\mu} \;=\; \mu(U^\star, u^\star) + n \tag{47}$$

where n is the measurement noise (whose covariance matrix is supposed to be known). We assume that we have initial estimates U_{init} and u_{init} for the parameters in question. While the estimate for u is never changed, we want to iteratively improve the available estimate for U. Thus we ask how to optimally update an "old" estimate U_{old} to obtain a "new" estimate

$$U_{\text{new}} \;=\; U_{\text{old}} + \delta U\,. \tag{48}$$

To assess the quality of an arbitrary estimate (U, u), we introduce the residual vector

$$\rho(U, u) \;=\; \widehat{\mu} - \mu(U, u) \tag{49}$$

which is a list of the differences between the actually obtained and the theoretically expected measurements. To properly measure the size of the residual vector, we weight

the different measurements according to their respective accuracies; i.e., we introduce the scalar quantity

$$Q(U,u) \;:=\; \rho(U,u)^T W \rho(U,u) \tag{50}$$

with the weighting matrix

$$W \;=\; \mathrm{Cov}[n]^{-1}\,. \tag{51}$$

Denoting by $\sqrt{W}$ the unique upper triangular matrix M such that $W = M^T M$ (practically obtained by performing the Cholesky decomposition of W) we can write

$$Q(U,u) \;=\; \| \sqrt{W}\rho(U,u) \|^2; \tag{52}$$

thus in the case of uncorrelated measurements Q is simply the sum of the squares of the weighted residuals, where the weighting factor for any measurement is the reciprocal of the standard deviation of this measurement. Now an update step δU as in (48) is considered optimal if it minimises the size of the resulting "new" residual vector

$$\begin{aligned} \rho_{\mathrm{new}} \;&=\; \rho(U_{\mathrm{new}}, u_{\mathrm{init}}) \;=\; \rho(U_{\mathrm{old}} + \delta U, u_{\mathrm{init}}) \\ &\approx\; \rho(U_{\mathrm{old}}, u_{\mathrm{init}}) + (\partial\rho/\partial U)(U_{\mathrm{old}}, u_{\mathrm{init}})\delta U \\ &=\; \rho_{\mathrm{old}} - A(U_{\mathrm{old}}, u_{\mathrm{init}})\delta U \end{aligned} \tag{53}$$

where

$$A(U,u) \;:=\; \frac{\partial \mu}{\partial U}(U,u) \tag{54}$$

denotes the matrix of partial derivatives of the measurements with respect to the solve-for parameters. Thus, using first-order approximations, we want to choose the update δU such that

$$Q_{\mathrm{new}} = \| \sqrt{W}\rho_{\mathrm{new}} \|^2 = \| \sqrt{W}\rho_{\mathrm{old}} - \sqrt{W}A\delta U \|^2 \tag{55}$$

(where $A := A(U_{\mathrm{old}}, u_{\mathrm{init}})$) becomes minimal. It is well known that if A has maximal rank this minimisation problem has the unique solution

$$\delta U \;=\; (A^T W A)^{-1} A^T W \rho_{\mathrm{old}}\,. \tag{56}$$

However, the matrix $A^T W A$ is often ill-conditioned; thus for numerical reasons it is not recommended to perform the matrix inversion in (56) in a straightforward way. Instead, we determine (typically by using a sequence of Householder transformations) an orthogonal matrix P such that

$$P\sqrt{W}A \;=:\; R \;=\; \begin{bmatrix} R_1 \\ 0 \end{bmatrix} \tag{57}$$

has upper triangular form (where R_1 is an upper triangular square matrix (whose size is given by the number of solve-for parameters). We let

$$\xi := P\sqrt{W}\rho_{\mathrm{old}}; \tag{58}$$

since applying an orthogonal matrix does not effect the norm of a vector, (55) becomes

$$\begin{aligned} Q_{\mathrm{new}} \;&=\; \|\xi - R\,\delta U\|^2 \;=\; \left\| \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix} - \begin{bmatrix} R_1\delta U \\ 0 \end{bmatrix} \right\|^2 \\ &=\; \|\xi_1 - R_1\delta U\|^2 + \|\xi_2\|^2; \end{aligned} \tag{59}$$

it is clear that this last expression is minimised by letting

$$\delta U \;:=\; R_1^{-1}\xi_1. \tag{60}$$

Note that (60) yields (56) because

$$\begin{aligned} &R_1^{-1}\xi_1 \;=\; (R_1^T R_1)^{-1} R_1^T \xi_1 \;=\; (R^T R)^{-1} R^T \xi \;= \\ &(A^T \sqrt{W}^T P^T P \sqrt{W} A)^{-1} A^T \sqrt{W}^T P^T P \sqrt{W}\rho_{\mathrm{old}} \end{aligned} \tag{61}$$

which, using $P^T P = \mathbf{1}$, becomes $(A^T W A)^{-1} A^T W \rho_{\mathrm{old}}$. Thus we know how the update step (48) should be performed. Since in each step we linearised about the current estimate, iteration of the procedure is necessary. To monitor convergence, we note from (53) that we can predict which residual vector can be expected in the next iteration (to be performed with U_{new} instead of U_{old}), namely

$$\rho_{\mathrm{expected}} \;=\; \rho_{\mathrm{old}} - A(U_{\mathrm{old}}, u_{\mathrm{init}})\delta U\,. \tag{62}$$

We consider convergence to be achieved if the difference between the residual vectors expected for and actually obtained in the next iteration becomes "small"; i.e., if

$$\max_{1\le i\le N} |\, \big(\sqrt{W}(\rho_{\mathrm{obtained}} - \rho_{\mathrm{expected}})\big)_i \,| \;<\; \varepsilon \tag{63}$$

for some predefined convergence margin $\varepsilon > 0$. It remains to assess the accuracy of the estimate obtained. After convergence, all remaining residuals are supposed to stem exclusively from the measurement noise and the uncertainty in the consider parameter estimate u_{init} (whereas the final estimate obtained for U is supposed to be the true value $U^\star$). Then, if $\delta u := u_{\mathrm{init}} - u^\star$ is the error in the estimate for u, the residual vector becomes

$$\begin{aligned} \rho \;&=\; \widehat{\mu} - \mu(U^\star, u^\star + \delta u) \\ &=\; \mu(U^\star, u^\star) + n - \mu(U^\star, u^\star + \delta u) \\ &\approx\; \mu(U^\star, u^\star) + n - \mu(U^\star, u^\star) - (\partial\mu/\partial u)(U^\star, u^\star)\delta u \\ &=\; n - (\partial\mu/\partial u)(U^\star, u^\star)\delta u \\ &\approx\; n - (\partial\mu/\partial u)(U^\star, u_{\mathrm{init}})\delta u\,. \end{aligned} \tag{64}$$

Making the (natural) assumption that the measurement noise and the error in the consider parameter estimate are uncorrelated and writing

$$B \;:=\; \frac{\partial\mu}{\partial u}(U^\star, u_{\mathrm{init}}), \tag{65}$$

we find from (64) that

$$\begin{aligned} \mathrm{Cov}[\rho] \;&=\; \mathrm{Cov}[n] + \mathrm{Cov}[B\,\delta u] \\ &=\; W^{-1} + B\mathrm{Cov}[\delta u]B^T \end{aligned} \tag{66}$$

and hence from (56) that

$$\begin{aligned} \mathrm{Cov}[\delta U] &= (A^T W A)^{-1} A^T W \mathrm{Cov}[\rho] W A (A^T W A)^{-1} \\ &= (A^T W A)^{-1} + D \mathrm{Cov}[\delta u] D^T \end{aligned} \tag{67}$$

where $D := (A^T W A)^{-1}(A^T W B)$. Note that the first summand in (67) represents the parameter estimation inaccuracy due to the noise in the measurements whereas the second summand represents the parameter estimation inaccuracy due the consider parameter uncertainty.

Up to this point, the treatment of a standard linear estimation problem has been described. Our situation is somewhat different because the estimation parameters include rotation matrices which cannot be taken as elements of a linear space, but must be treated as elements of a nonlinear manifold. (This fact is somewhat disguised in the two-dimensional case because an element of the rotation group in dimension 2 is uniquely characterized by a real number, namely the rotation angle, but becomes important when a generalization to the three-dimensional case is sought.) As is explained in [6] and [7], this problem can be overcome by modifying the estimation procedure in such a way by only allowing, in each step, only update vectors in the tangent space of the manifold in question at the currently best estimate. Once the update vector is found, it is "wrapped around" the manifold by applying to it the exponential mapping of the manifold at the currently best estimate; in other words, the update is applied along the geodesic determined by the tangent vector found by the linear estimation process described before. This is visualized in the following figure, in which the nonlinear parameter to be estimated is exemplified by a unit vector e and in which the update vector δe is represented by a tangent vector to the manifold (here a sphere) at the currently best estimate e_{old}).

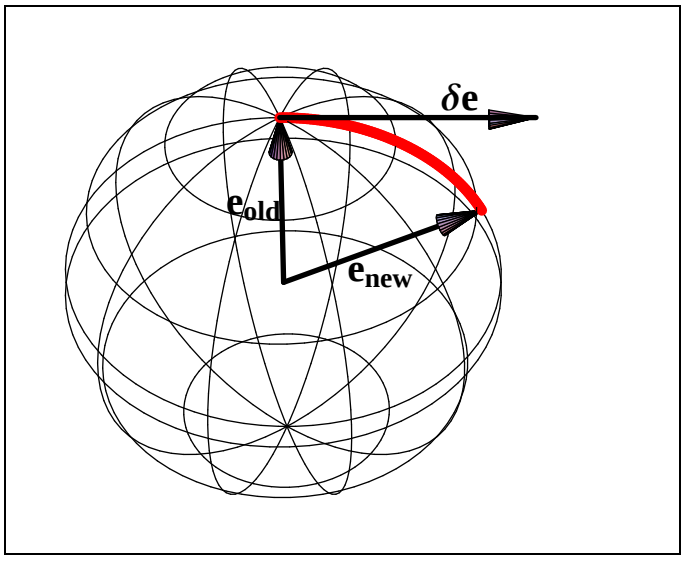

Fig. 6: Visualization of a nonlinear update step along a geodesic (here a great circle on a sphere).

Similarly, in the gradient algorithm described in section IV the gradient is formed with respect to the Riemannian structure on the manifold of estimation parameters (which coincides with the orthogonal projection of the gradient formed in the ambient space onto the tangent space at the currently best estimate).

VII. Conclusion

A problem in optical navigation was studied in which landmark position coordinates are available on a global map with low resolution and on local maps with high resolution. Two different approaches were studied which serve to optimally merge the two kinds of data, one based on successive linearizations and solutions of linear regression problems, the other based on a gradient algorithm. Both algorithms were found to yield extremely good solutions, the computational loads being similar in the two approaches. The methods carry over to the three-dimensional case; this will be elaborated in future work. Similar methods have been successfully applied to other problems in space navigation; cf. [2] and [8].

Acknowledgment

Fruitful discussions with Oana Curtef (Ph. D. student, Hochschule RheinMain and Julius-Maximilians-Universität Würzburg) are gratefully acknowledged.

References

[1] R. Fletcher, *Practical methods of optimization*, 2nd edition, Wiley, 2000.

[2] L. Fraiture, K. Spindler, *Spherical Goniometry and Spacecraft Attitude Determination*, Elemente der Mathematik (57) 2002, pp. 137-157

[3] S. Krishnan et al., *Optimisation-on-a-manifold for global registration of multiple 3D point sets*, Int. J. Intelligent Systems Technologies and Applications Vol. 3, No. 3/4, 2007.

[4] J. Nocedal and S. J. Wright, *Numerical Optimization*, 2nd edition, Springer, New York 2006.

[5] T. Smith, *Optimization Techniques on Riemannian Manifolds*, Fields Institute Communications 3 (1994).

[6] K. Spindler, *Optimal Calibration of a Camera System*, WSEAS Journal on Systems Theory and Applications (11) 2007, pp. 125-130

[7] K. Spindler, *Optimal Calibration of an X-Ray Detection System*, WSEAS Journal on Systems Theory and Applications (11) 2007, pp. 229-236

[8] K. Spindler, *Estimation of Cometary Rotation Parameters based on Camera Images*, Proc. 20th Int. Symp. Space Flight Dynamics, Annapolis, September 2007

[9] K. Spindler, *A Unified Mathematical Treatment of Regression Problems in Image Processing*, Springer Lecture Notes in Artificial Intelligence 5108, July 2008, pp. 108-122.

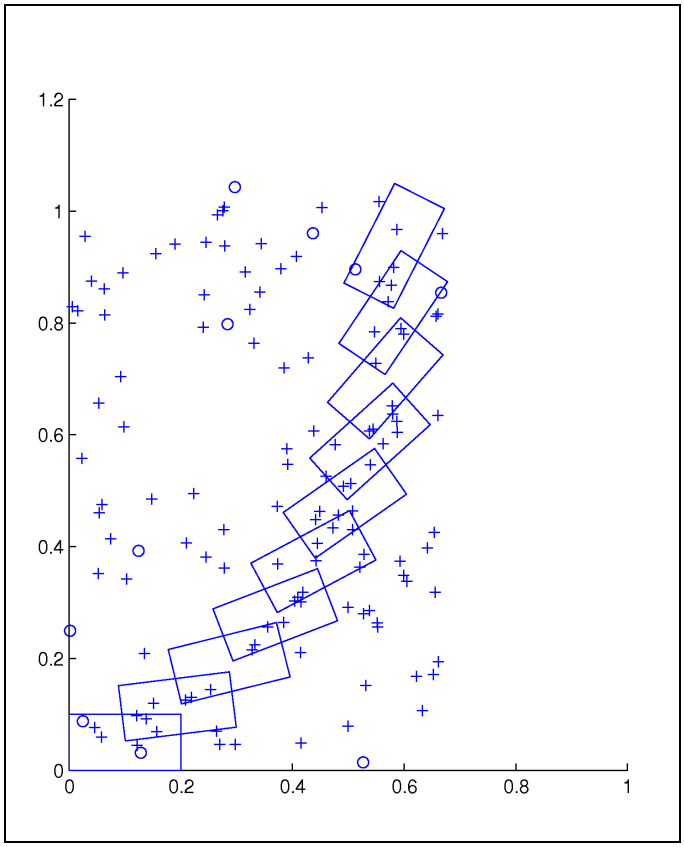

Fig. 7.

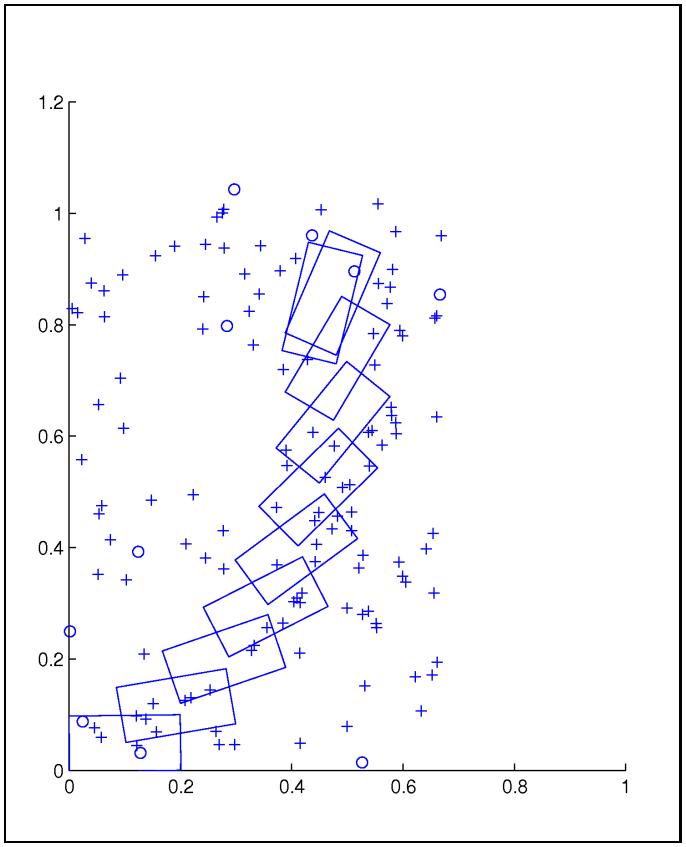

Fig. 9.

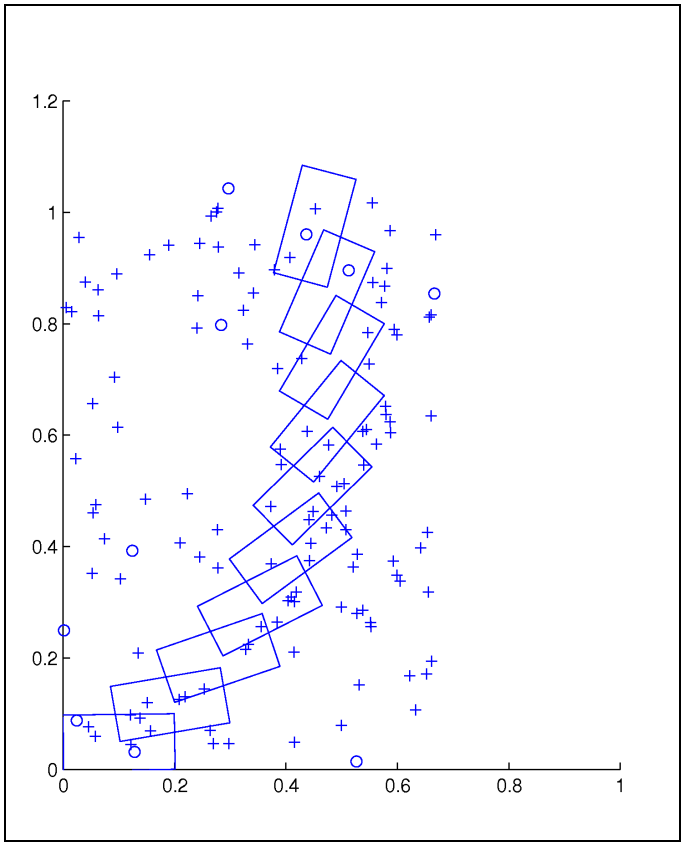

Fig. 8.

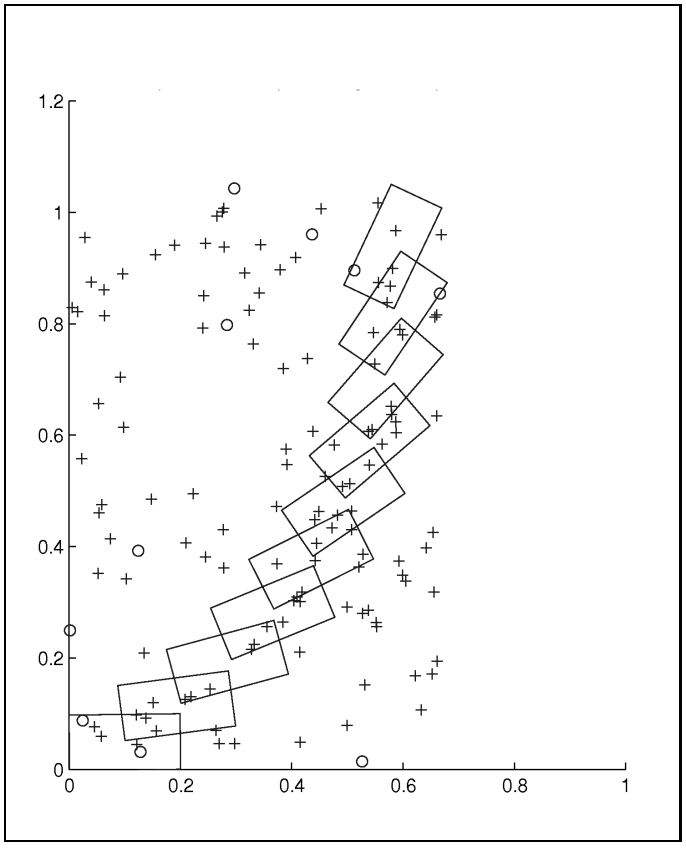

Fig. 10.

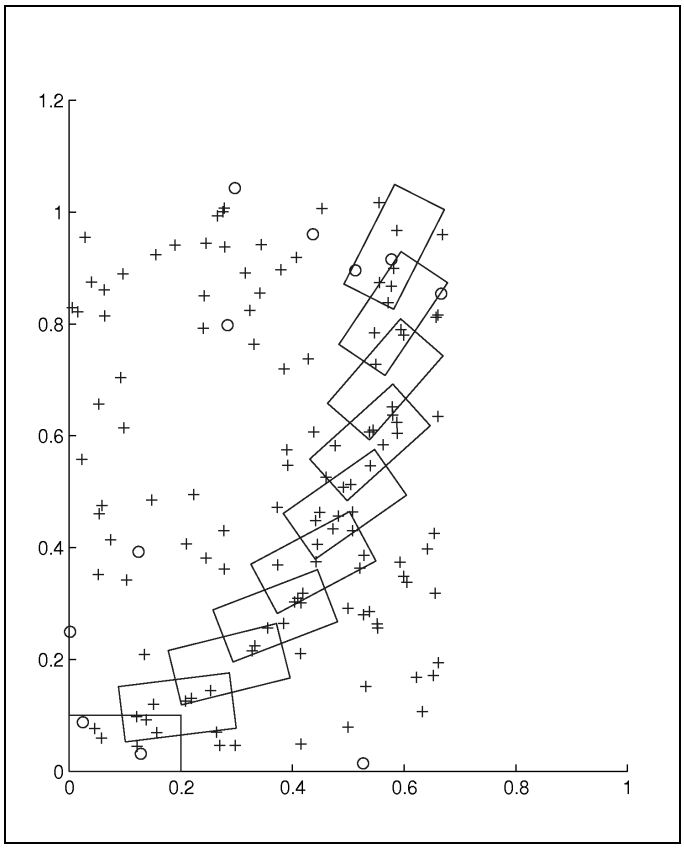

Fig. 11.

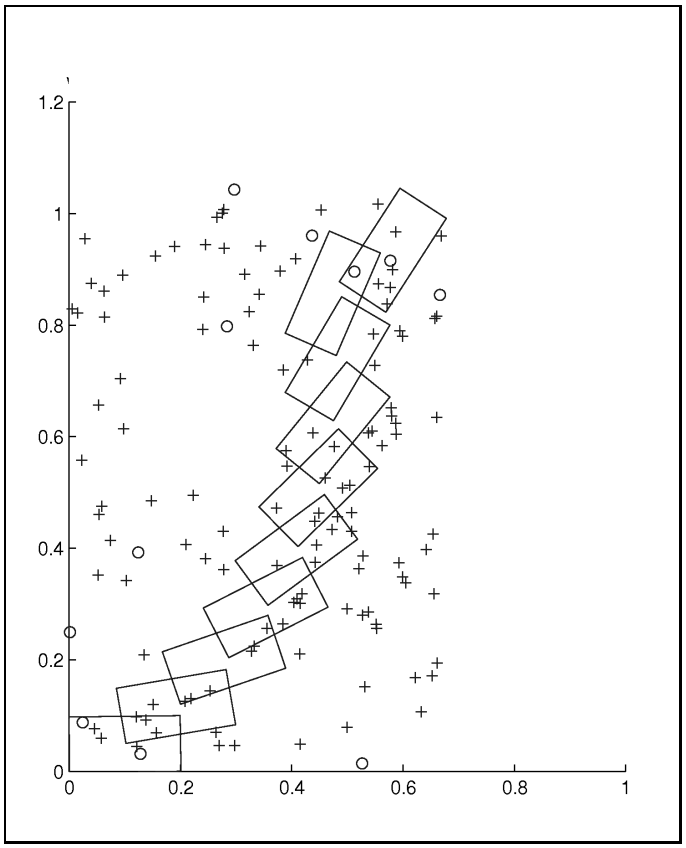

Fig. 13.

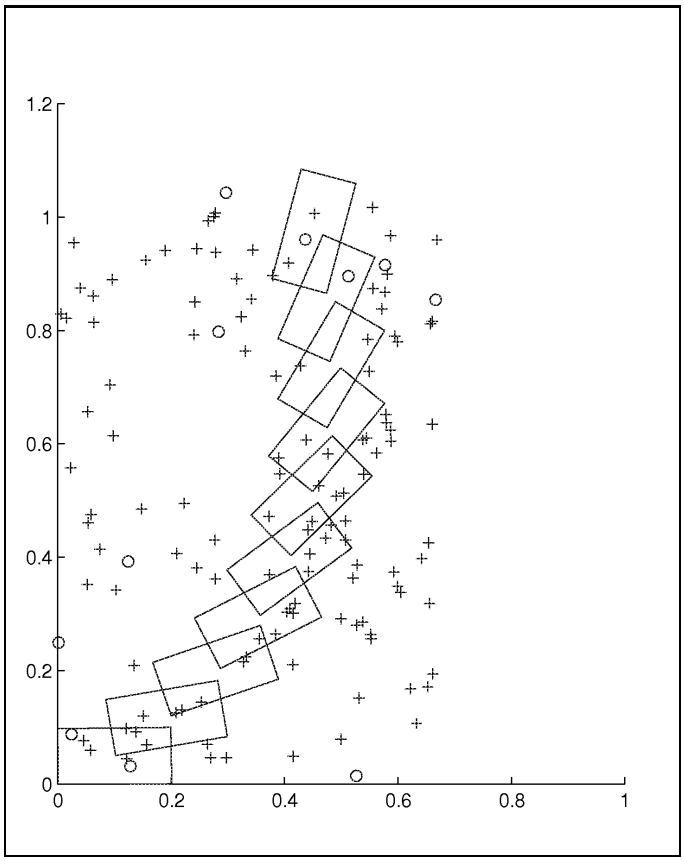

Fig. 12.

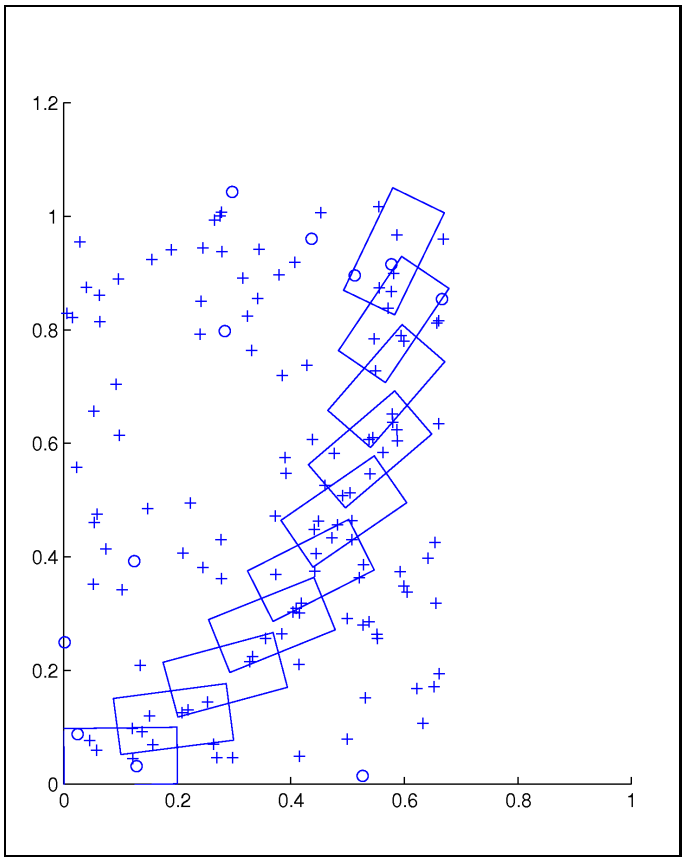

Fig. 14.