2008 10th Intl. Conf. on Control, Automation, Robotics and Vision
Hanoi, Vietnam, 17–20 December 2008

Detecting Spurious Features using Parity Space

David Törnqvist, Thomas B. Schön and Fredrik Gustafsson
Division of Automatic Control,
Linköping University
SE–581 83 Linköping, Sweden
{tornqvist, schon, fredrik}@isy.liu.se

***Abstract*—Detection of spurious features is instrumental in many computer vision applications. The standard approach is feature based, where extracted features are matched between the image frames. This approach requires only vision, but is computer intensive and not yet suitable for real-time applications. We propose an alternative based on algorithms from the statistical fault detection literature. It is based on image data and an inertial measurement unit (IMU). The principle of analytical redundancy is applied to batches of measurements from a sliding time window. The resulting algorithm is fast and scalable, and requires only feature positions as inputs from the computer vision system. It is also pointed out that the algorithm can be extended to also detect non-stationary features (moving targets for instance). The algorithm is applied to real data from an unmanned aerial vehicle in a navigation application.**

I. Introduction

Computer vision algorithms are today used in a large number of applications within areas of tracking, navigation, augmented reality and simultaneous localization and mapping (SLAM). In common for these applications is that features in the image frames are detected and used as observations in a (Kalman) filter to estimate camera pose and/or target pose. For a feature to be useful, it has to be stationary in space and clearly shaped so its position is unambiguously computed in consecutive frames by the feature detection algorithm. If the latter is not the case, the feature is called a *spurious* feature.

The classical approach is a stand-alone computer vision feature-based solution. One standard approach for avoiding spurious features is to use the so called random sample consensus (RANSAC) algorithm [2]. A problem with this algorithm is that it is not directly suitable for real-time algorithms. Some interesting recent work on reducing the computational complexity have been presented in [13]. To mention a few applications, where RANSAC type outlier detection is used we have [8], [12], [14]. For image to image comparisons the difference between the predicted feature location and its actual location can be used to detect outliers. Another idea is to make use of the epipolar geometry to detect outliers, simply by computing the distance to the epipolar lies for each correspondence. If this distance is larger than a certain threshold the feature is detected as being a spurious feature [16].

For real-time applications with limited computational resources, methods based on correlation maximization of patch templates in each frame for a large number of features should be avoided.

The approach we suggest is prediction based, combining IMU measurements with a model to obtain the predictions. The detection part is an application of the stochastic parity space approach in fault detection. The stochastic parity space approach as described in [5] extends the classical parity space approach in fault detection [1] to stochastic measurement errors and process noise. The proposed algorithm consists of the following components:

- An IMU for predicting how the features move from frame to frame.
- A list of feature points for each frame is used as measurement vector. Here, for instance, the Harris detector [6] can be used.
- A sliding window where the feature points are monitored for outliers.
- A parity space approach where the feature observations are projected onto a smaller subspace which eliminates unknown inputs, attenuates the noise and highlights outliers.

In a way, the analytical redundancy in the parity space extends the concept of projective geometry to a sequence of frames, utilizing a temporal motion model for the camera pose.

The algorithm is applied to airborne images in a SLAM application, where the goal is to navigate an unmanned aerial vehicle (UAV) back to home-base without infrastructure, as described in [9]. This case is particularly challenging, since the ground scene is rich of spurious features. Yet, the algorithm is quite successful in detecting spurious features.

The outline is as follows. Section II surveys the stochastic parity space approach to fault detection in linear systems. Section III presents a kinematic model suitable for fusion of IMU and camera information, and gives a state space model suitable for the parity space approach. Section IV illustrates the method on artificial data from the realistic scenario, on which the method is evaluated on real data in Section V.

II. Parity Space

If a model of a system is available, it is possible to determine in which part of the output space an output from a healthy system can reside. If there is signal energy in other parts of the output space, it is not in accordance with the model and a fault must therefore be present. This is exploited within the parity space framework [1], where an orthogonal projection of the output is used to create a residual that is non-sensitive to healthy outputs.

A. Modeling

Let us consider a discrete-time dynamical system in the form

$$x_{t+1} = F_t x_t + G_t^u u_t + G_t^f f_t + G_t^v v_t, \tag{1a}$$

$$y_t = H_t x_t + H_t^u u_t + H_t^f f_t + e_t, \tag{1b}$$

where x_t is the state vector, u_t an input signal, f_t a fault vector, y_t a measurement, v_t process noise and e_t the measurement noise. Since we are interested in monitoring how the system behaves over time, the system can be written in batch form. The batch form is often used in fault detection and diagnosis [1], [3], [4]. Stack L signal values to define the batched signal vectors like $\mathbb{Y} = \left(y_{t-L+1}^T, \ldots, y_t^T\right)^T$, for all signals. Then, the outputs in that window will be given by

$$\mathbb{Y} = \mathcal{O}_t x_{t-L+1} + \bar{H}_t^u \mathbb{U} + \bar{H}_t^f \mathbb{F} + \bar{H}_t^v \mathbb{V} + \mathbb{E}, \tag{2}$$

with the extended observability matrix

$$\mathcal{O}_t = \begin{pmatrix} H_{t-L+1} \\ H_{t-L+2} F_{t-L+1} \\ \vdots \\ H_t \prod_{k=t-1}^{t-L+1} F_k \end{pmatrix}. \tag{3}$$

The matrices determining how the remaining signals affect the system are described by

$$\bar{H}_t^s = \begin{pmatrix} H_{t-L+1}^s & 0 & \cdots & 0 \\ H_{t-L+2} G_{t-L+1}^s & H_{t-L+2}^s & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ H_t \prod_{k=t-1}^{t-L+2} F_k G_{t-L+1}^s & \cdots & H_t G_{t-1}^s & H_t^s \end{pmatrix}, \tag{4}$$

where $s = \{u, f, v\}$. For simplicity, denote $\mathbb{Z} = \mathbb{Y} - \bar{H}_t^u \mathbb{U}$, implying that the system (2) can be written as

$$\mathbb{Z} = \mathcal{O}_t x_{t-L+1} + \bar{H}_t^f \mathbb{F} + \bar{H}_t^v \mathbb{V} + \mathbb{E}. \tag{5}$$

If the system can be written in the form (5), where $\mathsf{Cov}(\mathbb{V} + \mathbb{E}) = S$, we can compute an estimate of the initial state according to

$$\hat{x}_{t-L+1} = \mathcal{O}_t^\dagger \mathbb{Z}. \tag{6}$$

In this work, the standard parity space approach with orthogonal projection is used. This implies that the initial state in the test window is estimated using the pseudo-inverse as shown above. An interesting extension would be to make a minimum variance estimate of the initial state which corresponds to an oblique projection in the parity space framework, see [15]. Note that in the special case $S = \sigma I$, (6) is in fact the minimum variance estimate.

B. Residual Generation

A residual for detecting abnormal behavior of the output y_t can now be formed as

$$\tilde{r} = \mathbb{Z} - \mathcal{O}_t \hat{x}_{t-L+1} = \mathbb{Z} - \underbrace{\mathcal{O}_t \mathcal{O}_t^\dagger}_{\mathcal{P}_{\mathcal{O}_t}} \mathbb{Z} = \underbrace{(I - \mathcal{P}_{\mathcal{O}_t})}_{\mathcal{P}_{\mathcal{O}_t^\perp}} \mathbb{Z}. \tag{7}$$

Since $\mathcal{P}_{\mathcal{O}_t^\perp}$ is rank deficient, the information in $\tilde{r}$ can be expressed using a lower dimensional residual

$$r = \mathcal{B}_{\mathcal{O}^\perp}^T \mathbb{Z}, \tag{8}$$

where the columns of $\mathcal{B}_{\mathcal{O}^\perp}$ forms an orthonormal basis for $\mathcal{R}(\mathcal{O}^\perp)$ (This can be computed with an SVD). The residual r has the covariance

$$\mathsf{Cov}(r) = \mathcal{B}_{\mathcal{O}^\perp}^T S \, \mathcal{B}_{\mathcal{O}^\perp}. \tag{9}$$

A normalized residual is then given by

$$\bar{r} = \underbrace{(\mathcal{B}_{\mathcal{O}^\perp}^T S \, \mathcal{B}_{\mathcal{O}^\perp})^{-1/2} \mathcal{B}_{\mathcal{O}^\perp}^T}_{\triangleq W^T} \mathbb{Z}. \tag{10}$$

C. Statistical tests

To test whether a fault has occurred or not, the following hypothesis test is formed,

$$\mathcal{H}_0 : \ \bar{r} \sim N(0, I), \tag{11a}$$

$$\mathcal{H}_1 : \ \bar{r} \sim N(W^T \bar{H}^f \mathbb{F}, I). \tag{11b}$$

To decide between the hypotheses, a GLR test will be performed using the following test statistic

$$\begin{aligned} L &= 2 \log \left(\frac{\sup_{\mathbb{F}} p(\bar{r} - W^T \bar{H}^f \mathbb{F})}{p(\bar{r})} \right) \\ &= \sup_{\mathbb{F}} 2 \log \frac{\frac{1}{(2\pi)^{n/2}} e^{-\frac{1}{2}\|\bar{r} - W^T \bar{H}^f \mathbb{F}\|_2^2}}{\frac{1}{(2\pi)^{n/2}} e^{-\frac{1}{2}\|\bar{r}\|_2^2}} \\ &= \sup_{\mathbb{F}} -\|\bar{r} - W^T \bar{H}^f \mathbb{F}\|_2^2 + \|\bar{r}\|_2^2, \end{aligned} \tag{12}$$

where n denotes the dimension of $\bar{r}$. This function is maximized for $\mathbb{F} = (W^T \bar{H}^f)^\dagger \bar{r}$, which gives

$$\begin{aligned} L &= -\|\bar{r} - \underbrace{W^T \bar{H}^f (W^T \bar{H}^f)^\dagger}_{\mathcal{P}_{W^T \bar{H}^f}} \bar{r}\|_2^2 + \|\bar{r}\|_2^2 \\ &= \bar{r}^T \mathcal{P}_{W^T \bar{H}^f} \bar{r}. \end{aligned} \tag{13}$$

To choose suitable thresholds for the test statistics above, it is necessary to compute their distributions. While having Gaussian noise, the test statistics will be chi-square distributed variables. For a thorough review of statistics in signal processing, see e.g., [10]. The normalized residual is distributed as

$$\bar{r} \sim \mathsf{N}(W^T \bar{H}^f \mathbb{F}, I), \tag{14}$$

where $\mathbb{F} = 0$ under the null hypothesis (11). The test statistic is then distributed as the non-central χ^2-distribution

$$L = \bar{r}^T \mathcal{P}_{W^T \bar{H}^f} \bar{r} \sim \chi_\nu'^2(\lambda), \tag{15}$$

where $\nu = \mathrm{rank}(W^T \bar{H}^f)$ and

$$\lambda = (W^T \bar{H}^f \mathbb{F})^T \mathcal{P}_{W^T \bar{H}^f} W^T \bar{H}^f \mathbb{F} = (W^T \bar{H}^f \mathbb{F})^T W^T \bar{H}^f \mathbb{F}. \tag{16}$$

Note that $\lambda = 0$ in the fault-free case and then the test statistic is distributed according to the central χ^2-distribution $L \sim \chi_\nu^2$. The threshold is then chosen from the χ^2-distribution so that the fault-free hypothesis is rejected erroneously only with a small probability.

III. Detection of Spurious Features

In this work we assume that stationary features are observed through a pinhole camera. As states in the model, the three-dimensional position of the feature $m_{i,t}$ is used and the full state vector is $m_t = \left(m_{1,t}^T, \ldots, m_{n_m,t}^T\right)^T$. The pinhole camera [11], [7] is assumed to be calibrated and have unit focal length. The system equations can be written as

$$m_{t+1} = m_t, \tag{17a}$$

$$\begin{aligned} y_{i,t} &= \underbrace{\frac{1}{\mathsf{z}_{i,t}^c}\begin{pmatrix}\mathsf{x}_{i,t}^c \\ \mathsf{y}_{i,t}^c\end{pmatrix}}_{h_i(m_{i,t}^c)=h_i(m_{i,t},p_t)} + e_{i,t} \\ &= \frac{1}{\mathsf{z}_{i,t}^c}\begin{pmatrix} I & 0\end{pmatrix} R_t(m_{i,t}-p_t) + e_{i,t}, \end{aligned} \tag{17b}$$

where

$$m_{i,t}^c = \begin{pmatrix}\mathsf{x}_{i,t}^c \\ \mathsf{y}_{i,t}^c \\ \mathsf{z}_{i,t}^c\end{pmatrix} = R_t(m_{i,t}-p_t). \tag{18}$$

Here, R_t is the rotation matrix between the world and the camera coordinate frame, p_t is the translation between the coordinate frames and $m_{i,t}^c$ the feature position in the camera coordinate frame. This model could be extended to include dynamics for m_t if we want to model moving objects as well. To use the parity space framework, the measurement equation is simply linearized according to,

$$y_{i,t} = h_i(m_{i,t}^c) + e_{i,t} \approx h_i(\hat{m}_t^c) + H_{i,t}(m_{i,t}^c - \hat{m}_{i,t}^c) + e_{i,t}^c, \tag{19}$$

for $j = 1, \ldots, n_m$, where $e_{i,t}^c$ is the noise from the camera sensor and the Jacobian matrix $H_{i,t}$ is

$$H_{i,t} = \left.\frac{\partial h_i(m_{i,t}^c)}{\partial m_{i,t}^c}\right|_{m_{i,t}^c=\hat{m}_{i,t}^c} = \frac{1}{\hat{\mathsf{z}}_{i,t}^c}\begin{pmatrix} 1 & 0 & -\frac{\hat{\mathsf{x}}_{i,t}^c}{\hat{\mathsf{z}}_{i,t}^c} \\ 0 & 1 & -\frac{\hat{\mathsf{y}}_{i,t}^c}{\hat{\mathsf{z}}_{i,t}^c}\end{pmatrix}. \tag{20}$$

Note that $\hat{m}_{i,t}^c = \begin{pmatrix}\hat{\mathsf{x}}_{i,t}^c & \hat{\mathsf{y}}_{i,t}^c & \hat{\mathsf{z}}_{i,t}^c\end{pmatrix}^T$ denotes the one-step ahead prediction from the filter. Evaluation of (19) gives

$$\begin{aligned} y_{i,t} &\approx h_i(\hat{m}_{i,t}^c) + H_{i,t}(m_{i,t}^c - \hat{m}_{i,t}^c) + e_{i,t} \\ &= \frac{1}{\hat{\mathsf{z}}_{i,t}^c}\begin{pmatrix}\mathsf{x}_{i,t}^c \\ \mathsf{y}_{i,t}^c\end{pmatrix} + \underbrace{\frac{\hat{\mathsf{z}}_{i,t}^c - \mathsf{z}_{i,t}^c}{\hat{\mathsf{z}}_{i,t}^c}\begin{pmatrix}\hat{\mathsf{x}}_{i,t}^c/\hat{\mathsf{z}}_{i,t}^c \\ \hat{\mathsf{y}}_{i,t}^c/\hat{\mathsf{z}}_{i,t}^c\end{pmatrix}}_{e_{i,t}^z} + e_{i,t}^c. \end{aligned} \tag{21}$$

The term $e_{i,t}^z$ has zero mean and covariance

$$\mathsf{Cov}(e_{i,t}^z) = \frac{P_{\mathsf{z}_{i,t}^c}}{(\hat{\mathsf{z}}_{i,t}^c)^2}\begin{pmatrix}\hat{\mathsf{x}}_{i,t}^c/\hat{\mathsf{z}}_{i,t}^c \\ \hat{\mathsf{y}}_{i,t}^c/\hat{\mathsf{z}}_{i,t}^c\end{pmatrix}\begin{pmatrix}\hat{\mathsf{x}}_{i,t}^c/\hat{\mathsf{z}}_{i,t}^c \\ \hat{\mathsf{y}}_{i,t}^c/\hat{\mathsf{z}}_{i,t}^c\end{pmatrix}^T, \tag{22}$$

where $P_{\mathsf{z}_{i,t}^c} = \mathsf{Cov}(\hat{\mathsf{z}}_{i,t}^c)$. Hence, we can make the approximation that this term is only a contribution to the measurement noise. Define the total measurement noise as $e_{i,t} = e_{i,t}^z + e_{i,t}^c$. Using this approximation, the measurement equation can be written as

$$y_{i,t} \approx \frac{1}{\hat{\mathsf{z}}_{i,t}^c}\begin{pmatrix}\mathsf{x}_{i,t}^c \\ \mathsf{y}_{i,t}^c\end{pmatrix} + e_{i,t} = \frac{1}{\hat{\mathsf{z}}_{i,t}^c}\begin{pmatrix} I & 0\end{pmatrix} R_t(m_{i,t}-p_t) + e_{i,t}. \tag{23}$$

Now, define $\hat{Z} = \mathrm{diag}(\hat{\mathsf{z}}_{1,t}^c I_2, \hat{\mathsf{z}}_{2,t}^c I_2, \ldots)$ and premultiply the total measurement equation with this matrix. This equation can then be written as

$$\begin{aligned} \hat{Z}y_t = &\underbrace{\mathrm{diag}\left(\begin{pmatrix} I & 0\end{pmatrix} R_t, \begin{pmatrix} I & 0\end{pmatrix} R_t, \ldots\right)}_{H_t} m_t \\ &- \underbrace{\begin{pmatrix}\begin{pmatrix} I & 0\end{pmatrix} R_t \\ \begin{pmatrix} I & 0\end{pmatrix} R_t \\ \vdots \end{pmatrix}}_{H_t^u} \underbrace{p_t}_{u_t} + \hat{Z}e_t, \end{aligned} \tag{24}$$

This linearized model can be used in the parity space framework from Section II, use the notation $x_t = m_t$, $F_t = I$ and the rest of the definitions in (24).

A. Known Translation

In case the translation is known, it is modeled as a deterministic input. The system can be written in batch form (2), but using $\mathbb{Y}$ as stacked $\hat{Z}y_t$ and observing that $\bar{H}^v\mathbb{V} = 0$ since the system has no process noise (17a),

$$\mathbb{Y} = \mathcal{O}x_{t-L+1} + \bar{H}^u\mathbb{U} + \bar{H}^f\mathbb{F} + \mathbb{E}. \tag{25}$$

The residual is then formed according to Section II-B

$$\bar{r}_1 = \underbrace{(\mathcal{B}_{\mathcal{O}^\perp}^T S\, \mathcal{B}_{\mathcal{O}^\perp})^{-1/2}\, \mathcal{B}_{\mathcal{O}^\perp}^T}_{\triangleq W^T}\mathbb{Z}, \tag{26}$$

where $\mathbb{Z} = \mathbb{Y} - \bar{H}^u\mathbb{U}$. For $n_m = 3$, $L = 3$ and $R_0 = I$ (without loss of generality), $\mathcal{O}$ will have the following form

$$\mathcal{O} = \begin{pmatrix} H_{t-2} \\ H_{t-1} \\ H_t \end{pmatrix} = \begin{pmatrix} \begin{pmatrix} I & 0\end{pmatrix} & 0 & 0 \\ 0 & \begin{pmatrix} I & 0\end{pmatrix} & 0 \\ 0 & 0 & \begin{pmatrix} I & 0\end{pmatrix} \\ \begin{pmatrix} I & 0\end{pmatrix} R_1 & 0 & 0 \\ 0 & \begin{pmatrix} I & 0\end{pmatrix} R_1 & 0 \\ 0 & 0 & \begin{pmatrix} I & 0\end{pmatrix} R_1 \\ \begin{pmatrix} I & 0\end{pmatrix} R_2 & 0 & 0 \\ 0 & \begin{pmatrix} I & 0\end{pmatrix} R_2 & 0 \\ 0 & 0 & \begin{pmatrix} I & 0\end{pmatrix} R_2 \end{pmatrix}. \tag{27}$$

The rank of this matrix is $3 \cdot 3 = 9$ and it is easy to see that the rank in the general case is $\mathrm{rank}(\mathcal{O}) = 3n_m$. The number of rows of $\mathcal{O}$ is $2n_mL$, so the dimension of the orthogonal column-space $\mathcal{O}^\perp$ is $2n_mL - 3n_m = (2L-3)n_m$. This is also the dimension of the residual $\bar{r}_1$ in (26).

B. Unknown Translation

When the translation is unknown, the predictions of the feature position in the camera coordinate frame will become more uncertain. The covariance in (22) depends on these predictions. In order to limit this we will make the following pragmatic approximation

$$\mathsf{Cov}(e_{i,t}^z) = \frac{P_{\mathsf{z}_{i,t}^c}}{(\hat{\mathsf{z}}_{i,t}^c)^2}\begin{pmatrix}\hat{\mathsf{x}}_{i,t}^c/\hat{\mathsf{z}}_{i,t}^c \\ \hat{\mathsf{y}}_{i,t}^c/\hat{\mathsf{z}}_{i,t}^c\end{pmatrix}\begin{pmatrix}\hat{\mathsf{x}}_{i,t}^c/\hat{\mathsf{z}}_{i,t}^c \\ \hat{\mathsf{y}}_{i,t}^c/\hat{\mathsf{z}}_{i,t}^c\end{pmatrix}^T \approx \frac{P_z}{(\hat{\mathsf{z}}_{i,t}^c)^2} y_{i,t}y_{i,t}^T. \tag{28}$$

This can be motivated by the fact that the sensor noise is fairly small and this noise term depends on the position of the feature in the image, see section IV for an illustration. In the

detection problem, the translation can no longer be seen as a known input to the system and the residual should therefore be insensitive to it. Hence, the batched system is rewritten as

$$\mathbb{Y} = \mathcal{O}x + \bar{H}^u\mathbb{U} + \bar{H}^f\mathbb{F} + \mathbb{E} = \underbrace{\begin{pmatrix}\mathcal{O} & \bar{H}^u\end{pmatrix}}_{\tilde{\mathcal{O}}}\begin{pmatrix}x\\ \mathbb{U}\end{pmatrix} + \bar{H}^f\mathbb{F} + \mathbb{E}. \tag{29}$$

A residual can now be constructed by an orthogonal projection onto $\tilde{\mathcal{O}}$ as

$$\bar{r}_2 = \underbrace{(\mathcal{B}_{\tilde{\mathcal{O}}^\perp}^T\, S\, \mathcal{B}_{\tilde{\mathcal{O}}^\perp})^{-1/2}\, \mathcal{B}_{\tilde{\mathcal{O}}^\perp}^T}_{\triangleq \tilde{W}^T}\, \mathbb{Y}, \tag{30}$$

where the columns of the orthonormal matrix $\mathcal{B}_{\tilde{\mathcal{O}}^\perp}$ spans the orthogonal image space to $\tilde{\mathcal{O}}$. For $n_m = 3$, $L = 3$ and $R_0 = I$ (without loss of generality), $\tilde{\mathcal{O}}$ would have the following form

$$\tilde{\mathcal{O}} = \begin{pmatrix}\mathcal{O} & \mathcal{O}_2\end{pmatrix} \tag{31}$$

where $\mathcal{O}$ was given in (27) and

$$\mathcal{O}_2 = \begin{pmatrix} \begin{pmatrix}I & 0\end{pmatrix} & 0 & 0 \\ \begin{pmatrix}I & 0\end{pmatrix} & 0 & 0 \\ \begin{pmatrix}I & 0\end{pmatrix} & 0 & 0 \\ 0 & \begin{pmatrix}I & 0\end{pmatrix} R_1 & 0 \\ 0 & \begin{pmatrix}I & 0\end{pmatrix} R_1 & 0 \\ 0 & \begin{pmatrix}I & 0\end{pmatrix} R_1 & 0 \\ 0 & 0 & \begin{pmatrix}I & 0\end{pmatrix} R_2 \\ 0 & 0 & \begin{pmatrix}I & 0\end{pmatrix} R_2 \\ 0 & 0 & \begin{pmatrix}I & 0\end{pmatrix} R_2 \end{pmatrix} \tag{32}$$

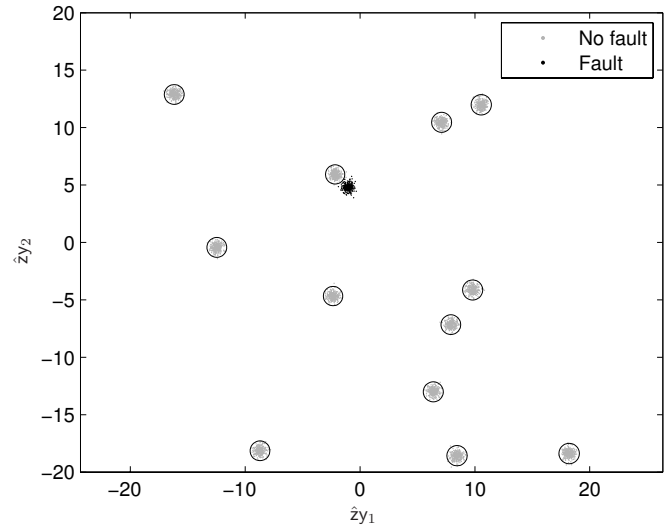
Using row operations, (32) can be transformed into

$$\begin{pmatrix} 0 & 0 & 0 & \begin{pmatrix}I & 0\end{pmatrix} & 0 & 0 \\ \begin{pmatrix}I & 0\end{pmatrix} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \begin{pmatrix}I & 0\end{pmatrix} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \begin{pmatrix}I & 0\end{pmatrix} R_1 & 0 \\ \begin{pmatrix}I & 0\end{pmatrix} R_1 & 0 & 0 & 0 & 0 & 0 \\ 0 & & \begin{pmatrix}I & 0\end{pmatrix} R_1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \begin{pmatrix}I & 0\end{pmatrix} R_2 \\ \begin{pmatrix}I & 0\end{pmatrix} R_2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \begin{pmatrix}I & 0\end{pmatrix} R_2 & 0 & 0 & 0 \end{pmatrix} \tag{33}$$

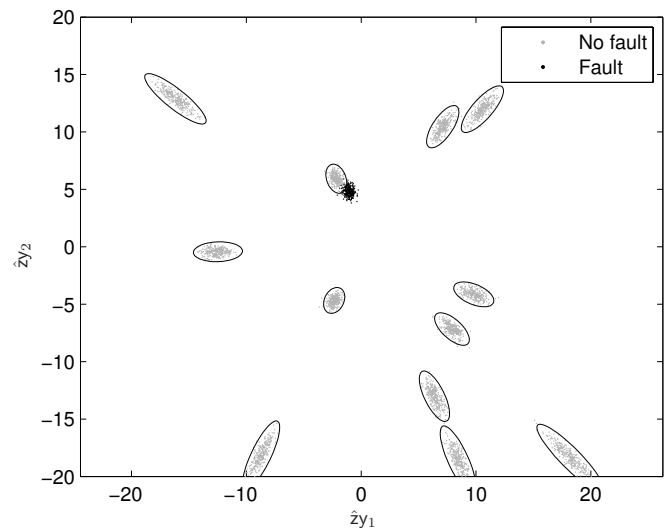
The rank of this matrix is $3 \cdot 2 + 2 \cdot 3 = 12$, or in the general case $3(n_m - 1) + 2L$. Since the number of rows are $2n_c l$, the dimension of the orthogonal image space to $\tilde{\mathcal{O}}$ will be $2n_m L - 3(n_m - 1) + 2L = (2L - 3)n_m - 2L + 3$. This will also be the dimension of the residual $\bar{r}_2$ in (30).

IV. Simulations

For the simulations, a map of features similar to the real flight scenario in Section V is created. The scenario is a helicopter flying at an altitude of 58 m, with a downward-looking camera attached to it. The features are uniformly spread out over an area of 40×40 m and the altitude is generated from a Gaussian distribution $\mathsf{N}(58, 1)$. A fault is introduced with the direction $k(-1 \quad 1)^T$ for one of the features, the magnitude k is varied between the values $\{0, 0.02, 0.03\}$. The measurement noise is distributed as $e_{i,t} \sim \mathsf{N}(0, 0.005^2)$. Figure 1 shows 300 measurements $\hat{Z}y_t$ as in (24), both with exact knowledge of $\mathsf{z}_{i,t}^c$ ($\mathsf{Cov}(\hat{\mathsf{z}}_{i,t}^c) = 0$) and when there is uncertainty present ($\mathsf{Cov}(\hat{\mathsf{z}}_{i,t}^c) = 3^2$). The difference between these cases is if the noise term $e_{i,t}^{\mathsf{z}}$ in (22) is present or not. A faulty measurement is also shown, where the fault magnitude is 0.02. It can be noted that the measurement noise is increased radially when $\mathsf{Cov}(\hat{\mathsf{z}}_{i,t}^c)$ is increased. To make diagnosis of spurious features possible, the features are divided into groups of three. Detection is then made for each group and if a fault is detected, the measurements from the group are discarded.



(a) $\mathsf{Cov}(\hat{\mathsf{z}}_{i,t}^c) = 0$, i.e., $\hat{\mathsf{z}}_{i,t}^c = \mathsf{z}_{i,t}^c$



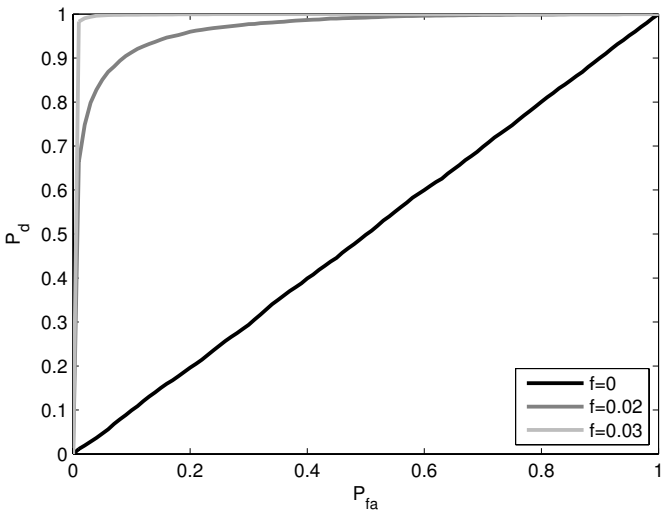
(b) $\mathsf{Cov}(\hat{\mathsf{z}}_{i,t}^c) = 3^2$

Figure 1. Simulated measurements $\hat{Z}y_t$ of the features are shown (300 realizations). The ellipsoids show the confidence region corresponding to 3 standard deviations.

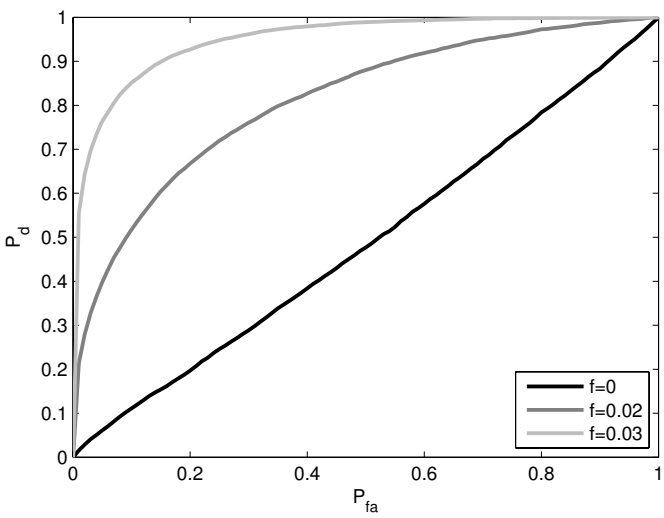
A. Known Translation

If the translation is known, the dimension of the residual will be $(2L - 3)n_m = 3$. This will also be the degrees of freedom in the χ^2-test used for detection, see Section II-C. In Figure 2, the Receiver Operation Characteristics (ROC) curves are plotted. The ROC-curve shows how the probability of

detection P_{d} for a fault varies versus the probability of false alarm P_{fa}. The ROC-curve is a straight line for the no fault case. This is natural, since $P_{\mathrm{d}} = P_{\mathrm{fa}}$ for any threshold if there is nothing to detect. It can be noted that the probability of detection decreases with uncertainty of z^c. This is due to the increased measurement noise, illustrated in Figure 1.



(a) $\mathsf{Cov}(\hat{\mathsf{z}}^c_{i,t}) = 0$, i.e., $\hat{\mathsf{z}}^c_{i,t} = \mathsf{z}^c_{i,t}$

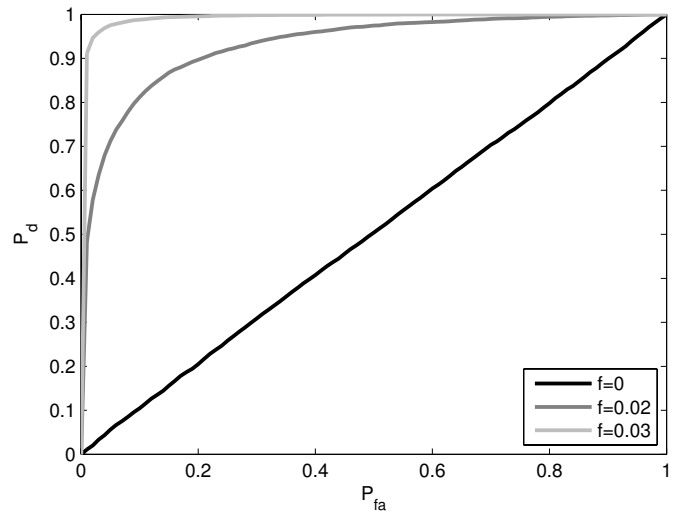


(b) $\mathsf{Cov}(\hat{\mathsf{z}}^c_{i,t}) = 3^2$

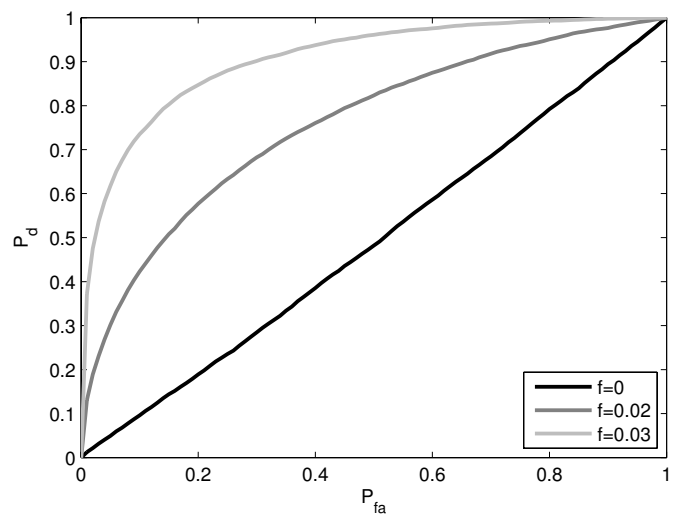
Figure 2. ROC-curves showing the detection performance with known translation.

B. Unknown Translation

In the case of unknown translation, the degrees of freedom in the χ^2-test is decreased to $(2L-3)n_m - 2L + 3 = 2$. This somewhat reduces the power of the test which can be seen in Figure 3. Given a certain P_{fa}, the P_{d} is lower here compared to the case of known translation in Figure 2. The strength with this test is that no knowledge of the translation is needed. In reality, the estimate of the position change between two frames is often poor.



(a) $\mathsf{Cov}(\hat{\mathsf{z}}^c_{i,t}) = 0$, i.e., $\mathsf{z}^c_{i,t} = \mathsf{z}^c_{i,t}$



(b) $\mathsf{Cov}(\hat{\mathsf{z}}^c_{i,t}) = 3^2$

Figure 3. ROC-curves showing the detection performance with unknown translation.

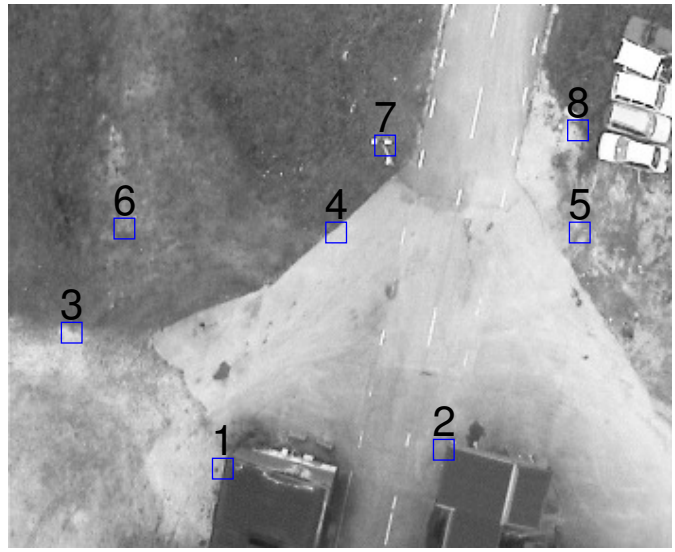
V. Real World Experiment

In this example, simultaneous localization and mapping (SLAM) is performed using a Yamaha RMAX helicopter with an IMU and a camera looking at the ground. The task is to be able to localize the helicopter by combining inertial measurements with the camera image. During the process, landmarks are found in the image and the 3D positions of these landmarks are estimated. The setup and the SLAM algorithm are explained in [9].

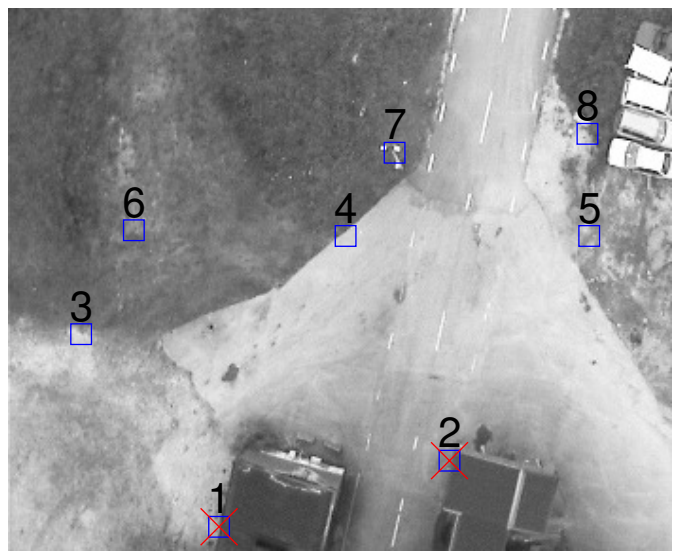
The detection of spurious features within the SLAM framework is the focus of the present work. Features in two consecutive frames are compared. The IMU computes the

change in rotation of the camera between the frames, but the translation is considered unknown since the integrated accelerations gives high uncertainty. The theory derived in Section III-B can therefore be applied. To form the test, the features are divided into several groups. If a test indicates that there is an inconsistency in one of the groups, these features will be discarded.

The image sequence and IMU data that we have used for this paper are collected when flying a 300 m distance at around 60 m altitude. Without using this detection method, the estimated position diverges, since spurious features contaminate the filter. When spurious features are detected, the position is estimated without problems. Figure 4 shows an example of when spurious features are detected. The eight features in view are grouped into three groups and a consistency test between image one and two is performed for each group. In this example, the group consisting of feature 1 and 2 is not consistent and the features are therefore discarded. Feature 1 is detected on an edge and the feature tracker makes a common mistake, which is to move the feature along this edge.



(a) Features detected in the first image.



(b) Features detected in the second image, the group consisting of feature 1 and 2 are detected as inconsistent with the first image.

Figure 4. Example of spurious feature detection.

VI. Conclusions and Future Work

The contribution of this paper has been to show how the parity space framework from the statistical fault detection literature can be applied to find spurious features. The characteristics and performance have been shown in simulation examples. Furthermore, it has also been successfully implemented in a real-world example where it is necessary to estimate the helicopter position.

Acknowledgment

This work was supported by the strategic research center MOVIII, funded by the Swedish Foundation for Strategic Research, SSF.

References

[1] A. Y. Chow and A. S. Willsky. Analytical redundancy and the design of robust failure detection systems. *IEEE Transactions on Automatic Control*, 29(7):603–614, 1984.

[2] Martin A. Fischler and Robert C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM*, 24(6):381–395, 1981.

[3] J. J. Gertler. *Fault Detection and Diagnosis in Engineering Systems*. Marcel Dekker, Inc., 1998.

[4] F. Gustafsson. *Adaptive Filtering and Change Detection*. John Wiley & Sons, New York, USA, 2000.

[5] Fredrik Gustafsson. Statistical signal processing approaches to fault detection. *Annual Reviews in Control*, 31(1):41–54, 2007.

[6] C. Harris and M. Stephens. A combined corner and edge detector. In *Proceedings of the 4th Alvey Vision Conference*, pages 147–151, Manchester, UK, 1988.

[7] R. Hartley and A. Zisserman. *Multiple View Geometry in computer vision*. Cambridge University Press, Cambridge, UK, 2 edition, 2003.

[8] T. Kanade, A. Amidi, and Q. Ke. Real-time and 3D vision for autonomous small and micro air vehicles. In *Proceedings of the 43th Conference on Decision and Control (CDC)*, Paradise Island, Bahamas, December 2004.

[9] R. Karlsson, T. B. Schön, D. Törnqvist, G. Conte, and F. Gustafsson. Utilizing model structure for efficient simultaneous localization and mapping for a UAV application. In *Proceedings of IEEE Aerospace Conference*, Big Sky, MT, USA, March 2008.

[10] Steven M. Kay. *Fundamentals of Statistical Signal Processing: Detection Theory*, volume 2. Prentice-Hall, Inc, 1998.

[11] Y. Ma, S. Soatto, J. Kosecka, and S. S. Sastry. *An invitation to 3-D vision – from images to geometric models*. Interdisciplinary Applied Mathematics. Springer, 2006.

[12] M. Maimone, Y. Cheng, and L. Matthies. Two years of visual odometry on the mars exploration rovers. *Journal of Field Robotics*, 24(3):169–186, 2007.

[13] D. Nistér. Preemptive RANSAC for live structure and mostion estimation. *Machine Vision and Applications*, 16(5):321–329, December 2005.

[14] D. Nistér, O. Neroditsky, and J. Bergen. Visual odometry for ground vehicle applications. *Journal of Field Robotics*, 23(1):3–20, 2006.

[15] David Törnqvist and Fredrik Gustafsson. Eliminating the initial state for the generalized likelihood ratio test. In *Proceedings of IFAC Symposium SAFEPROCESS*, Beijing, China, August 2006.

[16] P. H. S. Torr, A. Zisserman, and S. J. Maybank. Robust detection of degenerate configurations for the fundamental matrix. In *Proceedings of the 5th International Conferences on Computer Vision*, pages 1037–1042, Boston, USA, 1995.