Multiple Input Structures with Robust Estimator  MISRE

Xiang Yang    Peter Meer    Jonathan Meer

convolutional neural network  parametric estimation
Robust parametric estimation estimates a *mathematical relation* for the inlier's objective function using a given algorithm with no training necessary.

But

* the user has to give threshold(s) before the estimation and do not process independently the structures.

Should be this way:

a *set* of 2D images viewed in 3D and segmented with 3D spheres *without* task depended thresholds.
Input measurements \( y_i = [y_{i,1} \ y_{i,2} \ldots y_{i,l}]^\top \). Objective function \( f(y_i) \).

\( f(y_i) \) is solved by linearization and elemental subsets.

The objective function \( f(y_i) \) is regarded as a linear relation

\[
f(y_i) \rightarrow x_i^\top \theta - \alpha \quad i = 1, \ldots, n
\]

The carrier vector \( x_i \) contains both the \( y_i \) and \( y_{i,j}y_{i,k} \) (computer vision) and the \( l \) unknowns in \( f(y_i) \) give rise to \( m \) variables in \( \theta \).

An elemental subset is the minimum number of equations needed for the solution of \( m \) variables. Number of points chosen = \( m \), for the moment.

\[
x_i^\top \theta - \alpha = 0 \quad i = 1, \ldots, m.
\]

elemental subset \( \rightarrow \theta, \alpha \quad \text{Ambiguity is reduced if } ||\theta|| = 1. \)
example: Fundamental matrix between two 2D images uses object point correspondences to solve for the $3 \times 3$ matrix $F$

$$f(y) = [x' \ y' \ 1] \ F \ [x \ y \ 1]^\top \quad y = [x \ y \ x' \ y']^\top$$

gives eight carriers $x = [x \ y \ x' \ y' \ xx' \ xy' \ x'y \ yy']^\top$

matrix $F \rightarrow$ vector $\theta$ and scalar $\alpha$

$$x_i^\top \ \theta - \alpha = 0 \quad i = 1 \ldots 8 \quad ||\theta|| = 1 \quad \text{an elemental subset}$$
example: 2D ellipse

The original variables \( y = [x \ y]^\top \) are on the ellipse

\[
f(y) = (y - y_c)^\top Q(y - y_c) - 1
\]

if \( Q \) is a \( 2 \times 2 \) positive definite symmetric matrix \( y_c \) is the ellipse center. \( y_c \rightarrow \theta_1, \theta_2 \). \( Q \rightarrow \theta_3, \theta_4, \theta_5 \).

The carrier vector \( x = [x \ y \ x^2 \ xy \ y^2]^\top \) gives

\[
\theta_1 x + \theta_2 y + \theta_3 x^2 + \theta_4 xy + \theta_5 y^2 - \alpha
\]

and a valid elemental subset must satisfy

\[
4\theta_3 \theta_5 - \theta_4^2 > 0
\]

If the inlier scale is given, both ellipses will be entirely recovered only if all the inlier points have similar noise.
The user has to give \textit{before} the estimation $M$, the number of elemental subsets...

...and the \textit{inlier scale}. 
RANSAC

given: inlier scale; \( M \) number of trials

Repeat \( M \) times:

* choose an elemental subset
* find the linear model estimate
* assume the estimate valid for all \( n \) points
* distances less than the scale are inliers.

Largest consensus set gives the RANSAC estimate.

Do Total Least Squares (TLS) with all the inliers: \( \hat{\theta}^{tls}, \hat{\alpha}^{tls} \).

Project back to the original space and find the original estimates.

**example:**

2D line estimation

elemental subset has two points

\[ \theta_1 x + \theta_2 y - \alpha = 0 \]
RANSAC may fail
* if the scale is incorrectly guessed by the user
* if an image sequence have big changes of the scale
* if the outliers are asymmetric
* if there are multiple inlier structures

A single scale is not enough in many cases.
Multiple Input Structures with Robust Estimator

MISRE

Each structure (inlier or outlier) estimated independently.

Each structure has three steps:
* scale estimation
* refinement with mean shift
* compute the structure's density

When the remaining data is not enough for a structure: STOP

Sort the structures based on the decreasing densities.

The user decides on the number of detected inlier structures.
Building the linear relation for the estimation

The \( l \) unknowns in the objective function \( f(y) \) have covariance \( \sigma^2 \mathbf{I}_{l \times l} \) where \( \sigma \) is unknown and different for each structure (iteration).

The input vector \( y_i \) can have several carrier vectors \( x_{i}^{[c]} \) \( c = 1 \ldots \zeta \). Each column of the \( m \times l \) Jacobian matrix \( J_{x_i^c | y_i} \) have the derivatives of the carrier vector in one original variable. \( x_{i}^{[c]} \sim J_{x_i^c | y_i} y_i \)

\( m \times m \) covariance of \( x_{i}^{[c]} \) is \( \sigma^2 C_{i}^{[c]} = \sigma^2 J_{x_i^c | y_i} J_{x_i^c | y_i}^\top \) \( c = 1 \ldots \zeta \)

example: \( \zeta = 1 \) fundamental matrix

\[
f(y) = \begin{bmatrix} x' & y' & 1 \end{bmatrix} F \begin{bmatrix} x & y & 1 \end{bmatrix}^\top
\]

\[
x = \begin{bmatrix} x & y & x' & y' & xx' & xy' & x'y & yy' \end{bmatrix}^\top
\]

\[
J_{x_i | y_i}^\top = \begin{bmatrix}
1 & 0 & 0 & 0 & x_i & y_i & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & x_i & y_i & 0 \\
0 & 0 & 1 & 0 & x_i & 0 & y_i & 0 \\
0 & 0 & 0 & 1 & 0 & x_i & 0 & y_i
\end{bmatrix}
\]
example: $\zeta = 2$

2D homography between two 2D images is a plane correspondence found through a $3 \times 3$ matrix

$$H = \begin{bmatrix} h_1 & h_2 & h_3 \end{bmatrix}^T \quad y = \begin{bmatrix} x & y & x' & y' \end{bmatrix}^T$$

$$\begin{bmatrix} x'_h \\ y'_h \\ w'_h \end{bmatrix} - \begin{bmatrix} h_1^T \\ h_2^T \\ h_3^T \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = 0 \quad x' = \frac{x y 1}{x y 1} h_1 \quad y' = \frac{x y 1}{x y 1} h_2$$

written with the unknown vec $H^T$ and the carriers $x^{[1]}$, $x^{[2]}$

$$\begin{bmatrix} -x & -y & -1 & 0 & 0 & 0 & x' x & x' y & x' \\ 0 & 0 & 0 & -x & -y & -1 & y' x & y' y & y' \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \\ h_3 \end{bmatrix} = 0$$

Jacobian matrices $J_{x_2|y_i}^{[1]}, J_{x_2|y_i}^{[2]}$ are $9 \times 4$ and $\alpha = 0$.

An elemental subset

$$\begin{bmatrix} \begin{bmatrix} x_1^{[1]} \end{bmatrix}^T \\ \begin{bmatrix} x_1^{[2]} \end{bmatrix}^T \\ \vdots \\ \begin{bmatrix} x_4^{[1]} \end{bmatrix}^T \\ \begin{bmatrix} x_4^{[2]} \end{bmatrix}^T \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \\ h_3 \end{bmatrix} = 0$$

has 8 equations.
Elemental subset $x_i^{[c]\top} \theta - \alpha = 0 \quad i = 1, \ldots, m \rightarrow \theta$ and $\alpha$

and projects $x_i^{[c]}$ into $z_i^{[c]} = x_i^{[c]\top} \theta \quad i = 1, \ldots, n$.

Mahalanobis distance from $\alpha$ is $d_i^{[c]} = \frac{|x_i^{[c]\top} \theta - \alpha|}{\sqrt{\theta \top C_i^{[c]} \theta}} \geq 0$

without the unknown $\sigma$.

For each $y_i$ retain only the largest Mahalanobis distance

$\tilde{c}_i = \arg \max_{c=1 \ldots \zeta} d_i^{[c]} \quad d_i^{[\tilde{c}_i]} = \tilde{d}_i = \frac{|\tilde{x}_i \top \theta - \alpha|}{\sqrt{\theta \top \tilde{C}_i \theta}} \geq 0$

Input $y_i$ corresponds to the carrier vector $\tilde{x}_i$ in the estimation.
Scale estimation
Refinement with mean shift
Compute the structure's density

All MISRE estimators use the same two constants.

The $M$ elemental subset trials are given by the user.

An iteration has $n \leq n_T$ datapoints
where $n_T$ is the total number of datapoints.

First constant

Each iteration starts from $n_\epsilon$ points which is the larger between:
* $0.05n_T$ from the total number of datapoints, or
* five times number of unknowns $m$ in an elemental subset.

A second condition quick-in only when the data is relative small
and the elemental subset uses a large carrier vector.
In each sequence the Mahalanobis distances are **ascendingly** ordered 
\[ \tilde{d}_{[i]}(j) \quad i = 1 \ldots n \]  
and are  \[ j = 1 \ldots M \]  sequences.

The **working sequence** \( \tilde{d}_{[i]}(w) \) is the one with the 
**minimum sum** of the Mahalanobis distances

\[
\min_{j \in M} \sum_{i=1}^{n_{\varepsilon}} \tilde{d}_{[i]}(j) \quad \text{for the first } n_{\varepsilon} \text{ points}
\]

with the elemental subset returning \( \hat{\theta}_w, \hat{\alpha}_w \).
example of how MISRE will be build:
two ellipses $n_{in} = 200$, $n_{out} = 200$. $M = 2000$ $n_e = 30$

$\sigma_g = 5$ or $10$
Gaussian standard deviations
inlier scale $\sim 2.5\sigma_g$

first iteration

number of datapoints $n_T = 200+200+200 = 600$
the starting number $n_e = 0.05 \times 600 = 30$ points
Divide the working sequence in equal Mahalanobis distances $\Delta d_\eta$.

$\Delta d_\eta$ have $n_k$ points in the $k$-th segment $k = 1, 2, \ldots$

$\Delta d_\eta$ corresponds to $\eta \geq 0.05n_T$ points starting from $\hat{\alpha}_w$.

Expansions are independent and increase with $0.01n_T$ each time.
Second constant

* If the average number of processed points in \( k \) segments is **larger than twice** the number of points in the \((k+1)\)-th segment, the expansion terminates.

This condition is verified for \( k = 1, 2, \ldots \) at each \( \Delta d_\eta \)

\[
\frac{1}{k} \sum_{i=1}^{k} n_i > 2 n_{k+1}
\]

\[\Delta d_5, k_{t_5} = 8, \hat{\sigma} = 8.06 \quad \Delta d_{10}, k_{t_{10}} = 5, \hat{\sigma} = 11.10 \]

scale should be \( \sim 2.5 \times 5 \approx 12.5 \)
Region of interest is defined from $n_\varepsilon$, corresponding to $\Delta d_5$ until the first $\eta\%$ where the second condition holds.

Largest expansion gives the scale estimate

$$\hat{\sigma} = \max_{\eta=5\%, \ldots, \eta_f} k_{t_{\eta}} \Delta d_{\eta}$$

$\hat{\sigma} = 12.54$ in the example. $n_{\hat{\sigma}}$ points between $\hat{\alpha}_w \pm \hat{\sigma}$. 
Scale estimation

Refinement with mean shift
Compute the structure's density

Mean shift is an iterative procedure finding the modes of the distribution function of a given window.

Interested in the mode \(*\) closest to the point marked \(\mid\) for \(\hat{\alpha}_w\).

Convergence achieved after only a few iterations.
Setting up the mean shift

\[
K(u) = \begin{cases} 
\frac{3}{4}(1 - u^2) & |u| \leq 1 \\
0 & \text{otherwise}
\end{cases}
\]

Epanechnikov kernel

\[
g(u) = -K'(u^2) = -\kappa'(u) = -\kappa' \left( (z - \tilde{z}_i)^\top \tilde{B}_i^{-1} (z - \tilde{z}_i) \right)
\]

\[
\kappa(u) = \begin{cases} 
1 - u & u \leq 1 \\
0 & u > 1
\end{cases}
\]

\[
g(u) = 1 \quad \text{from } n_\hat{\sigma} \text{ points choose another } N = M/10 \text{ elemental subsets.}
\]
Mean shift is applied to all $n$ points.

$$\arg \max_{\hat{\alpha}} \sum_{i=1}^{n} \kappa \left( (z - \tilde{z}_i)\top \tilde{B}_i^{-1} (z - \tilde{z}_i) \right) \quad z \to \hat{\alpha}$$

Current value is $z_{old}$ and all points $\tilde{z}_i$ contribute equally

$$z_{new} = \frac{\sum_{i=1}^{n} g(u_i) \tilde{z}_i}{\sum_{i=1}^{n} g(u_i)}$$

if $|z_{old} - \tilde{z}_i| \leq \sqrt{\tilde{B}_i}$ \quad $g(u_i) = 1$

if $|z_{old} - \tilde{z}_i| > \sqrt{\tilde{B}_i}$ \quad $g(u_i) = 0$

After $N$ trials, the window having the most points $\tilde{z}_i$ at the convergence is the mode $\hat{\alpha} = z_{final}$.

$n_{\hat{\alpha}}$ points converge to $\hat{\alpha}$. 
Nonrobust total least squares (TLS) estimates the structure from all the \( n_{\hat{\alpha}} \) points

\[
\hat{\mathbf{x}}_i^T \mathbf{\theta} - \alpha = 0 \quad i = 1, \ldots, n_{\hat{\alpha}} \quad \rightarrow \quad \hat{\mathbf{\theta}}^{\text{tls}} \quad \hat{\alpha}^{\text{tls}} \quad \hat{\sigma}^{\text{tls}}
\]

\( n_{st} \) points between \( \hat{\alpha}^{\text{tls}} \pm \hat{\sigma}^{\text{tls}} \)

\( \hat{\sigma}^{\text{tls}} = 12.37 \quad n_{st} = 219 \) in the example.
The **density** for the structure is the ratio between the number of points and the scale of the structure.

$$\rho = \frac{n_{st}}{\sigma_{tls}}$$

$$\rho = 17.7 \text{ in the example}$$

$n_{st}$ are removed from the input and the processing of the next structure begins ..... until less than $n_e$ of points.
Sorting the structures

The detected structures are sorted in **descending** order based on the densities.

Significant inliers structures have much smaller scales and much larger densities.

The **user** has only to specify how many inlier structures are returned in the estimation.

Here retains the first two structures.
Why these constants?

*ellipses: inliers 33% elemental subset x 5 = 4.17% 100 trials*

With $M$ increasing till 1000 the output improves. We took 2000. For higher $M$-s the results no longer improve statistically. Pre/post-processing, not in MISRE, could only further help.
Summary of MISRE

For each structure:

- **Scale** is the largest expansion in the region of interest.
- **Refinement** returns $n_{st}$ points between $\hat{\alpha}_{tls} \pm \hat{\sigma}_{tls}$.
- **Density** of the structure is $\rho = n_{st}/\hat{\sigma}_{tls}$.

**Sorting** by decreasing densities.

**Separating** the inlier structures from outliers requires the user just to decide where the $\hat{\sigma}_{tls}$ increased a lot.

MISRE is as good as the RANSAC-type estimators if similar noise corrupts all inliers, but is superior when the noise is very different for each inlier structure.
MISRE degrades *gradually* when number of outliers increases while the RANSAC-type estimators fail completely. Will be discussed at the end of the talk.

Processing times are based on i7-2617M had a 1.5GHz clock.
2D lines

Canny edge detection 8072 points. \( M = 1000 \)

\( n_e = 404 \quad \text{t}_p = 4.35 \text{ seconds} \)

First three structures are inliers. Outliers <10 times larger scale.

Later will see how multiple inlier types can be also detected.
2D ellipses

Canny edge detection 4343 points. $M = 5000$

$n_e = 218 \quad t_p = 18.90 \text{ seconds}$

First three structures are inliers. Outliers <20 times larger scale.
Fundamental matrices

Parts (moving) together in 3D give a structure in 2D. \( M = 5000 \)

608 input pairs \( n_e = 8 \times 5 = 40 \) \( t_p = 1.75 \) seconds

First two structures are inliers (508 pairs red/green).

Outliers <15 timers larger scale (blue).
2D homographies

Correspondences in 3D may not correspond to correspondences in 2D for the homographies.

$M = 2000$

1910 input pairs $n_e=97$ $t_p=3.78$ seconds

First five structures are inliers (1747 pairs red/green/blue/cyan/yellow).

Outliers <50 times larger scale (purple).
Experiments with 3D point clouds

* From a sequence of 2D images
* build 3D tracks covering small parts of the entire 3D scene
* which are fused together into a single 3D point cloud.

Structure from Motion algorithm (SfM)
MISRE used repeatedly during this 3D reconstruction.

Processing time starts after the 3D point cloud was estimated.
3D point cloud was also generated with the Autodesk professional program ReMake.

Input images in 2D give a 3D mesh model.

In ReMake, the user should select before the estimation what are the different surfaces in the data.
In 3D many points can be outliers for the current estimation.

3D planes
Structure from Motion algorithm. \( M = 1000 \)

70 - 2D images give 23077 points in the 3D point cloud.

\( n_e = 1154. \quad t_p = 7.04 \text{ seconds.} \)

First six structures are inliers (21758 points).
Outliers much larger scale and smaller density.
3D spheres

\[
x = [X \ Y \ Z \ X^2 + Y^2 + Z^2]^T
\]

Autodesk ReMake \[ M = 1000 \]

36 - 2D images give 10854 points in the 3D point cloud.

\[ n_e = 543, \ t_p = 7.24 \text{ seconds}. \]

First two structures are inliers (3504 points).
Outliers have much larger scale and radius.
3D circular cylinders

Nine points solution $x = [X \ Y \ Z \ X^2 \ XY \ XZ \ Y^2 \ YZ \ Z^2]^T$ is valid for a cylinder when the elemental subsets have to satisfy additional relations resulting in only five degrees of freedom.

Autodesk ReMake $M = 2000$

22 - 2D images give 6500 points in the 3D point cloud

$n_e=325. \ t_p=12.65$ seconds.

First two structures are inliers (2262 points). Outliers have much larger scale and no height.
3D point cloud gives two spheres or three planes. Post-processing of both outputs together, after some reallocation of points, solves both tasks.

Better pre-processing can increase the number of inliers. Better post-processing can recover more inlier struct.
Limitations of MISRE

Every robust estimator fails if too many outliers are present. If a single inlier structure fails RANSAC fails too.

MISRE estimates each structure independently, therefore at the beginning only the "weakest" inlier structure, the lowest inlier density, become outliers.
five lines, "weakest" $n_{in}=100$ gaussian noise=15, $n_{out}=350$
"weakest" 6/100 are outliers

$n_{out}=500$ "weakest" 34/100, next one 3/100 too are outliers
three ellipses, smallest $n_{in}=200$ gaussian noise=$9$, $n_{out}=350$

smallest $10/100$, middle $6/100$ too are outliers

$n_{out}=800$ smallest $53/100$, middle $19/100$ too are outliers

exmp: outliers (blue) before inliers (cyan) two inliers interact
circle $n_{in}=200$ $n_{out}=1500$. Scale estimates correct around $\sim23$. Mean shift depends on the radius of the circle.

$r = 50$

incorrect often

$r = 200$

correct in 100 tests
MISRE, the Multiple Input Structures with Robust Estimator estimates each structure independently.

MISRE has a simple set-up with the same two constants for every estimation.

MISRE also works for inlier structures with vastly different standard deviations.

Paper in IEEE PAMI "Easy Access" number 9091905

Programs at https://github.com/MISRE
Thank You