ECCV '14 Zürich, September 6-12th, 2014

European Conference on Computer Vision

Match selection and refinement for highly accurate two-view SfM

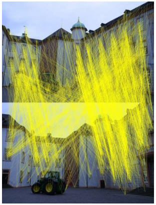
Zhe Liu, Pascal Monasse, Renaud Marlet

Université Paris-Est, LIGM (UMR 8049), ENPC, F-77455 Marne-la-Vallée, France

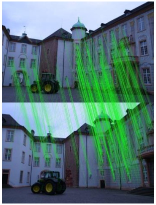
ECCV 2014, September 8-11

Standard two-view Structure from Motion

Feature detection & matching

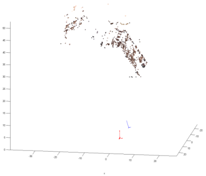


RANSAC for F estimation



Camera R & t estimation

3D reconstruction

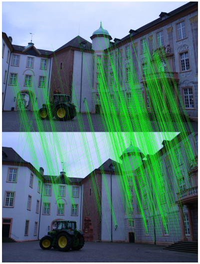


Point cloud

Match Selection

Quality vs quantity

Is using all inliers for F estimation the best thing to do?



- more matches with lower accuracy

  or
- less matches with higher accuracy

Goal: find large subset of most accurate matches

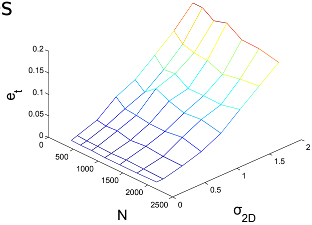
$\Rightarrow$ better SfM accuracy

Match Selection

Correlation of errors to quality and quantity

Experiments on real dataset:

- varying quality: σ_{2D}, match localization error
- varying quantity: N, number of matches



Measured errors:

- e_{3D}: 3D point location error
- e_R: camera rotation error
- e_t: camera translation error
- e_F: average epipolar error

Observations:

$$\log(e_{3D}), \log(e_R), \log(e_t) \approx \alpha \log(\sigma_{2D}) - \beta \log(N) \qquad (1)$$

with α and β depending on the match configuration

$$e_F \propto \sigma_{2D} \qquad (2)$$

Match Selection

Comparing errors

$$e_{3D}, e_R, e_t \propto \frac{e_F^{\alpha}}{N^{\beta}} \tag{3}$$

Knowing α/β is sufficient to compare errors:

$$\begin{aligned} e_{3D} &< e'_{3D} \\ e_R &< e'_R \\ e_t &< e'_t \end{aligned} \quad \Leftrightarrow \quad \frac{e_F^{\alpha}}{N^{\beta}} < \frac{e_F'^{\alpha}}{N'^{\beta}} \quad \Leftrightarrow \quad \frac{e_F^{\alpha/\beta}}{N} < \frac{e_F'^{\alpha/\beta}}{N'} \tag{4}$$

Thus, $M_{\mathrm{sub}} \subset M$ is better than M if:

$$\frac{e_F(M_{\mathrm{sub}})^{\alpha/\beta}}{|M_{\mathrm{sub}}|} < \frac{e_F(M)^{\alpha/\beta}}{|M|} \tag{5}$$

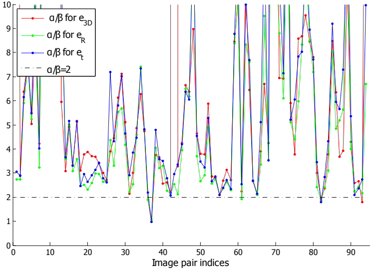
Match Selection

Comparing errors

Knowing a lower bound $\gamma \leq \alpha/\beta$ is enough to compare errors:

$$\frac{e_F(M_{\text{sub}})^{\gamma}}{|M_{\text{sub}}|} < \frac{e_F(M)^{\gamma}}{|M|} \quad \Rightarrow \quad \frac{e_F(M_{\text{sub}})^{\alpha/\beta}}{|M_{\text{sub}}|} < \frac{e_F(M)^{\alpha/\beta}}{|M|} \qquad (6)$$

Experiments:



- $\alpha/\beta \geq 2$ almost consistently
- $\gamma = 2$ thus safe for most scenes
- $\alpha/\beta = 2$ theoretical value for homography case

Match Selection

Exploring match subsets

Goal: find the optimal subset M^*_{sub} for estimating F

$$M^*_{\text{sub}} = \underset{M_{\text{sub}} \subset M}{\arg\min} \frac{e_F(M_{\text{sub}})^\gamma}{|M_{\text{sub}}|} \tag{7}$$

Problem: exploring all $M_{\text{sub}} \subset M$ is impractical

Our solution:

- score matches with some function $\phi : M \to \mathbb{R}$ (lower the better)
- sort matches according to ϕ: $i < j \Rightarrow \phi(m_i) < \phi(m_j)$
- consider N best matches $M_{\text{sub}} = \{m_1, \dots, m_N\}$ for all $N \leq |M|$
- in fact consider only a few values for N (discrete fractions of $|M|$)

Match Selection
Match ranking function

The choice of ϕ varies with the kind of feature.

For SIFT, match localization error correlates with:

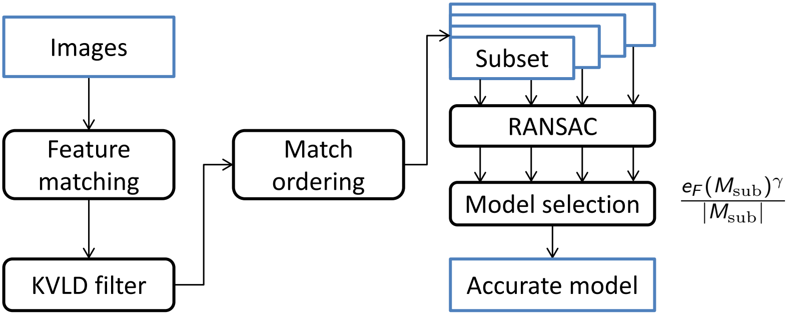
- the scale of detected features
- the descriptor difference

Our choice:

$$\phi(\mathbf{x}, \mathbf{x}') = \max(\mathit{scale}(\mathbf{x}), \mathit{scale}(\mathbf{x}')) \times d(\mathit{desc}(\mathbf{x}), \mathit{desc}(\mathbf{x}')) \quad (8)$$

Match Selection

Pipeline

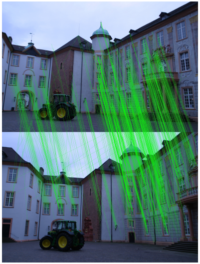


KVLD: robust photometric matching method that removes most outliers before M is sub-sampled (Liu & Marlet, BMVC 2012)

Match Selection

Conclusion for match selection

Is using all inliers for F estimation the best thing to do?



- more matches with lower accuracy

  or

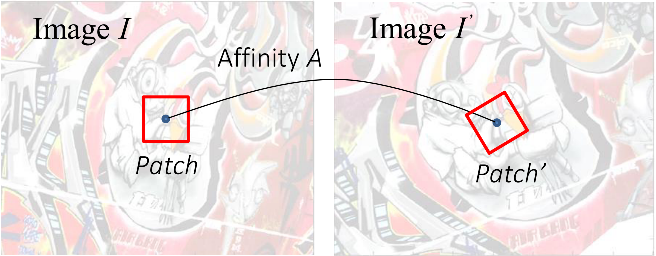
- less matches with higher accuracy

Goal: find large subset of most accurate matches

$\Rightarrow$ better SfM accuracy

Match Refinement

Least Square Matching (LSM)



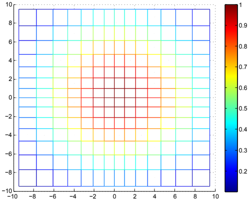
$$A^* = \arg\min_{A,f} \sum_{x \in Patch} \left| I(x) - f \circ I' \circ A(x) \right|^2 \qquad (9)$$

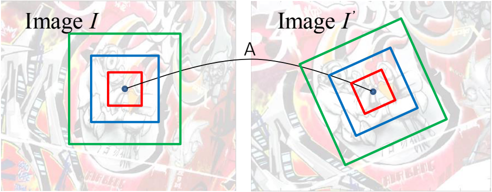
with $f(i) = ai + b$ linear radiometric adjustment

Match Refinement
LSM extension

Least Square Focused Matching (LSFM):

- irregular sampling grid focused on patch center
- image scale exploration for robustness to local minima

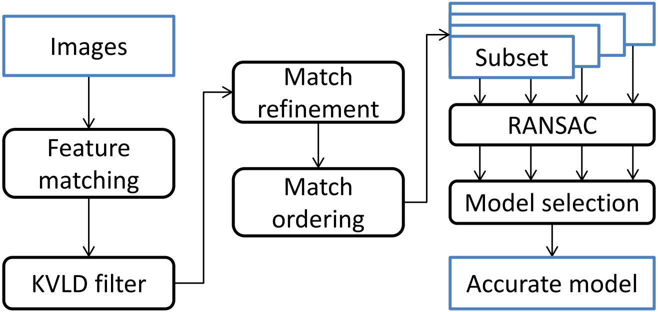




Match Selection with Match Refinement

Pipeline

Refinement before selection:



Match Selection with Match Refinement

Match ranking function

Feature point location adjusted $\Rightarrow$ no more correlation of errors with

- detection scale
- descriptor difference

Another scoring function ϕ required

New ϕ based on correlation of errors with

- dissimilarity of affine-transformed patches
- shearing of affinity

Experiments

Average rotation and translation errors

Strecha et al.'s datasets: 95 image pairs (CVPR 2008)

std: RANSAC-like only
MS: with match selection
MR: with match refinement
gain: std/(MR+MS)

e_R (deg $\times 10^{-2}$)	std	MS	MR	MR+MS	**gain**
RANSAC	16.4	9.52	10.3	8.87	**1.9**
MSAC	14.1	9.53	8.86	8.43	**1.7**
LO-RANSAC	16.4	9.54	10.3	8.97	**1.8**
MLESAC	15.8	7.81	9.50	7.76	**2.0**
ORSA	12.2	7.24	6.48	6.60	**1.9**
e_t (deg)	std	MS	MR	MR+MS	**gain**
RANSAC	1.85	1.09	1.23	1.04	**1.8**
MSAC	1.59	1.08	1.03	0.96	**1.6**
LO-RANSAC	1.83	1.10	1.21	1.05	**1.7**
MLESAC	2.16	0.95	1.09	0.87	**2.5**
ORSA	1.38	0.81	0.68	0.74	**1.9**

Experiments

Average rotation and translation errors

DTU robot datasets: 108 image pairs
(Aanæs et al., IJCV 2012)

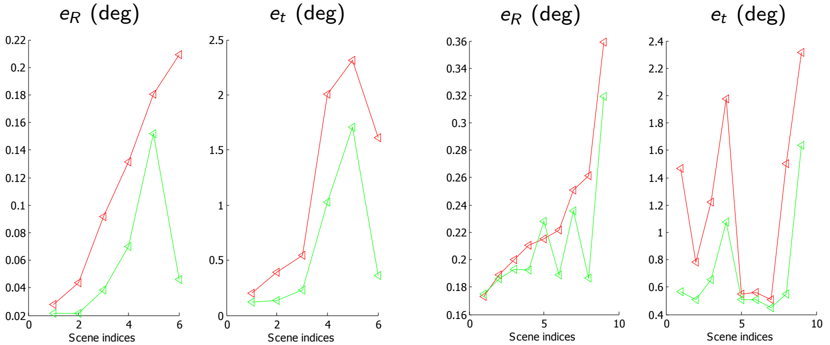
std: RANSAC-like only
MS: with match selection
MR: with match refinement
gain: std/(MR+MS)

e_R (deg $\times 10^{-2}$)	std	MS	MR	MR+MS	**gain**
RANSAC	26.5	22.3	21.5	21.3	**1.2**
MSAC	21.3	21.7	20.4	20.1	**1.1**
LO-RANSAC	26.8	22.2	21.5	21.3	**1.3**
MLESAC	21.8	22.6	20.8	20.2	**1.1**
ORSA	21.9	21.7	20.8	20.3	**1.1**
e_t (deg)	std	MS	MR	MR+MS	**gain**
RANSAC	3.83	2.12	1.81	1.02	**3.7**
MSAC	1.27	1.03	0.93	0.70	**1.8**
LO-RANSAC	3.89	2.14	1.76	1.02	**3.8**
MLESAC	2.02	1.34	1.23	0.77	**2.6**
ORSA	1.22	0.88	0.66	0.66	**1.8**

Experiments

Average rotation and translation errors

Variations with the kind of scene



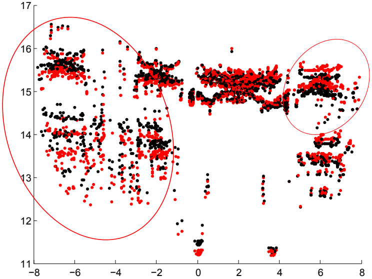
Strecha et al. (6 scenes) DTU robot (10 scenes)

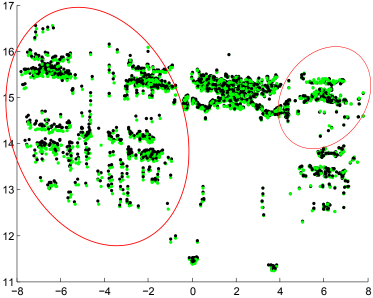
Estimation with **standard RANSAC** vs **MR+MS**

Experiments
3D point errors

Ground truth
Std RANSAC
MR+MS

Frontal view of point cloud



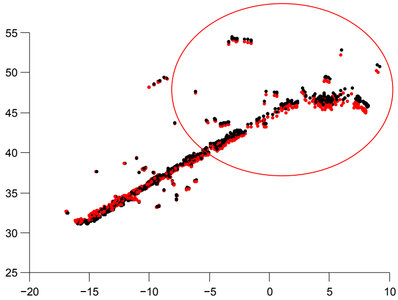


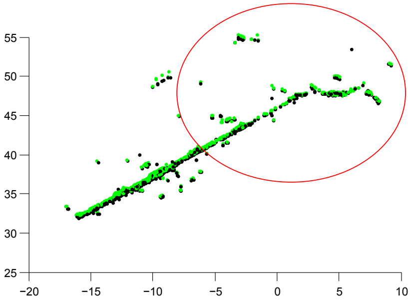
Experiments

3D point errors

Ground truth
Std RANSAC
MR+MS

Top view of point cloud





Conclusion

- Study of quality vs quantity of matches for 2-view SfM
  $\Rightarrow$ correlation of SfM errors with match number & location errors
- A new method for the selection of subsets of accurate matches
  $\Rightarrow$ improved SfM accuracy
- Combination with an improved LSM for match refinement
  $\Rightarrow$ even better SfM accuracy

Future work: extension to multi-view

- track selection/reduction (possible)
- track refinement (not trivial)

Source code available on Github!

Thank you!

Q & A

Additional

Exploring subsets

Problem: Exploring all $M_{\mathrm{sub}} \subset M$ for M^*_{sub} is impractical.
Our solution:
Assuming a ranking function:

$$\phi : M \to \mathbb{R} \text{ such that } \forall i < j \Rightarrow \phi(m_i) < \phi(m_j), \tag{10}$$

consider the fractions $M_{\mathrm{sub}}(N) = \{m_i \mid 1 \leq i \leq N\}$.
If ϕ is highly correlated to $e_{2D}(M, m)$, hence to $e_F(M, m)$, then

$$\begin{aligned} \min_{M_{\mathrm{sub}} \subset M} \frac{e_F(M_{\mathrm{sub}})^2}{|M_{\mathrm{sub}}|} &= \min_{N \leq |M|} \frac{1}{N} \min_{\substack{M_{\mathrm{sub}} \subset M \\ |M_{\mathrm{sub}}| = N}} e_F(M_{\mathrm{sub}})^2 \\ &\approx \min_{N \leq |M|} \frac{1}{N} e_F(M_{\mathrm{sub}}(N))^2 \end{aligned} \tag{11}$$

Additional

SIFT scoring function

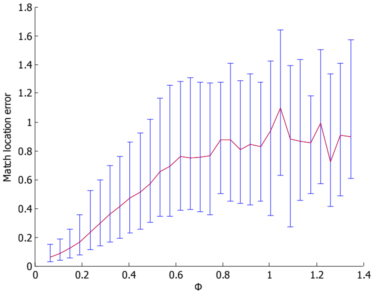


Figure: Correlation of σ_{2D} and ϕ.

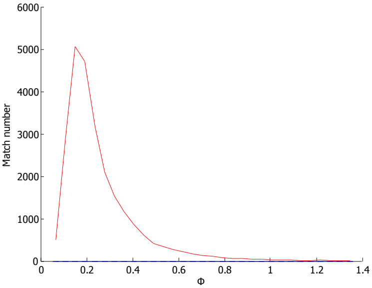


Figure: Histogram of ϕ