Robust Long-Term Aerial Video Mosaicking by Weighted Feature-based Global Motion Estimation

Holger Meuel · Stephan Ferenz · Florian Kluger · Jörn Ostermann

Goal and Problem Definition

- **Goal:** Highly accurate *long-term* mosaicking from aerial sequences
- Global motion estimation (GME) for frame registration using corner features, KLT feature tracker, RANSAC
- Accurate GME may be degraded by:
  - Nearly no features on unstructured areas (lawn)
  - Feature points at 3D structures not covered by the model (white ellipse)
  - Feature points at (slow) moving objects (red circle)

Problem: Drift induced by frame-to-frame-based estimation

Proposed Improvements

I. Weighted Feature-based Global Motion Estimation

Features often located on structured areas ⇒ enforce uniform distribution
Idea: Weight feature correspondence influence based on local feature density

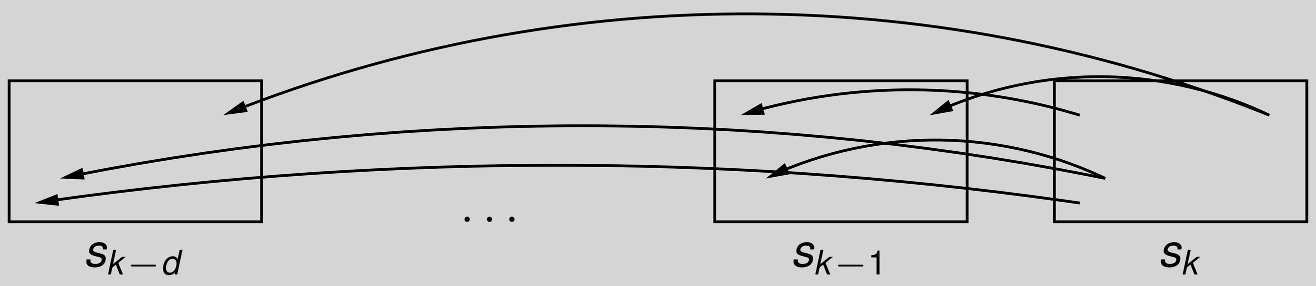
1. Extend least-square minimization in homography estimation for J inliers: $\min \sum_{j \in J} \left((\tilde{x}_{j,k-1} - x_{j,k-1})^2 + (\tilde{y}_{j,k-1} - y_{j,k-1})^2 \right) \cdot (W_{j,k})^2$
   with $(\tilde{x}_{j,k-1}, \tilde{y}_{j,k-1})$: estimated coordinates, $W_{j,k}$: weighting function depending on $x_{j,k}$ and $y_{j,k}$
2. Modeling $W_{j,k}$ using instance reweighting to approximate uniform distribution $p_e(x,y)$ of feature correspondences:
   $p_{\mathrm{feat},k}(x,y) = \frac{1}{J} \sum_{i=1}^{J} K(x - x_{i,k}, y - y_{i,k})$
3. Approximate kernel K by a Gaussian probability density function (pdf) to model the neighborhood of each feature:
   $p_g(x,y) = \frac{1}{2\pi\sigma_x\sigma_y} \exp\left[-\frac{1}{2} \left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2} \right) \right]$
4. Set σ_x and σ_y to mean value of pairwise distances of all feature correspondences and add scaling factor κ [1]
5. Calculate final weighting function $W_{j,k}$: $W_{j,k}(x_{j,k}, y_{j,k}) = \frac{p_e(x_{j,k}, y_{j,k})}{p_{\mathrm{feat},k}(x_{j,k}, y_{j,k})} = J \cdot \frac{2\pi\sigma_x\sigma_y}{\sum_{i=1}^{J} \exp\left[-\frac{1}{2} \left(\frac{(x_{j,k} - x_{i,k})^2}{\sigma_x^2} + \frac{(y_{j,k} - y_{i,k})^2}{\sigma_y^2} \right) \right]}$
   i. e. weight for each feature is the reciprocal of the real feature distribution

II. Increase of the Number of Features with Highest Possible Quality ("More features")

- **Idea:** Feature coverage over entire frame ⇒ always select the n-best features with n being a predefined number of features

III. Variable Tracking Distance

- Tracking over long temporal distances ⇒ local motion tends to be larger
- ⇒ RANSAC can better detect & remove features on moving objects
- ⇒ Reduced outliers ⇒ improved tracking accuracy



Experimental Results

Camera Captured Videos

- Aerial video captured from multicopter, 3000 frames, non-stabelized
- Hover on the spot, only small translational movement, but wobbling
- Accumulated GME error leads to inaccurate panorama image (left)
  - Severe stitching errors, broken lines, global deformation
- Improved GME leads to accurate panorama over 3000 frames (right)
  - Accurate stitching, consistent lines, no deformation
  - Accurate mosaicking of frames despite severe model violations

Left: Stitching without proposed techniques, including errors induced by inaccurate GME

Right: Final panorama using all proposed improvements for global motion estimation with uniform distribution and weight of $\kappa = 0.575$.

Results for Camera Captured Videos

Table: Results of different methods for the *Soccer* sequence, 3000 images ((∗): manual reference only for 100 images) and the *1500 m* sequence from TAVT.

Sequence	*Soccer* seq. DSSIM		*1500 m* seq. DSSIM	
Method	Mean	Max.	Mean	Max.
Manual Reference	0.036(∗)	0.060(∗)	—	—
Baseline (w/o proposed methods)	**0.120**	**0.146**	**0.067**	**0.156**
Weighting of correspondences	0.123	0.151	0.066	0.155
More features	0.094	0.129	0.065	0.133
Weighting & more features	0.094	0.128	0.064	0.133
Variable tracking	0.054	0.079	0.062	0.094
Weighting & variable tracking	0.045	0.071	0.063	0.099
Weighting & more feat. & var. track.	**0.035**	**0.051**	**0.061**	**0.088**

Conclusions

- Robust, highly accurate, long-term mosaicking over several thousand frames for aerial video with only little translational movement
- Combined approaches reduce stitching errors
- Highly improved subjective quality of panoramic images
- Structural dissimilarity (DSSIM) reduced from 0.12 to 0.035

References: [1] S.J. Reddi, A. Ramdas, B. Póczos, A. Singh, L. Wasserman: On the Decreasing Power of Kernel and Distance Based Nonparametric Hypothesis Tests in High Dimensions, Proc. of the AAAI Conf. on Art. Intel. pp. 3571-3577 (2015)

tnt
Institut für Informationsverarbeitung
Holger Meuel
meuel@tnt.uni-hannover.de
https://www.tnt.uni-hannover.de/staff/meuel

Leibniz Universität Hannover