Combined Guided Tracking and Matching with Adaptive Track Initialization

Anton Konouchine, Victor Gaganov, Vladimir Vezhnevets
Keldysh Institute of Applied Mathematics, Moscow State University
{ktosh,gagra,vvp}@graphics.cs.msu.ru

Abstract

Point feature tracking is the key step in solving such problems as camera calibration and 3d reconstruction. In this paper, we propose a new feature-tracking framework that is based on combination of guided tracking and matching approaches. The proposed framework raises the quality of tracking in terms of mean track lengths and fraction of successfully tracked features. We also propose an adaptive track initialization scheme based on spatial partitioning of detected features into bins that reduces the influence of dominant planes on outlier segmentation. Results are given for a number of real image sequences. The algorithm is demonstrated to outperform previous approaches.

Keywords: Feature Tracking, Guided Tracking, Guided Matching, Fundamental Matrix, Homography, Rematching, Dominant Planes

1. INTRODUCTION

One of the key problems in many computer-vision tasks such as 3d reconstruction and camera calibration is to establish a correspondence between points of different images of the same 3d scene. In general case it is impossible to compute such a correspondence for all pixels in the image. To lower the complexity of the problem a notion of *point feature* is introduced. Point x' is an image feature if its neighborhood is different from neighborhoods of all other neighboring points by selected measure:

$$\{\forall x : |x'-x| < r \rightarrow \rho(\Omega_x, \Omega_{x'}) > \varepsilon\},$$

where Ω_x is neighborhood of point x that is called *search window,* $\rho(\Omega_x, \Omega_{x'})$ is a measure of image distance. Consider $\{I_i\}, i = \overline{1,n}$ - is input image sequence, the sequence of point feature positions in image sequence $\{x'_i\}, i = \overline{1,n}$ is called *point feature track.*

Two general approaches exist for correspondence estimation. First is called feature matching and consists of two steps – independent detection of features in all frames and their matching in certain frame pairs (usually successive ones) [1]. The second step is called feature tracking. It relies on sequential tracking of feature positions, which have been detected on the first frame of the sequence [2].

In this paper we address several problems of feature tracking frameworks. First problem is correspondence estimation failures of tracking methods. Such failures significantly lower the length of some feature tracks and lead to generation of two different tracks for same image feature. The second problem is non-uniformity of feature detection in image, which arises from differences in texture density. Features from richly textured planar surfaces can outnumber all other features and create dominant subset, which lead to erroneous motion model selection and rejection a large number of correct tracks.

2. BACKGROUND

During the last 20 years a lot of different point feature detection algorithms had been developed. The most renown and widely used from them is Harris corner detector [3], which demonstrates low computational complexity and high repeatability, invariant to image rotations and noise. The feature matching is performed by comparing feature neighborhoods using selected image distance measure, e.g. cross-correlation or Sum of Squared Distances (SSD)[1]. One of the first feature tracking methods was Lucas-Kanade iterative algorithm [4]. Later it was modified to compensate affine deformations of feature search windows (e.g. during camera rotation and zooming) and changes of lighting conditions [2].

The most severe problem of feature track computation is erroneous correspondence estimation. Both feature-tracking and matching can match feature in one frame with wrong point in other frame. Additionally, some detected features arise not from 3d point on the surface of the scene object, but from some imaginary intersections of different objects. Such tracks are called outliers and should be rejected. For outlier track segmentation several tracking frameworks were proposed such as multiple-hypothesis tracking [5]. All of them are based on robust estimation of two- and three-view relations like fundamental matrix, homography and tri-focal tensor [6]. Matches that are marked as outlier during robust model estimation are removed, and their respective tracks are finished.

The most sophisticated tracking framework was proposed by Gibson in [7]. It uses feature tracking algorithm for correspondence estimation, and based on adaptive selection of key-frames, which partition the image sequence into segments. Consider the key-frame I_i, the next key-frame I_j is selected such as that is satisfy one of the following criteria:

a) The length of the sequence is reached or 10 frames is processed

b) 50% of tracks is lost during tracking between I_i and I_j

c) Robustly estimated homography from movement of features between I_i and I_j has large overall RMS

The last criterion is the most important. The algorithm of Gibson relies on assumption, that when robustly estimated homography fits the data poorly, then corresponding structure is far from the

degenerate, and fundamental matrix $F_{i,j}$ can be reliably calculated for matches of frames I_i and I_j. Inlier tracks are then used for estimating $\{F_{i,k}\}, k = \overline{i+1, j}$. The tracking of features from frame I_i is then repeated. For each frame I_k fundamental matrix $F_{i,k}$ is used to identify those features that moves a significant distance from corresponding epipolar line. The tracking of these features is stopped.

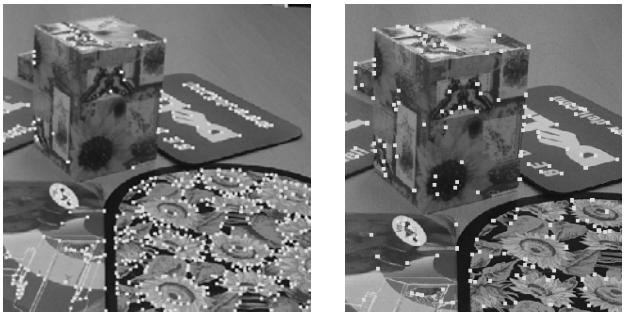
3. DRAWBACKS OF FEATURE TRACKING

Feature tracking methods were originally developed for tracking in video sequences that generally have a high frame rate, exceeding 10 frames per second. The camera movement between frames is usually small, so feature displacement between frames is also small. Feature tracking methods use feature position x_i in frame I_i as prediction $\tilde{x}_{i+1}$ of feature position x_{i+1} in next frame I_{i+1}, which is used as initialization for iterative search procedure. For video sequences, such initialization precision is sufficient. However, when image sequence is captured by photo camera, frame rate is much lower and camera movement is larger, which lead to large displacement of some features. If the displacement of feature is large between I_i and I_{i+1}, than $\tilde{x}_{i+1}$ can fall to smoothly colored regions without rich texture like room wall. In this case, feature tracking methods fail to correctly estimate the feature position x_{i+1} in I_{i+1}. The repetition of the tracking as in Gibson's framework will result in the same tracking failure.

The second problem of feature-tracking frameworks arises from track initialization procedure. Generally a threshold M is set on the number of detected features. The features are sorted according to their respective quality, so only M best features are selected and used for track initialization. If richly textured objects presents in captured scene, a large fraction of features are detected in the image of this objects. If the object is planar, the matches for this group of features satisfy planar homography. When the fraction of those features is large enough they form a dominant subset and overrule correct model during robust estimation. This subset is called a dominant plane. Figure 1(a) shows an example of richly textured flat surface. Note that majority of features are detected on mouse pads and tea-cloth, thus forming a dominant plane.

4. PROPOSED METHOD

We propose a new feature-tracking framework. It is based on partitioning the image sequence into segments using key-frames like one proposed by Gibson, but uses guided tracking and matching to establish correct correspondences for those features, which have been tracked incorrectly during key-frame selection. We also propose a new uniform track initialization scheme that is based on spatial feature partitioning.



(a) (b)

Figure 1 Adaptive track initialization. (a) Standard track initialization scheme. (b) Adaptive track initialization scheme. Note that number of features on the surface of box is significantly increased in (b), compared to (a), while number of features on flat surface is diminished.

4.1 Guided tracking

To solve the problem of feature tracking failures, the feature position in I_{i+1} should be predicted with greater accuracy. We propose to use point transfer by planar homography $H_{i,i+1}$ between frames I_i and I_{i+1} for prediction. Consider I_k and I_j - are two selected keyframes, where $k \le i < i+1 \le j$, for which the fundamental matrix $F_{k,j}$ is robustly estimated. Let $\{T\}$ be a set of feature tracks that have been marked as inliers during robust estimation of $F_{k,j}$, then homography $H_{i,i+1}$ is calculated by least-squares method from $\{T\}$.

Let $\{y_i\}$ be a set of features from I_i, for which feature tracking algorithm failed to establish correct correspondences. Set of predicted feature positions $\{y_i'\}$ can then be calculated by homography mapping:

$$y_i' = H_{i,i+1} y_i$$

Feature tracking is then repeated to establish correspondences on image I_{i+1} for features $\{y_i\}$, using $\{y_i'\}$ as initialization. As shown in Figure 3, this prediction can correct some of feature tracking failures.

4.2 Guided matching

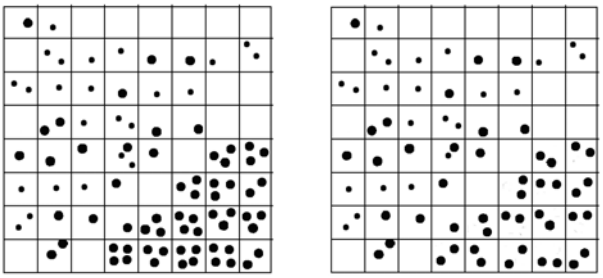
Feature tracking algorithm cannot directly benefit from known fundamental matrix. However, the epipolar constraint, written in form of fundamental matrix, can greatly limit the number of matches that are tested by feature matching methods [1]. The search of feature correspondences, that satisfy the known epipolar constrain is called guided matching.

We propose to use guided matching to establish correspondences for those features, for which even the guided tracking fails. $F_{k,i+1}$ is fundamental matrix for images

(I_k, I_{i+1}), where I_k is the last key-frame. Let x_k point in key-frame I_k that corresponds to x_i in frame I_i. Feature matching tests only feature matches (x_i, x_{i+1}), for which (x_k, x_{i+1}) satisfies the epipolar constraint induced by $F_{k,i+1}$.

4.3 Adaptive track initialization

To lower the probability of dominant subset appearance, the features should be detected uniformly in the image, so that same number of features is detected in each part of the image. In this case some features with lower quality will be selected in other part of the image, instead of high quality features from densely textured parts. Based on this idea we propose to partition the image in rectangular regions, which are called bins. Let N be the number of bins, then we select M/N best features for track initialization from each bin. The idea is illustrated in Figure 2.



(a) (b)

Figure 2 Partitioning of features into bins. The large points are features with high quality. (a) Partitioning into bins (b) Result of adaptive selection. In standard algorithm, features with best quality are selected, so only large points will remain. In this case, most of the tracks will be created from feature only in right-bottom corner of the image. In (b) features with lower quality are selected because they lie in different bin, and some of high-quality feature are neglected.

4.4 Algorithm summary

The proposed feature tracking framework is based on adaptive key-frame selection. However, experiments on real image sequences have shown, that robust estimation of homography is unstable when camera undergoes a considerable displacement between frames, which lead to different key-frame selection for each application to the image sequence. We estimate both fundamental matrix $F_{i,j}$ and homography $H_{i,j}$ and compare them using GRIC information criterion [8]. The new key-frame is found when either one of first two of Gibson's framework criteria is true, or when $F_{i,j}$ fits matches better then $H_{i,j}$ by GRIC. The proposed algorithm can be outlined as following:

1. Detection of point features in all frames $\{I_i\}, i = \overline{1, n_{total}}$
2. Partitioning of features intro bins and adaptive feature selection for track initialization for all frames
3. $i = 1$
4. Search for new key-frame I_j
5. If fundamental matrix $F_{i,j}$ fits better
    a. Second path using guided tracking and matching
2. If homography $H_{i,j}$ fits better
    a. Second path using guided tracking and matching
6. $i = j$
7. Repeat 3-6, until end of the sequence is reached

During second path guided tracking and matching is applied to features $\{x_i\}$ from I_i, for which tracking failed or that were rejected as outliers. The algorithm of second path for case of $F_{i,j}$ can be summarized as following:

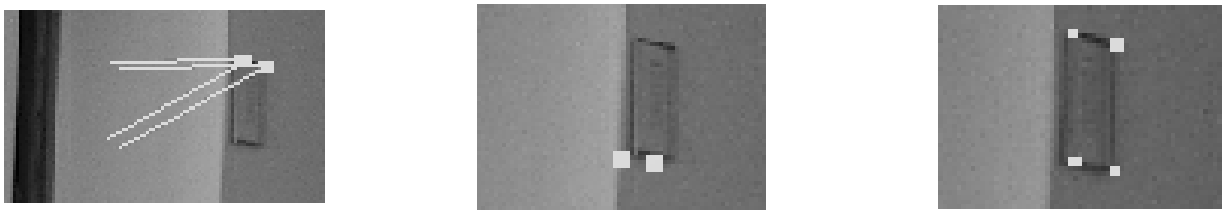
1. Calculate $\{F_{i,k}\}, \{H_{k-1,k}\}, k = \overline{i+1, j}$
2. For $k = \overline{i+1, j}$
    a. Apply guide tracking for (I_{k-1}, I_k) using $H_{k-1,k}$
    b. Apply guided matching for (I_{k-1}, I_k) using $F_{i,k}$
    c. Reject features using $F_{i,k}$ for which both guided tracking and matching fails.

If $H_{i,j}$ fits data better the same algorithm is applied, but homography $H_{i,k}$ is used for guided matching and outlier segmentation.

5. EXPERIMENTS ON REAL DATA

Several real image sequences were captured with Canon IXUS 500 camera. The scenes were constructed from a set of man-made objects arranged on a top of the table. The used objects have detailed textures that provide a large number of features. For adaptive track initialization testing planar objects with rich textures was used. Such objects produce thousands of distinct features that form a dominant feature subset and lead to wrong model type (homography) selection and erroneous inlier tracks rejection, as seen in Figure 1(a).

For correct comparison of proposed algorithm with one by Gibson in terms of mean track length and number of inlier tracks, adaptive track initialization has not been used. The results are shown in Table 1. As can be clearly seen, the number of inlier tracks is slightly lower, but the mean track length is larger for the proposed algorithm. When feature tracking fails in Algorithm of Gibson, a new feature track is initialized thus increasing the number of inlier tracks. In the same case, our method compensates the tracking failures by guided matching and tracking. This lowers the number of inlier tracks but raises the mean track length.



(a) (b) (c)

Figure 3 Guided feature tracking (a) Correspondences estimated by KLT tracker. (b) Feature position predicted by planar homography (c) Results of feature tracking by guided tracking. Note that KLT tracker without correct prediction matched both upper and lower corners of the frame to the upper ones.

As was shown in Figure 3 because of guided tracking our algorithm can correctly match features that move very far between the successive frames. The increased tracked length and features with large displacements between frames significantly increase the quality of scene structure and camera motion estimation.

The adaptive track initialization scheme was tested on image sequence with large densely textured planar object. Without adaptive initialization the 90% of the features were detected on the surface of this object if total 1000 features were detected. When the threshold on number of features was raised to 1500, still more then 83% of the features were detected on the surface of the planar object. In both cases planar homography was confidently selected by GRIC as correct motion model and most tracks out of the planar object were rejected as outliers. When adaptive track initialization was applied, only 60% of the 1000 detected features lied on planar object, as shown on Figure 1(b). In this case, fundamental matrix was selected as best model. The feature tracks from other objects were not removed.

Sequence	Mean track length	Number of inlier tracks
"Cup" sequence, (Gibson framework)	7.3	1746
"Cup sequence" (Proposed framework)	7.9	1697
"Box sequence" (Gibson framework)	8.6	2754
"Box sequence" (Proposed framework)	9.1	2703

Table 1 Comparison of proposed feature tracking framework with one by Gibson

6. CONCLUSION

In this paper a new feature-tracking framework has been proposed. It has been demonstrated that developed algorithm can efficiently compensate the feature tracking errors using combination of guided matching and tracking approaches. It has been shown to provide equal or superior mean track length compared to existing feature tracking frameworks. Also it has been demonstrated to efficiently enforce the uniformity of feature distribution through the image, which significantly reduces the problem of dominant planes.

Our method differs from existing methods in several ways. First, after new key-frame is selected, it exploits both guided tracking and rematching for features, which has been lost during common tracking or marked as outliers during fundamental matrix fitting. For guided tracking, a planar homography is estimated by least-squares method using all inlier features. Point transformed positions are used as initialization point for tracking algorithm. Second, it partitions all detected features into number of bins, and selects for tracking the same number of features from every bin. Third, during key-frame searching it estimates both homography and fundamental matrix, and compare them using GRIC criterion.

7. REFERENCES

[1]Z.Zhang, R.Deriche, O.Faugeras, Q.T.Luong. *"A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry"* AI Journal, vol.78:87-119, 1994

[2] J.Shi, C.Tomasi *"Good Features to Track"* CVPR Proc, pp. 593-600, 1994

[3]C.Harris, M.Stephens. *"A Combined Corner and Edge Detector"*, pp.147-151, 1988

[4]B.D.Lucas, T.Kanade *"An Iterative Image Registration Technique with an Application to Stereo Vision"* International Joint Conference on Artificial Intelligence, pp 674-679, 1981

[5] P.H.S.Torr, A.W.Fitzgibbon, A.Zisserman: *"Maintaining Multiple Motion Model Hypotheses Through Many Views to Recover Matching and Structure"*. ICCV Proc, pp. 485-491, 1998

[6] P.Torr, A.Zisserman, *"Robust Computation and Parametrization of Multiple View Relations"*, In Proc. ICCV'98, pp.727-732, 1998.

[7] S.Gibson, J.Cook, T.Howard, R.Hubbold *"Accurate Camera Calibration for Off-line, Video-Based Augmented Reality"*, In Proc. ISMAR'02, 2002

[8] P.Torr. *"An assesment of information criteria for motion model selection"*. In Proc. *CVPR*, pages 47-53, 1997

About authors

Anton Konouchine has received his specialist degree in computer science in 1997 in MSU. He is now a PhD student in Keldysh Institute of Applied Mathematics. His email address is ktosh@graphics.cs.msu.ru

Victor Gaganov is 4-th grade student in Computer Science department, MSU. His email address is gagra@graphics.cs.msu.ru

Vladimir Vezhnevets, PhD, has received his specialist degree in Moscow State University. He received his PhD in 2002 in Moscow State University. His email address is vvp@graphics.cs.msu.ru