



LAWRENCE  
LIVERMORE  
NATIONAL  
LABORATORY

# Non-Parametric Sequential Frame Decimation for Scene Reconstruction in Low-Memory Streaming Environments

D. Knoblauch, M. Hess-Flores, M. Duchaineau, K.  
I. Joy, F. Kuester

July 26, 2011

7th International Symposium on Visual Computing  
Las Vegas, NV, United States  
September 26, 2011 through September 28, 2011

## **Disclaimer**

---

This document was prepared as an account of work sponsored by an agency of the United States government. Neither the United States government nor Lawrence Livermore National Security, LLC, nor any of their employees makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States government or Lawrence Livermore National Security, LLC. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States government or Lawrence Livermore National Security, LLC, and shall not be used for advertising or product endorsement purposes.

# Non-Parametric Sequential Frame Decimation for Scene Reconstruction in Low-Memory Streaming Environments

Daniel Knoblach<sup>1</sup>, Mauricio Hess-Flores<sup>2</sup>, Mark A. Duchaineau<sup>3</sup>, Kenneth I. Joy<sup>2</sup>,  
and Falko Kuester<sup>1</sup>

<sup>1</sup> University of California, San Diego, USA, {dknoblau, fkuester}@ucsd.edu

<sup>2</sup> University of California, Davis, USA, {mhessf, kjoy}@ucdavis.edu

<sup>3</sup> Lawrence Livermore National Laboratory, Livermore, USA, duchaine@llnl.gov

**Abstract.** This paper introduces a non-parametric sequential frame decimation algorithm for image sequences in low-memory streaming environments. Frame decimation reduces the number of input frames to increase pose and structure robustness in Structure and Motion (SaM) applications. The main contribution of this paper is the introduction of a sequential low-memory work-flow for frame decimation in embedded systems where memory and memory traffic come at a premium. This approach acts as an online preprocessing filter by removing frames that are ill-posed for reconstruction before streaming. The introduced sequential approach reduces the number of needed frames in memory to three in contrast to global frame decimation approaches that use at least ten frames in memory and is therefore suitable for low-memory streaming environments. This is moreover important in emerging systems with large format cameras which acquire data over several hours and therefore render global approaches impossible.

In this paper a new decimation metric is designed which facilitates sequential keyframe extraction fit for reconstruction purposes, based on factors such as a correspondence-to-feature ratio and residual error relationships between epipolar geometry and homography estimation. The specific design of the error metric allows a local sequential decimation metric evaluation and can therefore be used on the fly. The approach has been tested with various types of input sequences and results in reliable low-memory frame decimation robust to different frame sampling frequencies and independent of any thresholds, scene assumptions or global frame analysis.

## 1 Introduction

There has been a significant amount of research in the area of Structure and Motion (SaM) in recent years. The approaches have matured enough to allow for reliable reconstructions from image or video sequences. Lately more work has been introduced to automate these approaches. One important step in SaM is the decimation of input frames, in order to reduce the computation load but also discard frames that could possibly lead to pose and structure degeneracies. Most approaches introduced in previous work apply a global frame decimation for parts of the captured image sequences. This not only results in high memory consumption, as many frames have to be buffered, but is

also not suitable for streaming environments, as high delays are introduced. These kinds of environments include reconstructions based on image streams acquired through embedded systems, where memory and network bandwidths are restricted and therefore require a sequential frame decimation before streaming. This paper introduces a new low-memory, non-parametric, sequential work-flow to frame decimation without usage of thresholds or scene-dependent knowledge for streaming environments. To enable this sequential work-flow a new frame decimation metric optimized for non-parametric, sequential frame decimation is introduced. Compared to decimation metrics introduced in previous work the new metric is designed to have one global maxima representing a good keyframe pair at each evaluation step. The goal of this approach is to avoid input frames that result in error-prone reconstructions and reduce the total amount of streamed data by filtering non-suitable frames as soon as possible.

Reconstruction errors can be introduced by degenerate camera poses, numerical errors in triangulation with small baselines or bad correspondences caused by large baselines which result in more occlusions. Degenerate camera poses can result from little or no translation, in which case epipolar geometry estimation is respectively ill-posed or undefined. Numerical errors in triangulation on the other hand are introduced by small baselines in relation to the depth of the viewed scene. This is because near-parallel rays yield triangulated points that have a potentially large uncertainty in their computed positions. Lastly, large baselines introduce errors in correspondences as more occlusions appear.

The main contribution of the introduced approach is to reduce errors in multi-view reconstructions from the mentioned error sources with a low-memory sequential frame decimation work-flow and newly introduced error metric, where no thresholds are used nor assumptions about the scene are made. The error metric is based on errors evaluated from different camera motion models and an analysis of the number of obtained correspondences in relation to the number of possible features in the observed scene. The sequential frame decimation is especially suited for pre-streaming filtering of frames in low-memory, low bandwidth environments.

## 2 Previous Work

Structure and Motion from image sequences or videos has been a major focus in computer vision. Pollefeys et al. [1] introduced a multi-view reconstruction approach based on uncalibrated hand-held cameras. Nistér [2] reconstructed scenes with a hierarchy of trifocal tensors from an uncalibrated frame sequence. Every algorithm based on uncalibrated images extracts as a first step correspondences between frame pairs. These correspondences are then used to extract the epipolar geometry or in other words the fundamental matrix and the corresponding camera poses. The most often-used feature matching algorithms are the scale-invariant feature transform (SIFT) by Lowe [3] and the speeded-up robust features (SURF) by Bay et al. [4].

With the automation of SaM from image sequences and videos the challenge to find good image pairs for pose estimation and reconstruction becomes apparent. Nistér [5] introduced a frame decimation algorithm based on global motion estimation between frames and a sharpness measure to remove redundant frames. Ahmed et al. [6] recently



introduced a global frame decimation algorithm based on number of correspondences, geometric robust information criterion (GRIC) [7] and a point-to-epipolar line cost between frame pairs. Both approaches analyze the given frames in a global manner and are therefore well-suited for global decimation once all frames are available. These approaches also rely on empirically chosen thresholds for the frame decimation decision. Royer et al. [8] introduced a simple sequential frame decimation algorithm for robotic applications. Their frame decimation decision is based on the number of available correspondences between keyframes, and tries to decimate as many frames in-between keyframes without going below an empirically chosen number of correspondences. Torr et al. [9] use their previously introduced GRIC approach to improve the correspondence track extraction over several frames by analysing if the epipolar geometry or a homography is a better motion model for the given frames. The active search algorithm proposed by Davison [4] performs a global analysis of frames to decide which one adds the most new information to a multi-view reconstruction. Beder and Steffen [10] introduced an error metric to analyse the goodness of pairs of camera poses based on the uncertainty of reconstructed 3D points. This error metric is only focused on good camera poses but does not estimate the possible goodness of the correspondences.

### 3 Frame Decimation

This paper introduces a sequential frame decimation approach that reduces the amount of input images for the SaM algorithm by filtering frames that could result in degenerate camera poses and numerically unstable triangulations. At the same time it ensures that large baselines are avoided, as this results in less accurate correspondences based on more occlusions. The sequential frame decimation is well suited for low-memory, streaming environments.

Figure 1 shows the basic work-flow of this approach. The primary goal is to find reliable consecutive keyframes to allow a more robust multi-view reconstruction, which has the additional benefit of reducing the amount of data transferred to the reconstruction infrastructure. The first step of the algorithm is to extract correspondences between the last keyframe and the present candidate frame. A frame decimation metric is extracted, which is used for a sequential frame decimation decision. Never more than three frames have to be kept in memory thanks to this sequential approach. It will be shown in the remainder of this paper that this approach results in a reduction of error between consecutive pairwise reconstructions and in the final multi-view reconstruction, and finds a locally optimal keyframe sequence while keeping memory usage low.

#### 3.1 Sequential Frame Decimation

Most previous frame decimation approaches aimed to find the globally optimal keyframe sequence and therefore required large subsets of the input frame sequence. This means that the frames have to be obtained in advance or the frames have to be buffered for a global analysis. It is obvious that these kinds of approaches are not well suited for streaming or low-memory environments. The sequential frame decimation approach in

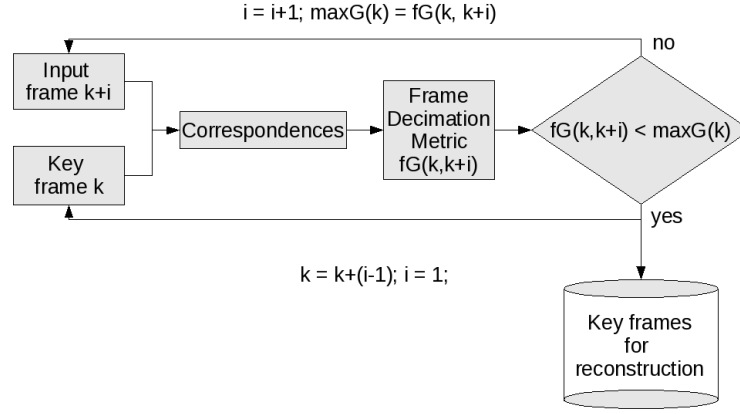


Fig. 1: Work flow sketch for the introduced non-parametric sequential frame decimation algorithm.

this paper overcomes these disadvantages by decimating frames on the fly. As a result only data from three frames, the last keyframe, the last frame and the present frame have to be managed in memory compared to at least ten frames in other approaches [6]. The introduced sequential work-flow can be used as long as the error metric is designed to have one global maximum, which corresponds to the next keyframe. This paper introduces a new decimation metric which incorporates a GRIC-based epipolar geometry versus homography residual comparison and the ratio of good correspondences to possible features at each evaluated frame pair. This is perfectly suitable for the introduced sequential frame decimation approach.

The sequential frame decimation is therefore performed by comparing the frame decimation metric  $fG$  of keyframe  $k$  with that of frame  $k + i$ . The counter  $i$  is increased as long as  $fG(k, k + (i - 1)) \leq fG(k, k + i)$ . As soon as the first frame pair with a positive value for  $fG$  and a significant decrease in error metric is found, the previous frame pair with a positive  $fG$  is chosen as a keyframe pair. To start the sequential frame decimation the first suitable frame of the input sequence is used as the first keyframe. The first possible key frame pair can be initialized based on Beder and Steffen [10] to have a valid starting point. Every frame pair with  $fG(k, k + i)$  which results in a local maximum is used for subsequent SaM calculations and frame  $k + i$  is used as the next start frame for further decimation. This work-flow is illustrated in figure 1. This sequential frame decimation approach gets rid of degenerate frames and reduces the number of frames with small baselines in which simple triangulation algorithms such as linear triangulation could fail, while keeping the baseline small enough to retain good correspondences.

### 3.2 Camera Pose and Structure Degeneracy Detection

There are two situations where relative camera pose estimation between two views results in degeneracies in which the epipolar geometry is not defined and therefore the fundamental matrix extraction fails. These cases are motion degeneracy and structure degeneracy. The motion degeneracy appears for example if the camera movement consists only of a rotation and no translation, or if scene points lie on certain quadric surfaces. The structure degeneracy appears if all the correspondences used to calculate the epipolar geometry are coplanar in the 3D scene. In these cases the fundamental matrix cannot be reliably computed, but the relative scene change between cameras can be described by a homography. Therefore, the comparison of residual error between these two scene representations gives an insight into the frame quality for pose and structure estimation. This comparison is performed based on the geometric robust information criterion (GRIC) [7]. GRIC is not only based on the goodness of fit but also on the parsimony of the two models. Equation 1 shows the relative comparison of fundamental matrix  $F$  and homography  $H$ .

$$relGRIC(F, H) = \frac{GRIC(H) - GRIC(F)}{GRIC(H)} \quad (1)$$

$GRIC(X)$  is defined by equations 2 and 3.

$$GRIC(X) = \sum_i \rho(e_i^2)_i + \lambda_1 dn + \lambda_2 k \quad (2)$$

$$\rho(e_i^2)_i = \min\left(\frac{e_i^2}{\sigma^2}, \lambda_3(r - d)\right) \quad (3)$$

The goodness of fit is represented by the sum of squared residuals  $e_i$  of  $F$  and  $H$  in relation to the input correspondences. The parsimony is based on  $d$ , the number of dimensions modeled ( $d = 3$  for  $F$  and  $d = 2$  for  $H$ ),  $k$ , number of degrees of freedom in the model ( $k = 7$  for  $F$  and  $k = 8$  for  $H$ ),  $r$ , dimension of the input data, which corresponds to  $r = 4$  in the case of 2D correspondences,  $\sigma^2$  is the variance of the residual errors, and similarly to [6] we set  $\lambda_1 = \log(r)$ ,  $\lambda_2 = \log(rn)$  and  $\lambda_3$  corresponds to a limit for the residual error, which was set to  $n$ , the number of correspondences between frames.

Based on the definition of  $relGRIC$  and the fact that  $F$  has high errors when there is a small baseline or there is no translation at all and the fact that  $H$  has low errors in these cases but high errors with wider baselines it can be said that a frame is a good input for SaM the higher the  $relGRIC$  is. It can also be said that if  $relGRIC < 0$  a bad input frame pair is found. In the sequential frame decimation work-flow all frame pairs with a  $relGRIC$  smaller than zero are directly decimated and not used for the sequential evaluation.

### 3.3 Correspondence Goodness

The value of  $relGRIC$  makes sure that frames with high degeneracy probability are excluded from further SaM steps. A closer look at  $relGRIC$  values and its definition

shows that its value will remain high or even increase as soon as  $H$  is no longer a good descriptor. In other words,  $relGRIC$  will have good values along longer baselines between cameras, as long as there are enough good correspondences to calculate the fundamental matrix. However, in the case of SaM reconstruction it is essential to find as many good correspondences as possible, and this usually occurs with smaller baselines since this results in fewer occlusions and matching errors. For this reason a weighting term  $cW$  for  $relGRIC$  is introduced, which represents the probability of finding good correspondences. Correspondences are extracted based on SURF [4] features. A good first guess for the size of the baseline and therefore the possible number of occlusions is taken by looking at the ratio between found features  $N_F$  in the source keyframe and the resulting correspondences  $N_C$  with the present target frame. This initial approximation gives adequate results but does not take the possibility of wrong correspondences into account. To make this weight more reliable, only inlier correspondences  $N_I$  from the RANSAC-based fundamental matrix calculation are used to calculate  $cW$ , as shown in equation 4.

$$cW = \frac{N_I}{N_F} \quad (4)$$

It can be seen that  $cW$  tends to decrease if the baseline grows. To make sure that the correspondences are a good representation of the given scene and cover as much of the scene as possible, a correspondence area  $cA$  versus image size  $iA$  ratio is introduced. The correspondence area  $cA$  is approximated by the axis-aligned bounding box of all the given inlier correspondences. To get a good reconstruction of the scene, this ratio  $aR$ , shown in equation 5, should stay as large as possible.

$$aR = \frac{cA}{iA} \quad (5)$$

### 3.4 Frame Decimation Metric

The last sections introduced ways to detect camera pose and structural degeneracies, and also evaluate camera baselines based on correspondence goodness. This chapter is going to introduce a new frame decimation metric  $fG$  that combines the latter results into a frame goodness for SaM estimation. The  $relGRIC$  metric is a relative measure of how good epipolar geometry describes the scene compared to a homography. In the case of  $relGRIC < 0$  this means that the frame pair is not well suited for reconstruction as there is either a camera pose degeneracy or the baseline is too small. On the other hand, if  $relGRIC > 0$  this means that this is a candidate frame pair for pose estimation. In this instance the main introduction of error in the reconstruction comes from the correspondences due in general to occlusions, which tend to increase with larger baselines. The introduced decimation metric takes both terms into consideration and weights the camera goodness by the correspondence goodness. This leads to the term for  $fG$  seen in equation 6.

$$fG = (cW * aR) * relGRIC(F, H) \quad (6)$$

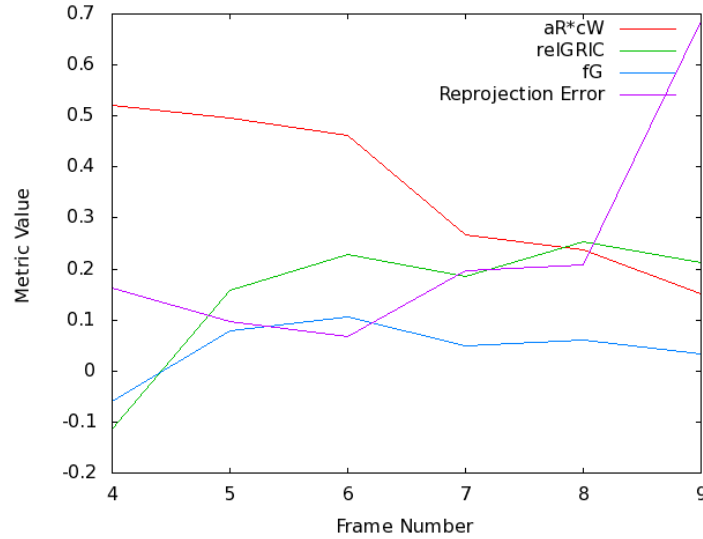


Fig. 2: Comparison of  $fG$  and its components ( $aR*cW$ ,  $relGRIC$ ) to reprojection error of the pairwise reconstruction from frame  $k = 0$  of the ‘kique’ sequence.

It can be seen that  $fG$  will have a high value if the  $relGRIC$  has a high value and the correspondence weight is high. This also means that  $fG$  will at most have the value of  $relGRIC$ , but also will decrease when the baseline grows. This behaviour represents the search for the sweet spot in baseline size based on the two main reconstruction error sources: camera pose and correspondences.

## 4 Results

The presented frame decimation algorithm has been tested with a large number of publicly available image sequences, consisting of different scene types and camera motions. Some examples of the tested sequences can be seen in figure 6. The ‘kique’ data set shown in figure 6a consists of aerial imagery taken while flying in a circle around a city neighbourhood. The ‘medusa’ data set [1] in figure 6b represents an image sequence taken by a hand-held camera moving in a half-circle around an archeological artifact, where the movement is very jittery at times. The data set ‘Leuven castle’ in figure 6c covers the outside of a building with a hand-held camera. Figure 6d shows data set ‘castle-P30’ [11] which covers a building from the inner courtyard. The ‘house’ data set [12] in figure 6e represents a carefully spaced image sequence taken in a circle around a toy house. It can be seen that these different target scenes and camera movements are all handled well by the introduced frame decimation approach.

An evaluation of the goodness metric is performed. In figure 2 the frame decimation values for the ‘kique’ data set are plotted in conjunction with the resulting reprojection

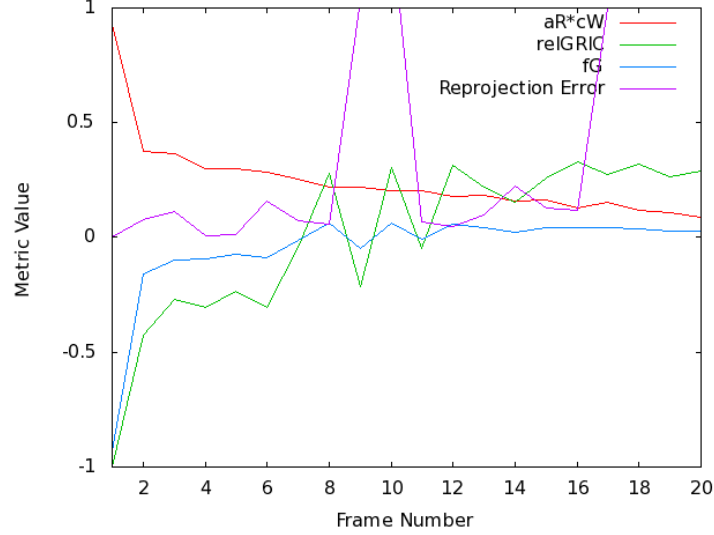


Fig. 3: Comparison of  $fG$  and its components ( $aR * cW$ ,  $relGRIC$ ) to reprojection error for each pairwise reconstruction with respect to keyframe  $k = 1$ , for the ‘medusa’ sequence.

error of the given frame pair reconstruction between the first suitable frame of the sequence and the following frames. It can be seen that the  $fG$  value for the keyframe pair search starting at frame  $k = 0$  has its best value at frame  $k + 3$ . This is in agreement with the lowest reprojection error representing the best reconstruction in this sequence. The plot shows also that frame  $k + 1$  is a bad frame for pose estimation as the  $fG$  has a negative value, which is supported by the reprojection error. A bigger baseline as suggested by the  $relGRIC$  (in frame  $k + 5$ ) is avoided thanks to the correspondence-based weighting. The frame decimation is run on the ‘kique’ data set and results in a sequence of keyframe pairs suitable for sequential multi-view reconstruction. By extracting 38 percent of the frames, degenerate camera poses are avoided while making sure that the baselines of the cameras stay small enough to reduce possible problems in correspondence calculation. This is all done on the fly with a sequential frame decimation work-flow designed to suit low-memory streaming environments. Figure 6a shows a selection of the first five extracted keyframes.

Another example of the tests that have been run is based on the ‘medusa’ data set [1]. For this test the input video has been decomposed into images with 10 frames per second. Figure 3 shows the resulting values in the frame decimation from the first frame  $k = 1$  in the sequence. Based on the introduced algorithm the next keyframe is frame  $k + 12$ , which is the local maximum of  $fG$ . The resulting dense pair-wise reconstruction for this keyframe pair can be seen in figure 4a. It can also be seen that the reprojection error for this frame pair is the local minimum. On the other hand there are similarly low reprojection errors in frames  $k + 4$  and  $k + 5$ . But as  $relGRIC$  and  $fG$  have negative values it can be seen that these frame pairs are degenerate. Figure 4b shows the

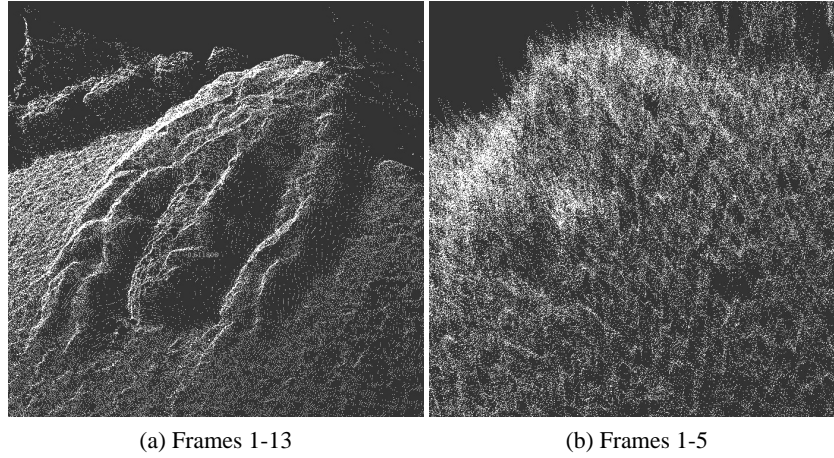


Fig. 4: Comparison of reconstructions between extracted keyframe pair 1-13 and low reprojection error frame pair 1-5 with linear triangulation for the ‘medusa’ sequence.

reconstruction result with frame  $k + 5$ . The reconstruction is clearly worse than with the found keyframe. This can be explained by the fact that the baseline is too small and therefore the noise in the point cloud, caused through numerical errors, is much bigger, even though the reprojection error is small. It becomes clear from this that pairwise reconstructions between keyframes should present low reprojection errors, but that reprojection error in itself should not be used as the decimation metric.

Between the frames  $k + 8$  to  $k + 11$  a zig-zagging in  $fG$  around zero and an explosion in reprojection error can be seen. It is assumed that in these areas the given correspondences and camera poses have high uncertainty. But these areas are decimated by the given algorithm, as negative  $fG$  values are not taken into account, which results in a good keyframe pair. The frame decimation was run on the entire sequence and results in good keyframe extraction. In this example the input sequence is decimated to 23 percent of the original frames. This high decimation rate can be explained by the slow camera movement compared to reconstruction target size and distance. The original baselines are small and therefore prone to numerical errors in triangulation. Figure 6b shows the first five of the extracted keyframes.

The given results show that the sequential work-flow in combination with the frame decimation metric  $fG$  results in good decimations. Based on the different movements and reconstruction targets covered in the tested data sets, it can be said that this approach is suitable for a large variety of sequences. A comparison between values of  $relGRIC$  and  $fG$  show that in cases where large areas of the target scene are covered by every frame the  $relGRIC$  and therefore the camera movement is dominant in the decimation decision. The correspondence-feature-ratio integrated into  $fG$  comes into play the less overlap there is in a given frame sequence. Two examples on opposite ends of these frame sequence attributes can be seen in the ‘medusa’ and ‘castle-P30’ data sets. The ‘medusa’ data set consists of a camera movement around the reconstruction target. This results in the fact that  $fG$  decimation differs from  $relGRIC$  only with larger base-

lines. In the ‘castle-P30’ case the  $fG$  metric results in more key frames as the overlap between frames is much smaller. Both these cases comply with our expectations and are valuable features for multi-view reconstructions.

Up until now it has been shown that the introduced algorithm decimates frames that are not suited for good SaM. Several tests have been performed with image sequences that have been carefully spaced before-hand, to allow good reconstructions. Figure 6e shows one of these sequences introduced by Zisserman et al. [12]. The frame decimation in this sequence results in no decimated frames and shows that there is no over-decimation.

To make sure that the introduced approach is independent of the spacing, or in other words, frame rate of the input data, tests have been conducted for a video sequence with different frames rates. This has been done with the ‘medusa’ data set [1], which was decoded into 5, 10, 15 and 20 fps to allow this comparison. The frame decimation was run for all the sequences, and the number of resulting keyframes and the corresponding number of input frames is given in figure 5a. It can be seen that the number of keyframes stays fairly constant. There is a small change in number of keyframes which can be explained by the fact that the introduced algorithm does not find a unique frame decimation, as the decimation is performed locally and sequentially, and the fact that in the lower frame rates possible keyframes are already excluded by the lack of spatial information.

To show how this frame decimation approach is well suited for sequential multi-view

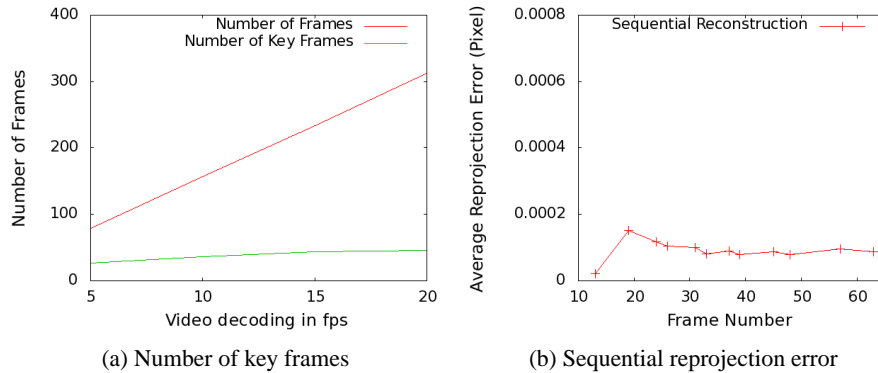


Fig. 5: Comparison between number of total frames and number of extracted keyframes for different ‘medusa’ sequence decodings and reprojection error during sequential reconstruction with decimated frames. The crosses represent keyframes and the corresponding reprojection error with every newly added keyframe is plotted.

reconstruction, the average reprojection error resulting after every addition of a new keyframe to the reconstruction is extracted. Figure 5b shows the average reprojection error in pixels after addition of every extracted keyframe. The keyframes are represented by the crosses. All other frames are decimated by the introduced approach. It can be seen that overall the reprojection error is small and also constant over time. This shows clearly that all the introduced keyframes are well suited for reconstruction. This



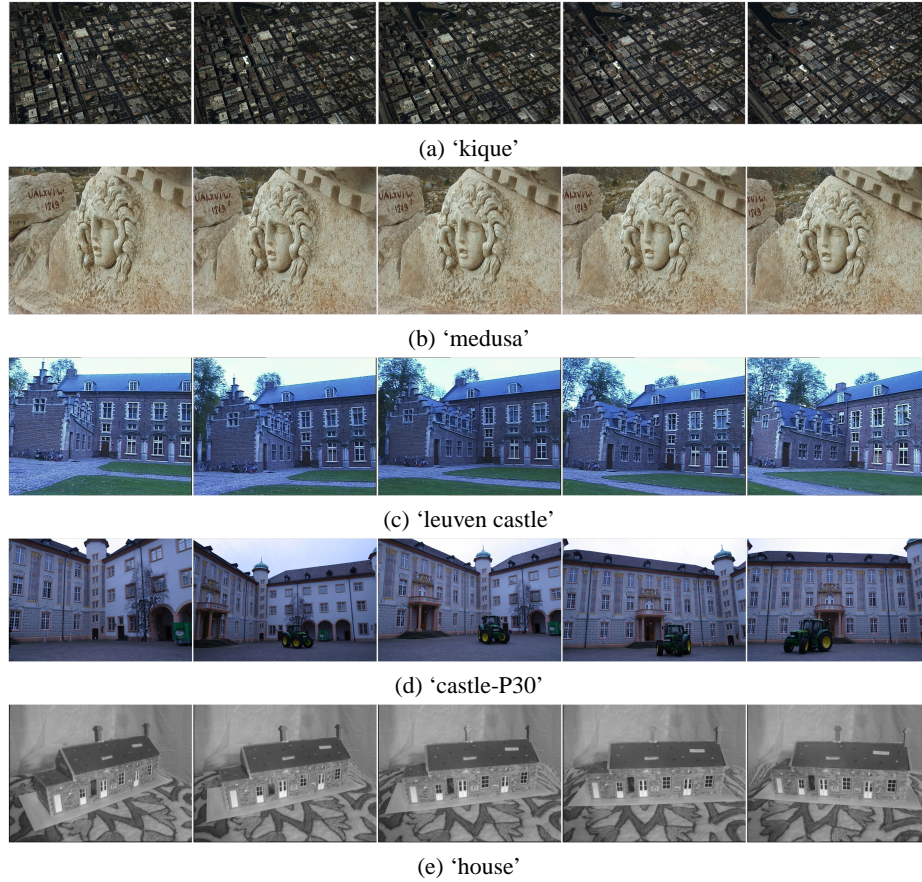


Fig. 6: First five keyframes extracted by non-parametric sequential frame decimation for different data sets.

means that degenerate camera poses as well as frames with large baselines and therefore bad correspondences are avoided.

## 5 Conclusion

This paper introduced a non-parametric sequential frame decimation algorithm for scene reconstructions in low-memory streaming environments. This frame decimation reduces the number of input images by eliminating frames that could lead to erroneous pose and structure estimates on the fly. The main contribution of this paper is the introduction of a sequential work-flow for frame decimation based on a newly introduced smooth frame goodness metric, which is designed to choose only one global maximum value at each keyframe evaluation. This allows for a sequential frame decimation with-

out usage of thresholds or scene assumptions. This metric is based on a ratio between good correspondences and possible features and the geometric robust information criterion (GRIC) for residual error relationships between epipolar geometry and homography estimation. The definition of this metric allows for a local error minimization evaluation and can therefore be used on the fly. Thanks to the sequential nature of this approach, less memory is used during the decimation evaluation as only three frames have to be kept in memory at any time. In comparison, previous work that suggested global decimation optimization uses at least ten frames at a time. This approach has been tested with multiple publicly available data sets representing different types of target scenes and camera movements. The results show reliable frame decimation robust to frame sampling rates and independent of any thresholds, scene assumptions or global frame analysis.

## References

1. Pollefeys, M., Van Gool, L., Vergauwen, M., Verbiest, F., Cornelis, K., Tops, J., Koch, R.: Visual modeling with a hand-held camera. *International Journal of Computer Vision* **59** (2004) 207–232 10.1023/B:VISI.0000025798.50602.3a.
2. Nistér, D.: Reconstruction from uncalibrated sequences with a hierarchy of trifocal tensors. In: *Computer Vision - ECCV 2000*. Volume 1842 of *Lecture Notes in Computer Science*. Springer Berlin / Heidelberg (2000) 649–663
3. Lowe, D.: Object recognition from local scale-invariant features. In: *ICCV*, Published by the IEEE Computer Society (1999) 1150–1157
4. Bay, H., Ess, A., Tuytelaars, T., Gool, L.V.: Speeded-up robust features (surf). *Computer Vision and Image Understanding* **110** (2008) 346 – 359 Similarity Matching in Computer Vision and Multimedia.
5. Nistér, D.: Frame decimation for structure and motion. In: *SMILE '00: Revised Papers from Second European Workshop on 3D Structure from Multiple Images of Large-Scale Environments*, London, UK, Springer-Verlag (2001) 17–34
6. Ahmed, M., Dailey, M., Landabaso, J., Herrero, N.: Robust key frame extraction for 3D reconstruction from video streams. In: *International Conference on Computer Vision Theory and Applications (VISAPP)*. (2010) 231–236
7. Torr, P.H.: Geometric motion segmentation and model selection. *Philosophical Transactions: Mathematical, Physical and Engineering Sciences* **356** (1998) 1321–1340
8. Royer, E., Lhuillier, M., Dhome, M., Lavest, J.M.: Monocular vision for mobile robot localization and autonomous navigation. *International Journal of Computer Vision* **74** (2007) 237–260 10.1007/s11263-006-0023-y.
9. Torr, P.H., Fitzgibbon, A.W., Zisserman, A.: The problem of degeneracy in structure and motion recovery from uncalibrated image sequences. *International Journal of Computer Vision* **32** (1999) 27–44 10.1023/A:1008140928553.
10. Beder, C., Steffen, R.: Determining an initial image pair for fixing the scale of a 3d reconstruction from an image sequence. In: *Pattern Recognition*. Volume 4174 of *Lecture Notes in Computer Science*. Springer Berlin / Heidelberg (2006) 657–666
11. Strecha, C., von Hansen, W., Van Gool, L., Fua, P., Thoennessen, U.: On benchmarking camera calibration and multi-view stereo for high resolution imagery. In: *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*. (2008) 1–8
12. Fitzgibbon, A.W., Cross, G., Zisserman, A.: Automatic 3D model construction for turn-table sequences. In Koch, R., Van Gool, L., eds.: *3D Structure from Multiple Images of Large-Scale Environments*, LNCS 1506. Springer-Verlag (1998) 155–170