УДК 004.932

Sharpening of Video Frames for Improvement of Feature Points Detection

S. A. Turko, I. V. Safonov

Some visual tracking tasks are unable to operate stable for video with blurred frames because feature points detection is degraded on blur images. In the paper we concentrate on blurriness due to refocusing of camera optical system of mobile devices. Our aim is evaluation of influence of blur and sharpening on Harris corner and FAST keypoints detectors. We propose a quality criterion in order to be able to make numerical assessment of the influence. For enhancement of blurred frames we investigate two sharpening filters: local tone mapping decreases of edge transition slope length, unsharp mask via bilateral filter increases local contrast along edges. On data from KITTI Vision Benchmark Suite as well as video captured by smartphone we demonstrate that sharpening improves feature points detection for both Harris corner and FAST methods. A positive impact for FAST is significant, it may improve robustness of various computer vision systems.

Key words: *sharpening, FAST keypoints, Harris corner, local tone mapping, unsharp mask.*

Introduction

One of the base problems of modern computer vision is a visual tracking. In particular, visual SLAM (Simultaneous Localization and Mapping) and odometry apply tracking of landmarks (or feature points) for pose detection [1]. Systems based on visual monocular or stereo SLAM are intended for robotics, augmented reality and navigation. However, current realizations of such systems lack the robustness needed to be useful in real life outside laboratory conditions. A tracking relies on a prior landmarks over a current video frame. Feature points have to be detected, matched and tracked frame-to-frame. The loss of a large number of tracks on successive frames leads to fail of a camera pose estimation and map creation, some computer vision applications are able to survive after break in tracking, but for visual SLAM and odometry it can be fatally.

There are a lot of causes which prevent keypoints detection and interrupt tracking, among them are sharp brightness alteration, motion blur due to rapid camera motion as well as movement of foreground or background regions, out-of-focus due to autofocusing/refocusing of optical system, handshaking, etc. A reducing of impact of such defects on video is actual problem in order to do visual SLAM applications more stable. Among mentioned above deformities, overcoming of motion blur is discussed in publications mainly. In the paper we concentrate on enhancement of video that comprises of sequences of frames affected by blur due to refocusing of camera optical unit. That case is typical when smartphone is used for indoor navigation by means of visual SLAM or odometry. Periodically depending on scene camera of smartphone tries to adjust focus, and it leads to blurring of several successive frames. Usually majority of tracks are interrupted on the blurred frames.

The aim of our study is to try to find answers on the following issues:

- How does blurriness affect on feature point detection and tracking?
- Is it possible to improve feature point detection on blurred video via detector parameters adjustment?

- Which is detector more robust for keypoints detection on blurred image?
- Is able a sharpening technique improve feature points detection?
- Which is a sharpening method perform better for that task?

Nowadays there are a lot of algorithms for calculation of feature points and local descriptors of images. Nonetheless, mobile and embedded systems require techniques which are capable to provide low power consumption, it is connected directly with low computational complexity. In practical application computationally inexpensive Harris corner detector and FAST method are used frequently rather more complex algorithms, such as SIFT, SURF, MSER, etc. Moreover, paper [2] demonstrates for visual SLAM task: in general Harris detector performs comparable or better to other interest point detectors and descriptors. In the paper we examine an influence of blurriness on Harris corner detector [3] and FAST technique [4] as well as propose an approach for improvement of detection of feature points by means of a sharpening of video frames. Figure 1 illustrates our concept.

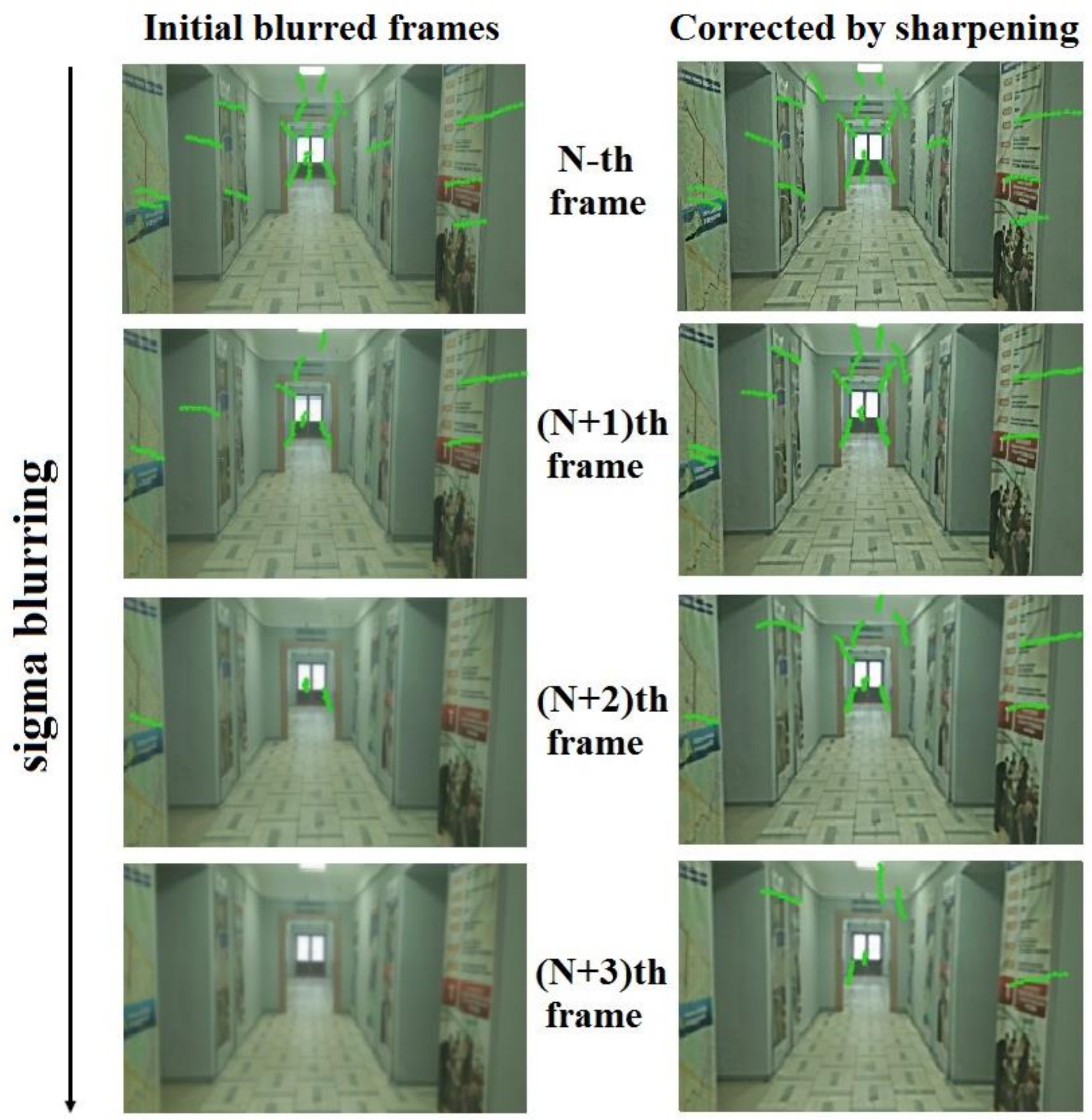

Figure 1. Sharpening of blurred frames improves tracking of landmarks

Related works

In the best of our knowledge, we do not familiar with publications devoted to issues connected with feature point detection on blurred frames due to refocusing of camera optics. Some relation to considering problem has studies which invent interest point detectors which are invariant to blurring. However such detectors have a high computational complexity, for instance, paper [5] describes detector based on Gabor multi-scale space.

The fact that number of detected features degrades with growing of blurriness may be used for construction of non-reference sharpness metric. Paper [6] poses a blind blurriness assessment comprising of the following stages: extraction Harris corner feature points from the blurred image and re-blurred image, then the two feature point maps are divided into blocks to generate block-wise maps and are combined to the feature point quantity similarity map, finally, an overall blurriness estimation is calculated by pooling of the similarity and visual saliency maps. The general idea of such approach is close to non-reference metric proposed by Crete [7], it is interesting to compare those two evaluation techniques. In any case, both methods allow to estimate sharpness of video frame.

As it is mentioned above, majority of existing publications are considering a performing of visual SLAM in a presence of motion blur. In [8] camera motion and 3D map are reconstructed by SLAM based on a few number of landmarks, and the information makes the estimation of motion blur. The blurred images are recovered by Lucy-Richardson deconvolution, and additional feature points are extracted from the restored image in order to repeat SLAM with a greater number of keypoints. So, SLAM and motion deblurring help each other. The approach has hard limitation: at least several landmarks have to be tracked for motion estimation. More practical way is a usage of inertial sensors for measuring of camera movement as it is suggested in [9]. A good camera movement estimation allows deblurring of frames affected by motion blur, but in the case of out of focus frames due to camera refocusing the approach does not work, it is necessary to obtain information directly from optical unit in order to do possible performing of deblurring.

It is worth mention dense tracking approach that becomes more and more popular nowadays. Paper [10] suggests to match regions by blurring of sharp image instead of deblurring of blur image and matching of sparse feature points. Method from the paper is intended for video with motion blur. Actually, it is good way for matching in the case of a lot of various video deformations, if you do not pay attention to a huge computational complexity. Nevertheless, an applicability of such approach for frames smoothed by Gaussian blur is questionable.

Figure of merit

We need in a figure of merit (or a quality criterion) in order to be able to estimate numerically influence of blurring and sharpening on feature point detection, matching and tracking. First of all, we selected several video files which do not contain blurred frames. For our experiments we borrowed fragments of two outdoor video from the KITTI Vision Benchmark Suite [11]. The dataset is intended for a lot of computer vision problems including visual odometry. The two test patterns are designated below as Video 1 and Video 2. Video 3 was captured by smartphone Samsung Galaxy S5. The video simulates an indoor navigation. Figure 1 shows several frames of Video 3. Each test pattern has duration about 10 s.

Next, we captured dozens of real HD video by cameras of various smartphones and searched frames which are blurred due to refocusing of optical system. The investigation of such frames revealed: a typical refocusing durations are less 1 s; smoothing of each separate frame can be modeled by a Gaussian blur; a variance of Gaussian blur grows almost linearly up to 7, after that it falls to 0. Based on those facts we distort our test video files as a sequence of refocusing procedures. Figure 2 illustrates our modelling of a given type of blur. *Ti* and variance of each refocusing chunk vary in the ranges [0.5, 1] s and [5, 7] correspondingly.

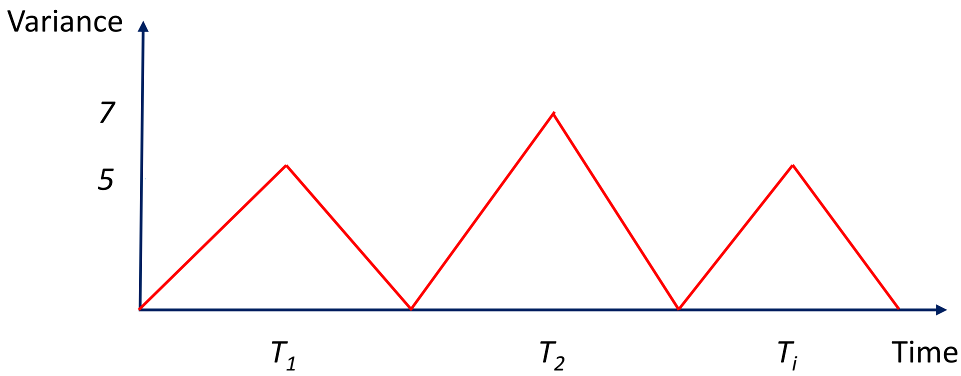

Figure 2. Refocusing simulation for video frames

Long tacks are preferable for visual SLAM and odometry, we take into account tracks with length greater than 10 that is tracks which we are able to track during 10 frames in sequence. For matching of interest points in adjacent frames Sum of Absolute Differences (SAD) of blocks 13x13 is used. We calculate tracks for initial good quality video and check presence the same tracks (or their long pieces) on distorted video. It is possible to have a small offset between corresponding key points when we compare tracks of initial and blurred video frames. We propose figure of merit as ratio of cumulative sum of lengths of long tracks of blurred video to cumulative sum of lengths of long tracks of initial video, whereas tracks on distorted video have to correspond to tracks on initial one:

$$Q = \frac{\sum_k P_k}{\sum_i E_i} \times 100\%, \tag{1}$$

where P_k is length of k^{th} track of processed pattern, *Ei* is length of i^{th} track of initial video.

For unprocessed or completely restored video Q=100%, for fatally distorted frames *Q*=0%. One can argue with proposed criterion, because it does not take into account new tracks which absence on initial frames and appear on processed ones. Theoretically it is true statement. However in practice a number of such new tracks is negligibly small. In general, proposed figure of merit coincides with our subjective assessments.

Feature points detection on blurred video

We estimated how that type of blur acts on detection of Harris corner and FAST keypoints. As we expected, a few number of interest points can be detected only. Tables below demonstrates that it is impossible to adjust of detectors thresholds T_H and T_F for Harris and FAST detectors accordingly in order to improve detection capability. Harris corner detector is a little bit more robust to blur than FAST. About 10% of tracks are survived for Harris corner detector that corresponds to about two dozen of tracks per video. Is it enough for visual SLAM/odometry? We have doubt. In the next chapters we try to improve situation by means of sharpening filters.

Table 1 – Figure of merit *Q* for Harris detector on blurred video

T_H	Video 1	Video 2	Video 3
0.05	6,47	9,76	15,78
0.08	5,22	8,64	14,20
0.11	5,12	7,86	15,38
0.14	4,89	7,90	13,20
0.17	6,54	8,79	12,67
0.2	5,68	7,94	12,87

Table 2 – Figure of merit *Q* for FAST detector on blurred video

T_F	Video 1	Video 2	Video 3
35	1,75	5,29	5,29
45	0,89	2,60	3,69
55	0,56	1,17	2,89
65	2,09	0,36	1,06
75	0	0,14	0,32
85	0	0	0

Sharpening algorithm

In spite of big advances of image restoration techniques, deconvolution algorithms are too slow for real-time video processing. Methods of image enhancements are more practical. Paper [12] proposes joint application of two image enhancement filters:

- Unsharp Mask (UM) via bilateral filter increases local contrast on the ends of edge transition slope;
- Local Tone Mapping (LTM) with an ordering decreases edge transition slope length.

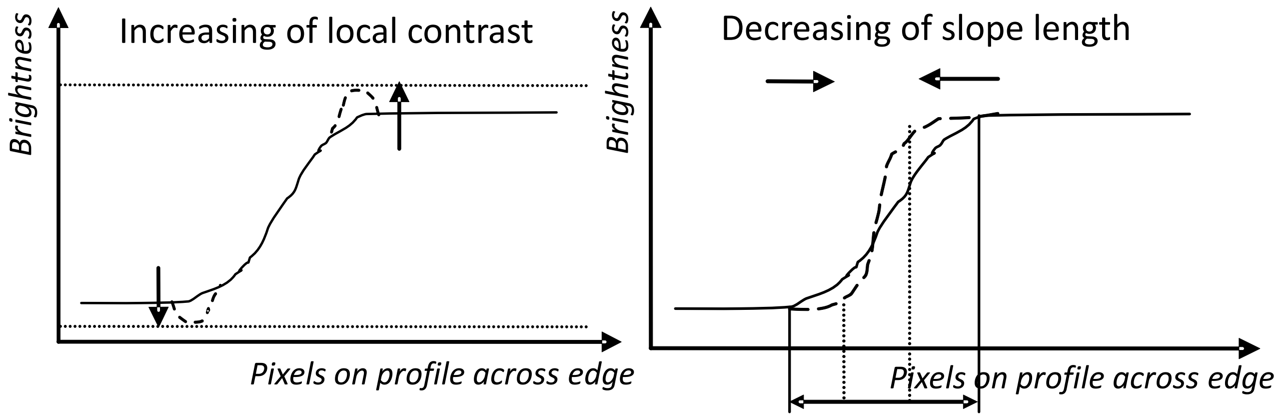

Figure 3. Illustration of UM and LTM filters
on brightness profile across blurred edge.

We adopted those two filters for sharpening of video. At first, an ordering in [12] is intended for processing of noisy images. Video for visual SLAM has a low noise level. Thus, we can drop an ordering, it give a positive impact on processing speed. At second, we add modification of color channels that follows after sharpening of brightness image. Brightness channel is calculated as:

$$Y(x,y) = MAX(R(x,y), G(x,y), B(x,y)), \tag{2}$$

where *(x,y)* is pixel coordinate, *R*, *G* and *B* are color channels of an image.

In LTM filter a square sliding window moves throughout a frame, and each pixel of *Y* is transformed by means of locally adaptive S-shaped curve:

$$Y_{LTM}(x,y) = \begin{cases} Y(x,y) : H(x,y) - L(x,y) \le T_{LTM} \\ L(x,y) + \dfrac{a(x,y)^2 \times (H(x,y) - L(x,y))}{a(x,y)^2 + (1 - a(x,y))^2} : otherwise \end{cases}, \tag{3}$$

where $a(x,y) = \dfrac{Y(x,y) - L(x,y)}{H(x,y) - L(x,y)}$, *L(x,y)* is minimal from *Y(x,y)* in local window $[x - R_{LTM}, y - R_{LTM}, x + R_{LTM}, y + R_{LTM}]$, *H(x,y)* is maximal form *Y(x,y)*, T_{LTM} is threshold for preventing processing of flat regions.

Unsharp mask is one of the most popular sharpening filter. The filter improves visual perception of an image by addition of combs along edges. However, the filter has several drawbacks: halo-artifacts forming, height of the combs depends on edge magnitude. UM via bilateral filter is free from those disadvantages. We use the following filter:

$$Y_{UM}(x,y)=\begin{cases} Y(x,y)+k(Y(x.y)-Y_{BF}(x,y)):\left|Y(x,y)-Y_{BF}(x,y)\right|\geq T_{UM} \\ Y(x,y)+\dfrac{k(Y(x,y)-Y_{BF}(x,y))}{2}:otherwise \end{cases}, \quad (4)$$

where Y_{BF} is filtered Y by bilateral filter, k is amplification factor, T_{UM} is threshold for preventing of noise level growing.

In contrast to a classical bilateral filter [13], that uses Gaussians as spatial kernel and photometric distance (or edge-stop function), we propose application of flat spatial kernel and edge-stop function, that on the one hand is similar to Gaussian and on the other hand does not tend to zero so rapidly. The filter blurs edges stronger in comparison with classical bilateral filter and weaker than a Gaussian smoothing. It prevent forming of halo-artifact that is typical for a conventional unsharp mask technique.

Our modification of a bilateral filter is calculated according to the following formulae:

$$Y_{BF}(x,y)=\frac{\sum_{i=-S/2}^{S/2}\sum_{j=-S/2}^{S/2}Y(x+i,y+j)D(Y(x+i,y+j)-Y(x,y))}{S^2\sum_{i=-S/2}^{S/2}\sum_{j=-S/2}^{S/2}D(Y(x+i,y+j)-Y(x,y))}, \quad (5)$$

where S is size of spatial kernel (or sliding window), D is photometric distance, that is calculated as:

$$D(z)=\frac{1}{\sqrt{1+\left|\dfrac{z}{1.5S^2}\right|^3}}. \quad (6)$$

LTM and UM filters can be used independently from each other. However joint application of both filters is capable provide a better outcomes. It is preferable to apply LTM before UM.

Finally, we need to make conversion of gray channel processed by sharpening filters to color image, because matching of feature points is carried out for color frames. Paper [14] poses a simple approach that preserves ratios between RGB channels with each other, accordingly the following expressions do not change hue and saturation of the processed color image in comparison with initial one:

$$R_e(x,y)=R(x,y)Y_e(x,y)/Y(x,y), \quad (7)$$

$$G_e(x,y)=G(x,y)Y_e(x,y)/Y(x,y), \quad (8)$$

$$B_e(x,y)=B(x,y)Y_e(x,y)/Y(x,y), \quad (9)$$

where Y_e is Y processed by sharpening filters.

Figure 4 shows example of consistent application of LTM and UM filters. One can see, edges become a sharper, processed frame has no annoying artifacts.

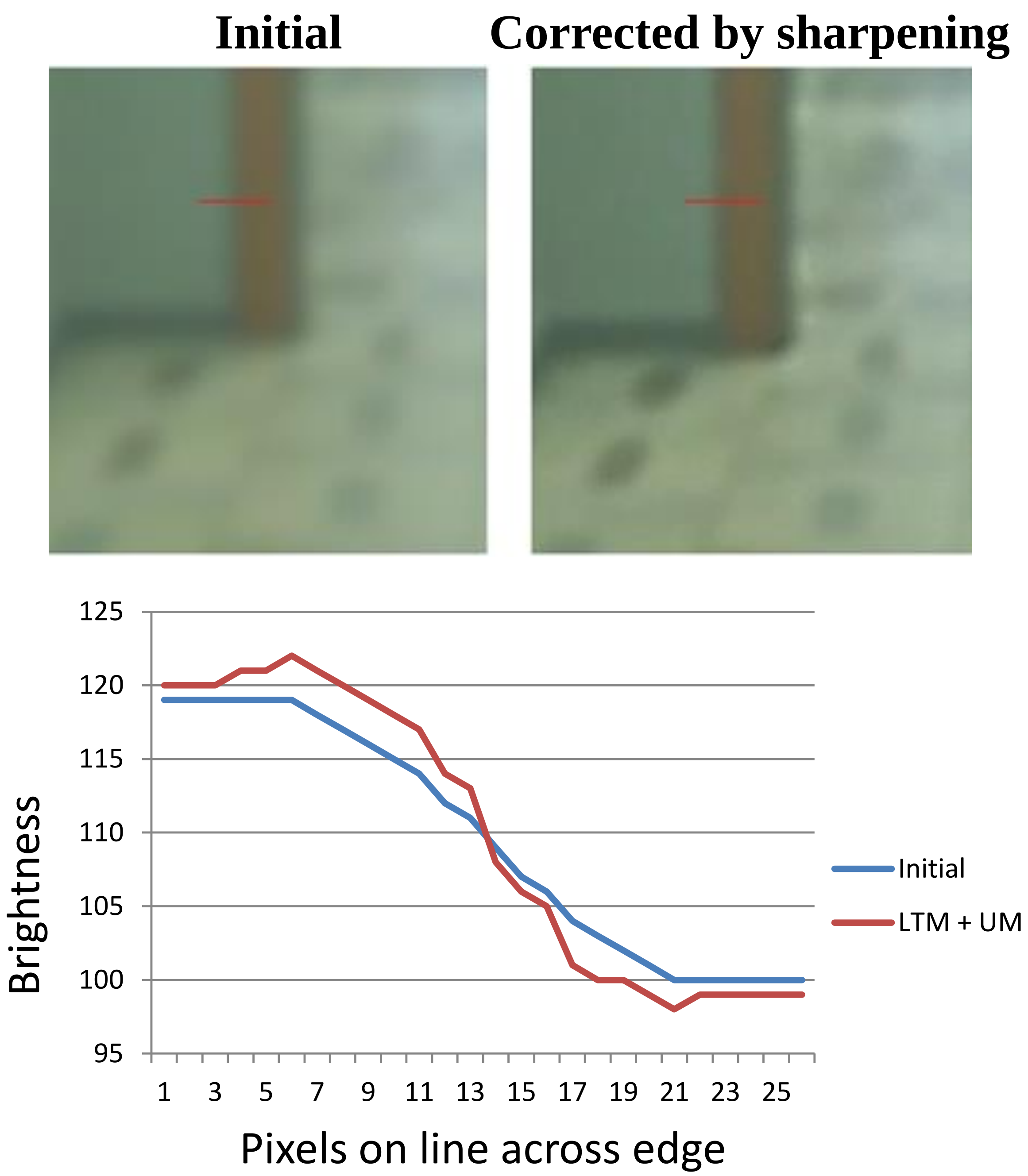

Figure 4. Example of sharpening by LTM+UM

Results

We processed our three blurred patterns by LTM, UM and combination (LTM+UM) of this filters with the following parameters: R_{LTM} =15, T_{LTM}=7, S=22, k=105 and T_{UM} =7. Such parameters ensure good visual quality of corrected video. Table below contains obtained outcomes for Harris corner and FAST detectors. For Harris corner detector an impact of LTM and UM is approximately equal. For FAST detector an impact of UM filter is in several times more in comparison with LTM. In both cases combination of LTM and UM filters provides a better results in comparison with outcomes of each separate filter.

Table 3 – Criterion *Q* for keypoints detectors on improved video

	Harris corner			FAST		
	LTM	UM	LTM+ UM	LTM	UM	LTM+ UM
Video 1	8,04	8,52	13,25	6,45	18,89	21,76
Video 2	11,30	8,08	12,89	5,45	17,30	28,56
Video 3	14,72	13,96	17,75	1,99	12,65	23,20

It is important to compare *Q* for blurred and sharpened video. Diagram on figure 5 allows doing that. In general, we can conclude: sharpening is unable to improve blurred video for Harris detector significantly; contrary, sharpening can improve detection ability of FAST detector in several times. Application of sharpening restores about 25% long tracks. In absolute numbers it corresponds to hundreds of tracks per 10 s of video. We ask ourselves again: is it enough for visual SLAM/odometry? It is hard to answer surely, but it should work in the most cases, although, it is not guarantee an appropriate improvement always.

Discussion and Future work

Our study demonstrates that sharpening is able to improve FAST feature points detection for blurred video due to refocusing of camera optical system. Of course, sharpening is a palliative solution that is unable to restore strongly blurred images. Nevertheless we encourage application of sharpening as preprocessing stage in visual SLAM and odometry tasks, where a break in tracking leads to inconsistency of algorithms often.

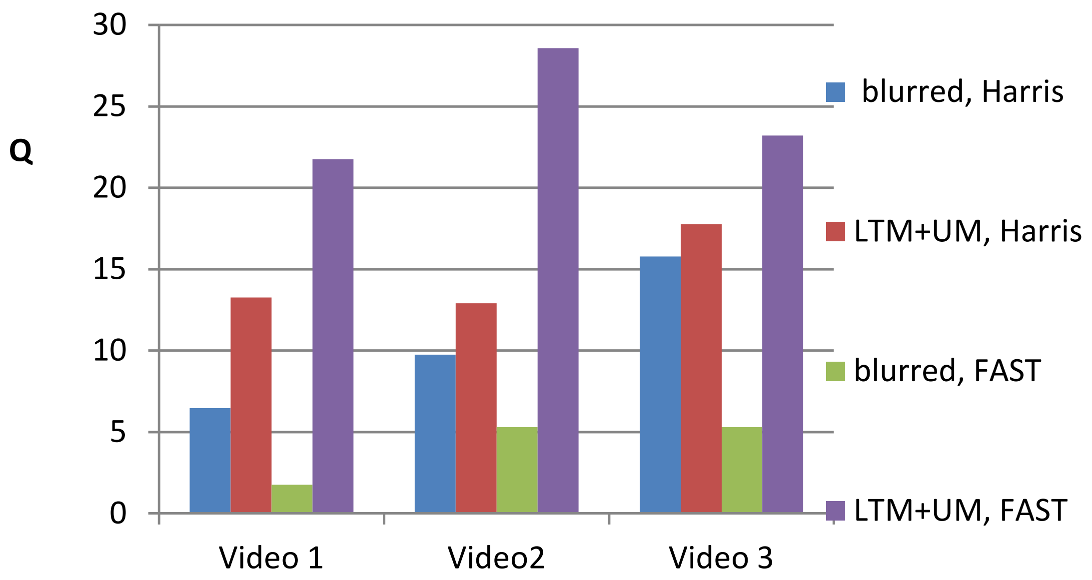

Figure 5. Impact of blur and sharpening on feature detection

Additional research is necessary for various sharpening algorithms as well as different feature points and descriptors. In particular, we revealed that increasing of amplification factor *k* of UM filter leads to growing of *Q* for both analyzed feature points algorithms, but video becomes to be unpleasant visually. Other topical problem is reduction of influence on feature points detection other types of blurriness including motion blur. According to our preliminary experiments, sharpening a little bit helps in the case of motion blur too. However, a deeper research is necessary. In the future we are going to investigate much more test video files for various types of blur as well as several sharpening techniques including adaptive approaches, where filter parameters are adjusted depending on blurriness of image. Also, an actual research subject is development of novel blind sharpness metrics based on feature points.

References

1. Munguia R., Gra A. Monocular SLAM for visual odometry. *Proc. of IEEE International Symposium on Intelligent Signal Processing (WISP 2007)*, 2007, pp. 1-6.

2. Gil A., Mozos O. M., Ballesta M., Reinoso O. A comparative evaluation of interest point detectors and local descriptors for visual SLAM. *Machine Vision and Applications*, 2010, vol. 21, no. 6, pp. 905-920.

3. Harris C., Stephens M. A combined corner and edge detector. *Proc. of Alvey vision conference*¸ 1988, vol. 15, p. 50.

4. Rosten E., Drummond T. Machine learning for high-speed corner detection. *Proc. of Computer Vision conf. (ECCV 2006)*, 2006, pp. 430-443.

5. Xie J., Cai Z. An effective blur invariant interest point detector. *Proc. of IEEE International Conference on Virtual Reality and Visualization (ICVRV 2013)*, 2013, pp. 281-284.

6. Cai H., Li L., Qian J., Pan J. S. Image Blur Assessment with Feature Points. *Journal of Information Hiding and Multimedia Signal Processing*, 2015, vol. 6, no. 3, pp. 482-490.

7. Crete F., Dolmiere T., Ladret P., Nicolas M. The blur effect: perception and estimation with a new no-reference perceptual blur metric. *Proc. of Electronic Imaging*, 2007, pp. 64920I-64920I-11.

8. Lee H. S., Kwon J., Lee K. M. Simultaneous localization, mapping and deblurring. *Proc. of IEEE International Conference on Computer Vision (ICCV 2011)*, 2011, pp. 1203-1210.

9. Klein G. S., Drummond T. W. Tightly integrated sensor fusion for robust visual tracking. *Image and Vision Computing*, 2004, vol. 22, no. 10, pp.769-776.

10. Jin H., Favaro P., Cipolla R. Visual tracking in the presence of motion blur. *Proc. of IEEE Computer Vision and Pattern Recognition conf. (CVPR 2005)*, 2005, vol. 2, pp. 18-25.

11. Geiger A., Lenz P., Urtasun R. Are we ready for autonomous driving? The kitti vision benchmark suite. *Proc. of IEEE Computer Vision and Pattern Recognition conf. (CVPR 2012)*, 2012, pp. 3354-3361.

12. Safonov I. V., Rychagov M. N., Kang K. M., Kim S. H. Adaptive sharpening of photos. *Proc. of Electronic Imaging*, 2008, pp. 68070U-68070U-12.

13. Tomasi C., Manduchi R. Bilateral filtering for gray and color images. *Proc. of IEEE International Conference on Computer Vision*, 1998, pp. 839-846.

14. Safonov I. V., Automatic correction of amateur photos damaged by backlighting. *Proc. of International Conference on Computer Graphics and Vision (GraphiCon 2006),* 2006, pp. 80-89.

Литература

1. Munguia R., Gra A. Monocular SLAM for visual odometry // IEEE International Symposium on Intelligent Signal Processing (WISP 2007). 2007. C. 1-6.

2. Gil A., Mozos O. M., Ballesta M., Reinoso O. A comparative evaluation of interest point detectors and local descriptors for visual SLAM // Machine Vision and Applications. 2010. Т. 21. № 6. С. 905-920.

3. Harris C., Stephens M. A combined corner and edge detector // Alvey vision conference. 1988. Т. 15. – С. 50.

4. Rosten E., Drummond T. Machine learning for high-speed corner detection // Computer Vision conf. (ECCV 2006). 2006. С. 430-443.

5. Xie J., Cai Z. An effective blur invariant interest point detector // IEEE International Conference on Virtual Reality and Visualization (ICVRV 2013). 2013. С. 281-284.

6. Cai H., Li L., Qian J., Pan J. S. Image Blur Assessment with Feature Points // Journal of Information Hiding and Multimedia Signal Processing. – 2015. Т. 6. № 3. С. 482-490.

7. Crete F., Dolmire T., Ladret P., Nicolas M. The blur effect: perception and estimation with a new no-reference perceptual blur metric // Electronic Imaging. 2007. С. 64920I-64920I-11.

8. Lee H. S., Kwon J., Lee K. M. Simultaneous localization, mapping and deblurring // IEEE International Conference on Computer Vision (ICCV 2011). 2011. С. 1203-1210.

9. Klein G. S., Drummond T. W. Tightly integrated sensor fusion for robust visual tracking // Image and Vision Computing. - 2004. Т. 22. № 10. С. 769-776.

10. Jin H., Favaro P., Cipolla R. Visual tracking in the presence of motion blur // IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2005). 2005. Т. 2. С. 18-25.

11. Geiger A., Lenz P., Urtasun R. Are we ready for autonomous driving? The kitti vision benchmark suite // IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2012). 2012. С. 3354-3361.

12. Safonov I. V., Rychagov M. N., Kang K. M., Kim S. H. Adaptive sharpening of photos // Electronic Imaging. 2008. С. 68070U-68070U-12.

13. Tomasi C., Manduchi R. Bilateral filtering for gray and color images // IEEE International Conference on Computer Vision. 1998. – С. 839-846.

14. Safonov I. V. Automatic correction of amateur photos damaged by backlighting // International Conference on Computer Graphics and Vision (GraphiCon 2006). 2006. С. 80-89.

Статья поступила 21 февраля 2016 г.

Information about Authors

Sergey A. Turko - a student of National Research Nuclear University MEPhI. His research interest includes image processing and computer vision. E–mail: sergey.turko94@gmail.com

Ilia V. Safonov – Ph.D. in Computer Sciences. Associate professor of the Department of Computer Science and Control Systems. National Research Nuclear University MEPhI. Field of research: signal and image processing, computer vision, pattern recognition. E–mail: ilia.safonov@gmail.com

Address: Russia, 115409, Moscow, Kashirskoe highway, 31.

Повышение резкости видеокадров для улучшения обнаружения ключевых точек

Турко С. А., Сафонов И. В.

Актуальность. *Системы компьютерного зрения, относящиеся к визуальной навигации, одометрии и слежению, работают нестабильно на видео с размытыми подпоследовательностями кадров, так как на нечетких изображениях детектируется существенно меньше ключевых (или особых) точек. В данной работе рассматривается размытие, возникающее в процессе расфокусировки или перефокусировки оптических систем мобильных устройств. Предлагается математическая модель такого типа размытия.* ***Целью исследования является*** *оценка влияния размытия и последующего повышения резкости на обнаружение особых точек детектором углов Харриса и алгоритмом FAST. Рассматриваются два фильтра повышения резкости: локальное преобразование тонов, которое сокращает протяженность контурного перепада, и фильтр нерезкого маскирования, повышающий локальный контраст вдоль перепада. В фильтре нерезкого маскирования вместо Гауссова размытия используется билатеральный фильтр.* ***Результаты работы*** *фильтров демонстрируются на данных из тестового набора KITTI Vision Benchmark Suite и на видео, снятом камерой смартфона. Согласно предложенному референсному критерию качества, применение к видео фильтров, повышающих резкость, ведет к росту доли обнаруживаемых особых точек для обоих рассматриваемых детекторов. Причем положительное влияние последовательного применения фильтров локального преобразования тонов и нерезкого маскирования на FAST детектор весьма значительно, что позволяет в ряде случаев повысить стабильность функционирования систем компьютерного зрения.*

Ключевые слова: *повышение резкости, детектор особых точек FAST, детектор углов Харриса, локальное преобразование тонов, фильтр нерезкого маскирования.*

Информация об авторах

Турко Сергей Александрович – студент. Национальный исследовательский ядерный университет МИФИ. Область научных интересов: обработка изображений и компьютерное зрение. E–mail: sergey.turko94@gmail.com

Сафонов Илья Владимирович – кандидат технических наук. Доцент кафедры информатики и процессов управления. Национальный исследовательский ядерный университет МИФИ. Области научных интересов: обработка сигналов и изображений, компьютерное зрение, распознавание образов. E–mail: ilia.safonov@gmail.com

Адрес: 115409, Россия, г. Москва, Каширское шоссе, 31.