

## VIDEO COMPLETION IN DIGITAL STABILIZATION TASK USING PSEUDO-PANORAMIC TECHNIQUE

M. N. Favorskaya<sup>a,\*</sup>, V. V. Buryachenko<sup>a</sup>, A. G. Zotin<sup>a</sup>, A. I. Pakhirka<sup>a</sup>

<sup>a</sup> Institute of Informatics and Telecommunications, Reshetnev Siberian State Aerospace University, 31, Krasnoyarsky Rabochy av., Krasnoyarsk, 660037 Russian Federation – (favorskaya, buryachenko, zotin, pahirka@sibsau.ru)

Commission II, WG II/5

**KEY WORDS:** Digital stabilization, Video completion, Motion inpainting, Scene restoration

### ABSTRACT:

Video completion is a necessary stage after stabilization of a non-stationary video sequence, if it is desirable to make the resolution of the stabilized frames equalled the resolution of the original frames. Usually the cropped stabilized frames lose 10-20% of area that means the worse visibility of the reconstructed scenes. The extension of a view of field may appear due to the pan-tilt-zoom unwanted camera movement. Our approach deals with a preparing of pseudo-panoramic key frame during a stabilization stage as a pre-processing step for the following inpainting. It is based on a multi-layered representation of each frame including the background and objects, moving differently. The proposed algorithm involves four steps, such as the background completion, local motion inpainting, local warping, and seamless blending. Our experiments show that a necessity of a seamless stitching occurs often than a local warping step. Therefore, a seamless blending was investigated in details including four main categories, such as feathering-based, pyramid-based, gradient-based, and optimal seam-based blending.

### 1. INTRODUCTION

Over the two last decades, many digital video stabilization methods were developed for various practical applications. These methods are categorized into 2D and 3D stabilization frameworks that are specified in different ways. While the 2D methods utilize the estimation of 2D motion model, smooth motion model for removal of unwanted jitters, and application of the geometric transforms to the video frames (Cai and Walker, 2009; Favorskaya and Buryachenko, 2015), the 3D methods use the structure-from-motion model with following video reconstruction based on the interpolated low frequency 3D camera path (Liu et al., 2009). In the both cases, the stabilized areas of frames are reduced in a comparison to the original non-stationary videos. Thus, the task of video enhancement (called as a video completion) in order to fill the blank regions on boundaries of the stabilized frames appears. The early known methods, such as a mosaicing (Litvin et al., 2003), a motion deblurring (Yitzhakyet al., 2000), a point spread function (Chan et al., 2005), a motion inpainting (Matsushita et al., 2006), among others, are utilized for restoration of every frame independently. However, the stitching provides the artifacts for the non-planar scenes, especially. Therefore, the improvement of the stitching procedures remains desirable.

Video inpainting is an extension of single image inpainting that can be classified into the Partial Differential Equation (PDE) image inpainting algorithms (Bertalmio et al., 2000) and the texture synthesis techniques (Liang et al., 2001). The PDE-based techniques analyze the gradient vectors around the inpainting regions with following diffusion of the adjacent linear structures into the missing areas. Nevertheless, the restored regions may become too blurred when enlarged and

such procedure is a time-consuming process. The texture synthesis techniques select the basic textural patches of the sample texture and past the cloned patches into missing area. The patch-based sampling algorithm is fast but suffers from mismatching features across patch boundaries. Hereinafter, the hybrid techniques were proposed. Thus, Criminisi et al. (Criminisi et al., 2004) proposed the efficient algorithm based on an isophote-driven image sampling process, including a simultaneous propagation of texture and structure information.

The main shortcoming is in that the discontinuous textures or edges may appear in some regions. For video inpainting, additionally a temporal continuity ought to be considered (Favorskaya et al., 2013). Three approaches of video inpainting algorithms prevail including the patch search in the spatiotemporal domain, separate processing of object and background (object-background splitting), and structural and texture classification.

The remainder of our paper is organized as follows. The overview of the related work is presented in Section 2. In Section 3, the proposed construction of pseudo-panoramic key frame is outlined. The suggested inpainting of missing boundary areas of the successive frames is described in Section 4. Section 5 provides the experimental results illustrated by the restored frames. Finally, conclusions are covered in Section 6.

### 2. RELATED WORK

Video stabilization includes three steps, such as the camera motion estimation, motion smoothing, and image warping, which are differently implemented by 2D and 3D methods. However, after jitters' compensation the missing boundary areas appear that requires a video completion in order to save a

\* Corresponding author

resolution of video sequence. Consider the main approaches of video inpainting.

The patch search inpainting is based on a volume matching colour components and spatiotemporal gradient vectors (Wexler et al., 2007). Nevertheless, this method was applicable to low-resolution videos because of the time-consuming during a volume matching process. Furthermore, the concept of recovering the motion information before the spatial reconstruction, using the patches, had been developed. Thus, Shih et al. (Shih et al., 2009) suggested a motion segmentation algorithm for object tracking based on separating the video layers. This permitted to process the slow and fast moving objects differently in order to avoid “ghost shadows” in the resulting videos. Shiratori et al. (Shiratori et al., 2006) formulated the fundamental idea of motion warping, when the motion parameter curves of an object are warped. The motion field transfer is computed on the boundary of the holes, progressively advancing towards the inner holes, using the copied source patches with the corresponding motion vectors. However, this approach suffers from such effects as deformation, object occlusion, and background uncovering.

In order to avoid the problem mentioned above, the technique of the object-background splitting with the following separate inpainting was developed in 2010s (Jia et al., 2004; Patwardhan et al., 2005; Jia et al., 2006; Patwardhan et al., 2007). Jia et al. (Jia et al., 2006) developed a video completion system, which was capable to synthesize the missing pixels completing the static background and moving cyclically objects. They introduced the special term “movelet” that means a structured moving object and provided a video repairing of large moving motion by sampling and aligning movelets. The Lambertian, illumination, spatial, and temporal consistencies were maintained after background completion. Particularly, the temporal consistency was preserved by the wrapping and regularization of moving regions. Levin et al. (Levin et al., 2004) studied an image stitching in gradient domain and introduced several cost functions to evaluate the similarity to the input images and visibility of a seam in the output image. Hereinafter, more complex algorithms were designed. Xia et al. (Xia et al., 2011) constructed the Gaussian mixture models for both background and foreground separately that saved the time for calculating the optical flow mosaics for the foreground objects only. This approach is worked well for individual inpainting objects but fails under highly complex contents or irregular moving objects in a scene.

The inpainting based on structural and texture classification is usually based on the PDE functions. The images are divided into the structural and textural regions and each type of regions is processed by different inpainting technique according to the assigned priority. Canny edge detector (Canny, 1986) is often employed to identify the edge pixels in adjacent frames. Fang and Tsai (Fang and Tsai, 2008) performed a temporal linear interpolation in the structural regions. For the textural regions, a propagation filling method is used to preserve the spatial consistency. Sangeetha et al. (Sangeetha et al., 2011) proposed a method, including the image decomposition (an image is represented as a cartoon-like version of the input image, where the large-scale edges are preserved but interior regions are smoothed), structural part inpainting (an image interpolation using a third order PDE), damaged area classification (a reduction a number of pixels to be synthesized), and improved exemplar based texture synthesis. The last step consists mainly of three iterative steps, until all pixels in the inpainted region

are filled, such as the computing patch priorities, propagating structure and texture information, and updating confidence values. However, the structural temporal linear interpolation cannot work well for rapidly moving objects, irregular motion paths, or large inpainting regions.

The short literature review shows that the ideal approach for a video inpainting does not exist. The good way is to use the multi-resolution and multi-layered video inpainting with the irregular patch matching and seamless blending in a frequency domain.

### 3. CONSTRUCTION OF PSEUDO-PANORAMIC KEY FRAME

The real-world videos can be described in different manner, for example, a steady-scene with small range of motion, a steady-scene with large range of motion, a homogeneous background with small range of motion, an outdoor-shooting with small moving objects, an indoor-shooting with forward translation, an on-road shooting with forward moving platform, and so on. A scene with a homogeneous background with small range of motion is depicted in Figure 1, while a scene obtained from an on-road shooting with forward moving platform is depicted in Figure 2. These figures show the results of video stabilization using three methods, such as Derivative Dynamic Time (DDT) warping based on angular integral projections (Veldandi et al., 2013), Fourier Radon (FRadon) warping (Mohamadabadi et al., 2012), and Differential-Radon (DRadon) curve warping (Shukla et al., 2017). As it can be seen, all methods produce the cropped frames that caused a necessity of video completion.

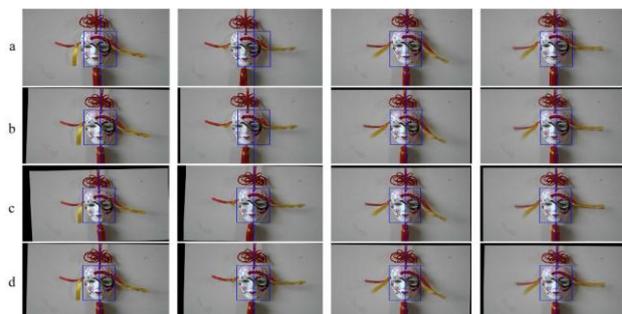


Figure 1. Scene with a homogeneous background with small range of motion: a original frames, b the DDT warping, c the FRadon warping, d the DRadon warping



Figure 2. Scene of on-road shooting with forward moving platform: a original frames, b the DDT warping, c the FRadon warping, d the DRadon warping

While a scene with a homogeneous background from Figure 1 requires the simplest restoration like a mosaic, a scene of on-

road shooting with forward moving platform demands the complex inpainting of many details.

Our approach deals with a preparing of pseudo-panoramic key frame during a stabilization stage as a pre-processing step for the following inpainting. Note that the extension of a view of field may appear due to the pan-tilt-zoom unwanted camera movement. When the stabilized video frames with the decreased sizes will be obtained and aligned respect to the original sizes, they can be reconstructed using the pseudo-panoramic key frame as a first approximation. Then a gradient-based inpainting, using the model of the multi-layered motion fields, is utilized in order to improve the details of the missing regions. A pseudo-panoramic key frame is an extension of the original key frame by the parts from the following frames with the random sizes. In the case of the non-planar frames with moving objects, several pseudo-panoramic key frames can be constructed but their number is very restricted between two successive key frames and helps to find the coinciding regions very fast. A schema of this process is depicted in Figure 3.

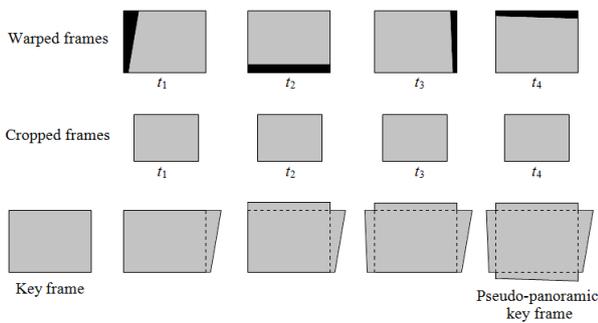


Figure 3. Scheme of pseudo-panoramic key frame receiving

The use of the pseudo-panoramic key frame/frames provides the background information in the missing boundary areas. Such approach leads to the simplified stitching step against the conventional panoramic image creation.

#### 4. INPAINTING OF MISSING BOUNDARY AREAS

Unlike the conventional problem statement of image/videos inpainting, a video completion of the stabilized video sequences deals with the missing areas near boundaries. This means that the source of reliable information is predominantly located on one side but not around the missing area, which is usually looks like as a hole with a random shape. Our approach is based on a multi-layered representation of each frame including the background and objects, moving differently. The proposed algorithm involves four steps, such as the background completion, local motion inpainting, local warping, and seamless blending, describing in Sections 4.1-4.4, respectively. Note that the background completion is always required, while other steps operate when it is necessary.

##### 4.1 Background Completion

The availability of video sequence permits to consider the task of video completion as an extension of 2D successive frames' completion to the spatiotemporal space. The temporal consistency in the filled areas is guaranteed. At this step, the proposed algorithm, using a pseudo-panoramic key frame, does not require to separate the video into the background and dynamic foreground objects, extract the motion layers, or use

the spatiotemporal patches from other frames. A stack of the non-cropped frames between the adjacent key frames provides reliable background information in the missing boundary areas. However, this is only the first but usually not final step of video completion. The background may be changed by the motion layers, as well as by artifacts of illumination.

##### 4.2 Local Motion Inpainting

Due to a local motion occupies a small area in the successive frames, it is reasonable to suppose that the motion information is sufficient to reconstruct the motion areas in the problematic boundaries. Thus, it is required to find the similarity in motion in order to provide a good matching of the motion patches. Let the estimated by the optical flow motion vector at point  $\mathbf{p} = (x, y, t)^T$  be denoted as  $(u(\mathbf{p}), v(\mathbf{p}))^T$  and the motion vector  $\mathbf{m}$  be defined as  $\mathbf{m}_p \equiv (u(\mathbf{p})t, v(\mathbf{p})t, t)^T$ . Then, according to Shiratori et al. (Shiratori et al., 2006) the error  $\varepsilon_p$  can be computed by equation 1:

$$\varepsilon_p = \arg \min_{(u(\mathbf{p}), v(\mathbf{p}))} \sum \left( u(\mathbf{p}) \frac{\partial I}{\partial x} + v(\mathbf{p}) \frac{\partial I}{\partial y} + \frac{\partial I}{\partial t} \right), \quad (1)$$

where  $I$  = an intensity function

$\frac{\partial I}{\partial x}, \frac{\partial I}{\partial y}$  = the derivatives along the spatial axes

$\frac{\partial I}{\partial t}$  = a derivative along the temporal axis

The dissimilarity measure of motion patches can be easily estimated calculating a distance between two motion vectors  $\mathbf{m}_p$  and  $\mathbf{m}_q$  using the angular difference  $\varepsilon_m(\mathbf{m}_p, \mathbf{m}_q)$  (Barron et al., 1994) using equation 2:

$$\varepsilon_m(\mathbf{m}_p, \mathbf{m}_q) = 1 - \frac{\mathbf{m}_p \cdot \mathbf{m}_q}{\|\mathbf{m}_p\| \|\mathbf{m}_q\|} = 1 - \cos \theta, \quad (2)$$

where  $\mathbf{m}_p$  and  $\mathbf{m}_q$  = the motion vectors in points  $\mathbf{p}$  and  $\mathbf{q}$   
 $\theta$  = an angle between these motion vectors

The angular error measure  $\varepsilon_m$  evaluates the differences in both directions and the differences in magnitudes of motion vectors. The dissimilarity measure between the source patch  $P_s$  and the target patch  $P_t$  as the weight can be calculated using equation 2. The best matching patch fills a corresponding area. This is an iterative procedure, which is executed until the averaged dissimilarity measure will be less a predefined threshold value.

##### 4.3 Local Warping

The pure warping means that the points are mapped to the points without changing their colors. This means that in rare cases such distortions can be based mathematically but usually they cannot be described by the known dependencies like the planar affine or planar perspective (also called homography) transforms.

Sometimes, especially when an optical flow is mismatched, a local warping helps to improve a visibility of the moving objects. Only small parts of the moving objects require such time-consuming step. In the most cases, the blurring or linear interpolation is enough, whereas the non-parametric sampling (Wexler et al., 2004), diffusion methods (Bugeau and

Bertalmio, 2009), or local homography fields for the complicated scenes (Liu and Chin, 2015) may be also used. The necessity of this step is determined by the type of a scene and requirements to the final results.

#### 4.4 Seamless Blending

Our experiments show that a necessity of a seamless stitching occurs often than a local warping step. Therefore, a seamless blending was investigated in details including four main categories, such as feathering-based, pyramid-based, gradient-based, and optimal seam-based blending. The application of simple  $\alpha$ -blending method provides a disappointing visibility.

The methods from feathering-based category perform a blending operation using an average value in each pixel of the overlapping region. The simplest way is to calculate an average value  $C(x)$  at each pixel:

$$C(x) = \sum_k w_k(x) \tilde{I}_k(x) / \sum_k w_k(x), \quad (3)$$

where  $\tilde{I}_k(x)$  = the warped frames  
 $w_k(x)$  = the weights, which are equal 1 at valid pixels and 0 otherwise.

However, the simple averaging fails under the exposure differences, misalignments, and presence of moving objects. Szeliski (Szeliski, 2006) mentioned that a superior method is to use the weighted averaging along with a distance map, when the pixels near the center are weighted heavily and weighted lightly near the edges. In terms of the Euclidean distance this expression has a view of equation 4:

$$w_k(x) = \left\| \arg \min_y \left\{ \|yy\| \mid \tilde{I}_k(x+y) \text{ is invalid} \right\} \right\|. \quad (4)$$

Such weighted averaging using a distance map is called feathering-based blending. The distance map values can be raised using the parameter  $w_k(x)$  in large power in equation 3. Peleg et al. (Peleg et al., 2000) proposed a visibility mask-sensitive version applying the familiar Vornoi diagram for a matching of each pixel to the nearest center in the set. However, it is difficult to obtain a balance between the smooth low-frequency exposure differences and preservation of sharp transitions. Also, these methods suffer from ghosting artifacts.

The main goal of pyramid-based blending is to separate the low frequency and high frequency components and to blend the low frequency component gradually over a wide area around the seam, while the high frequency component ought to concentrate over a narrow area around the seam (Pandey and Pati, 2013). This provides a gradual transition around the seam. Usually the methods from this category include two steps. First, a mask image associated with a source image is created using, for example, a grassfire transform (Xiong and Turkowski, 1998). Second, the mask image is converted into a low-pass pyramid using a Gaussian kernel. These steps are repeated for different levels with various sample density and resolution. The resultant blurred and subsampled masks serve as the weights to perform a per-level feathering. Then the final mosaic is interpolated and summarized according to equation 5:

$$LO(x, y) = GM(x, y) * LI_1(x, y) + (1 - GM(x, y)) * LI_2, \quad (5)$$

where  $LI_1(x, y)$  and  $LI_2(x, y)$  = the Laplacian pyramids of the warped source images  $I_1(x, y)$  and  $I_2(x, y)$

$GM(x, y)$  = the Gaussian pyramid of the mask image

$M(x, y)$

$LO(x, y)$  = the Gaussian pyramid of the output image

$O(x, y)$

Such algorithms demonstrate a reasonable balance between the smoothing out of the low frequency components and preserving sharp transitions to prevent blurring. The edge duplication is eliminated. Nevertheless, the double contouring and ghosting effects are possible.

The gradient-based methods utilize the idea of a suitably mixing the gradient of images. The humans are more sensitive to the gradients than the image intensities. Thus, these methods provide more pleasant visual results respect to the categories mentioned above. Levin et al. (Levin et al., 2004) developed the Gradient-domain Image Stitching (GIST) framework in two variants. The GIST1 computed the stitched image by minimizing a cost function as a dissimilarity measure between the derivatives of the stitched image and the derivatives of the input images. The GIST2 employed additionally the minimization of a dissimilarity measure between the derivatives of the input images and a field stitching. Xiong (Xiong, 2009) eliminated the ghosting artifacts that appear from the moving objects. The gradient vector field constructed by solving a Poisson equation with boundary conditions was utilized for the blending images. Szeliski et al. (Szeliski et al., 2011) proposed a method for fast Poisson blending and gradient domain compositing based on a multi-spline representation of the separate low-resolution offset field associated with each source image. The main shortcoming is that the gradient-based methods require higher computational resources and perfect alignment of initial images.

The gradient-based inpainting uses the multi-layered motion fields of foreground objects or the local homography fields for the more complicated scenes (Pérez et al., 2003). The following blending step removes some discontinuities during a filling process. Such techniques as a simple smoothing based on the pixels' intensities, the extrapolation, or the smoothing based on the Poisson equation are applied for the textural regions. The Poisson equation in the spatial domain has a view of simultaneous linear equations 6 for all  $p \in \Omega$ :

$$|N_p| f_p - \sum_{q \in N_p \cap \Omega} f_q = \sum_{q \in N_p \cap \partial\Omega} f_q^* + \sum_{q \in N_p} \text{div}_{pq}, \quad (6)$$

where  $\Omega$  = a missing area

$p$  = the pixels in the missing area

$q$  = the pixels in the known area

$|N_p|$  = a number of neighboring pixels  $N_p$

$f_p$  and  $f_q$  = the correct pixel values

$\text{div}_{pq}$  = a divergence of pixels  $p$  and  $q$

$\partial\Omega$  = a region, surrounding the missing area  $\Omega$  in the known areas

$f_q^*$  = a known value of pixel  $q$  in a region  $\partial\Omega$

Equations 6 form a classical, sparse, symmetric, positive-definite system. Because of the arbitrary shape of region  $\partial\Omega$ , the

iterative solvers ought to be attracted, for example, either Gauss-Seidel iteration. If a missing area  $\Omega$  contains pixels on the boundary of a frame,  $|N_p| < 4$ , then there are no boundary terms in the right hand side of equations 6, and equation 7 is obtained:

$$|N_p| f_p - \sum_{q \in N_p} f_q = \sum_{q \in N_p} \text{div}_{pq} \cdot \quad (7)$$

Note that the Poisson equation can be applied in the temporal domain, when the neighboring pixels are attracted from the adjacent frames.

The optimal seam-based algorithms suppose that a seam line placement should minimize the photometric differences between two images. The seam gradient becomes invisible, if the difference between two images on the seam line has zero value. The optimal seam-based algorithms consider the scene content in the overlapping region in a contradiction to the smoothing-based blending that permits to solve the problems like moving objects or parallax impact. Recently, different optimal seam finding methods have been developed. Gracías et al. (Gracías et al., 2009) proposed an automatic blending technique using the watershed segmentation that reduces the search space for finding the boundaries between the images and graph cut optimization with the guarantees of the globally optimal solution for each intersection region. The authors claimed the parallel implementation and memory-efficient technique even for large-scale mosaics.

The idea of multi-band blending was formulated by Burt and Adelson (Burt and Adelson, 1983). Let a weight function in Euclid space of an image be  $W(x, y) = w(x)w(y)$ , where functions  $w(x)$  and  $w(y)$  vary linearly between 0 at the edges and 1 at the image centre. The idea of a multi-band blending is to blend the low frequencies over a large spatial range and the high frequencies over a short range. The blending weights for each image are initialized by finding the set of points  $j$  for which image  $i$  using equation 8:

$$W_{\max}^i(x, y) = \begin{cases} 1 & \text{if } W^i(x, y) = \arg \max_j W^j(x, y) \\ 0 & \text{otherwise} \end{cases}, \quad (8)$$

Then the weight maps are blurred for each band in order to form the weights in each band. A high pass version of the image can be formed by equations 9:

$$\begin{aligned} B_{\sigma}^i(x, y) &= I^i(x, y) - I_{\sigma}^i(x, y) \\ I_{\sigma}^i(x, y) &= I^i(x, y) * g_{\sigma}(x, y) \end{aligned} \quad (9)$$

where  $B_{\sigma}^i(x, y)$  = the spatial frequencies in the range of wavelengths  $\lambda \in [0, \sigma]$

$I^i(x, y)$  = an image in current band

$I_{\sigma}^i(x, y)$  = an image convoluted with the Gaussian

$g_{\sigma}(x, y)$  = the Gaussian filter with a standard deviation  $\sigma$

For current band, the images must be convoluted with corresponding maximum weight functions:

$$W_{\sigma}^i(x, y) = W_{\max}^i(x, y) * g_{\sigma}(x, y) \quad (10)$$

where  $W_{\sigma}^i(x, y)$  = the blending weight for the wavelength from  $\lambda \in [0, \sigma]$

Therefore, each subsequent band  $k \geq 1$  is blended using the previous lower frequency band images and weights:

$$\begin{aligned} B_{(k+1)\sigma}^i(x, y) &= I_{k\sigma}^i(x, y) - I_{(k+1)\sigma}^i(x, y) \\ I_{(k+1)\sigma}^i(x, y) &= I_{k\sigma}^i(x, y) * g_{\sigma'}(x, y) \\ W_{(k+1)\sigma}^i(x, y) &= W_{k\sigma}^i(x, y) * g_{\sigma'}(x, y) \end{aligned} \quad (11)$$

where  $\sigma'$  = the Gaussian standard deviation of next band,  $\sigma' = \sqrt{2k+1}\sigma$

As a result, the subsequent bands have the same range of wavelengths and the final expression for each band is provided by equation 12

$$I_{k\sigma}^{\text{multi}}(x, y) = \frac{\sum_{i=1}^N B_{k\sigma}^i(x, y) W_{k\sigma}^i(x, y)}{\sum_{i=1}^N W_{k\sigma}^i(x, y)}, \quad (12)$$

where  $N$  = a number of sub-bands

The multi-band mosaic is obtained by summing images of all subsequent bands. This multi-band blending approach allows high frequency bands to be blended over short ranges and low frequency bands to be blended over long ranges. An example of multi-band mask is depicted in Figure 4.

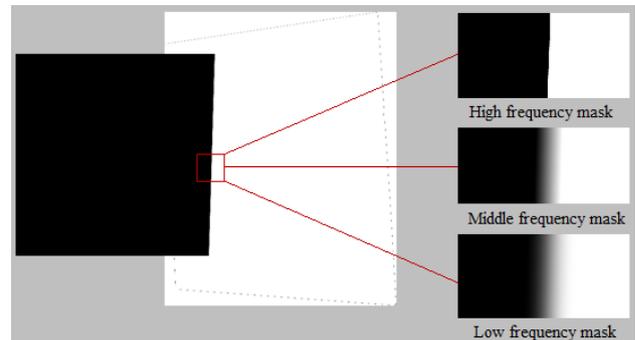


Figure 4. Multi-band masks for image blending

The steps of the local motion inpainting, local warping and seamless blending are executed for each motion layer if it is necessary. The proposed selective technique for video completion is reinforced by the look up table that stores the computed blending weights for the defined locations in a pseudo panoramic key frame.

## 5. EXPERIMENTAL RESULTS

The sampling of twelve video sequences was used in experiments. These video sequences contain one or many foreground objects in a non-planar scene and were obtained with an unstable camera movement. A description of some video sequences is performed in Table 1.



Figure 5. Results of processing of frame 20 from 00081 MTS - shaky original.avi, frame 20 from Gleicher1.avi, frame 15 from Gleicher2.avi, and frame 20 from Gleicher4.avi (from left to right): a original frames, b stabilized frames, c completion of missing regions using feathering-based blending, d completion of missing regions using pyramid-based blending, e completion of missing regions using gradient-based blending, f completion of missing regions using seamless multi-band blending

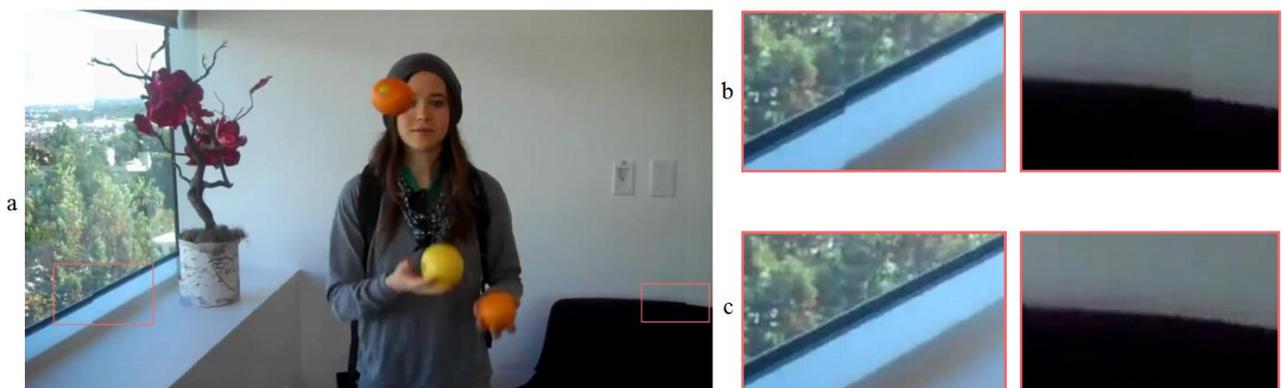


Figure 6. Results of seamless stitching in frame of a high resolution video sequence: a original frame, b fragments with artifacts, c fragments without artifacts obtained with seamless multi-band blending

Video	Size	Length	Description
00081 MTS - shaky original.avi <sup>1</sup>	pixels 640×360	frames 1230	Complex structure, moving objects
Gleicher1.avi <sup>2</sup>	640×360	480	Complex structure, unstable camera motion
Gleicher4.avi <sup>2</sup>	640×360	412	Slow camera motion, 3D scene structure
Sam_1.avi <sup>2</sup>	640×360	330	Fast motion of foreground object
New_gleicher.avi <sup>2</sup>	480×270	275	Non-planar scene, complex structure

Table 1. Short description of test sampling

The results of frames' processing are depicted in Figure 5. As it seems, the seamless multi-band blending provides better visual results even for video sequences with low resolution. The results of frame processing from our high resolution video sequence are depicted in Figure 6. The feathering-based blending causes noticeable visual artifacts that occur due to the different color settings and brightness of the original frames. Such disadvantages can be eliminated using of seamless multi-band blending combining the pixel intensity values on the boundaries of the successive frames.

Some problems appear, when the missing areas in the current frame are absent in the original frames. This happens, when a video sequence contains the fast motion or large changing in a scene. In these cases, the additional blurring of contours in the missing areas, small frame scaling, or texture reconstruction methods are recommended to apply.

## 6. CONCLUSIONS

Video completion can be considered as an extension of 2D frame completion to the spatiotemporal space. This is a necessary stage after video stabilization in order to hold the resolution of the reconstructed frame at the initial level. The algorithm was tested on several video sequences with different unwanted jitters. A special attention was paid to the seamless blending in the reconstructed area. The experiments show that an optimal seam-based blending provides the best visibility result for all types of scenes but for simple scenes the pyramid-based approach can be applied as a trade-off between visibility and computational cost.

## ACKNOWLEDGEMENTS

The reported study was funded by the Russian Fund for Basic Researches according to the research project № 16-07-00121 A.

## REFERENCES

Barron, J.L., Fleet, D.J., Beauchemin, S.S., 1994. Performance of optical flow techniques. *International Journal of Computer Vision*, 12(1), pp. 43-77.

Bertalmio, M., Sapiro, G., Caselles, V., Ballester, C., 2000. Image inpainting. In: *the 27th annual conference on Computer*

*graphics and interactive techniques, SIGGRAPH 2000*, New Orleans, LA, USA, pp. 417-424.

Bugeau, A., Bertalmio, M., 2009. Combining texture synthesis and diffusion for image inpainting. In: *The Fourth International Conference on Computer Vision Theory and Applications*, Porto, Portugal, pp. 26-33.

Burt, P.J., Adelson, E.H., 1983. A multiresolution spline with application to image mosaics. *ACM Transactions on Graphics*, 2(4), pp. 217-236.

Cai, J., Walker, R., 2009. Robust video stabilisation algorithm using feature point selection and delta optical flow. *IET Computer Vision*, 3(4), pp. 176-188.

Canny, J., 1986. A computational approach to edge detection. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, PAMI-8(6), pp. 679-698.

Chan, T.F., Yip, A.M., Park, F.E., 2005. Simultaneous total variation image inpainting and blind deconvolution. *International journal of imaging systems and technology*, 15(1), pp. 92-102.

Criminisi, A., Perez, P., Toyama, K., 2004. Region filling and object removal by exemplar-based image inpainting. *IEEE Transactions on Image Processing*, 13(9), pp. 1200-1212.

Fang, C.L., Tsai, T.H., 2008. Advertisement video completion using hierarchical model. In: *The IEEE Conference on Multimedia and Expo*, Hannover, Germany, pp. 1557-1560.

Favorskaya, M., Damov, M., Zotin, A., 2013. Accurate spatio-temporal reconstruction of missing data in dynamic scenes. *Pattern Recognition Letters*, 34(14), pp. 1694-1700.

Favorskaya, M., Buryachenko, V., 2015. Fuzzy-based digital video stabilization in static scenes. In: Tshirintzis GA, Virvou M, Jain LC, Howlett RJ, Watanabe T (Eds) *Intelligent Interactive Multimedia Systems and Services in Practice*, vol. 36 SIST, pp. 63-83.

Gracias, N., Mahoor, M., Negahdaripour, S., Gleason, A., 2009. Fast image blending using watersheds and graph cuts. *Journal Image and Vision Computing*, 27(5), pp. 597-607.

Jia, J., Wu, T.P., Tai, Y.W., Tang, C.K., 2004. Video repairing: inference of foreground and background under severe occlusion. In: *The IEEE Conference on Computer Vision and Pattern Recognition*, Washington, DC, USA, vol. 1, pp. 364-371.

Jia, J., Tai, Y.W., Wu, T.P., Tang, C.K., 2006. Video repairing under variable illumination using cyclic motions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(5), pp. 832-839.

Levin, A., Zomet, A., Peleg, S., Weiss, Y., 2004. Seamless image stitching in the gradient domain. In: *The 8th European Conference on Computer Vision*, Prague, Czech Republic, part IV, pp. 377-389.

Liang, L., Liu, C., Xu, Y.Q., Guo, B., Shum, H.Y., 2001. Real-time texture synthesis by patch-based sampling. *ACM Transactions on Graphics*, 20(3), pp. 127-150.

<sup>1</sup> <https://www.youtube.com/watch?v=NnyYXbk8MXU/>

<sup>2</sup> <http://www.cc.gatech.edu/cpl/projects/videostabilization/>

- Litvin, A., Konrad, J., Karl, W., 2003. Probabilistic video stabilization using kalman filtering and mosaicking. In: *The IS&T/SPIE Symposium on Electronic Imaging, Image, and Video Communications*, pp. 663-674.
- Liu, F., Gleicher, M., Jin, H., Agarwala, A., 2009. Content-preserving warps for 3D video stabilization. *ACM Transactions on Graphics*, 28(3), pp. 44.1-44.9.
- Liu, W.X., Chin, T.J., 2015. Smooth globally warp locally: video stabilization using homography fields. In: *The International Conference on Digital Image Computing: Techniques and Applications*, Gold Coast, Australia, pp 1-8.
- Matsushita, Y., Ofek, E., Ge, W., Tang, X., Shum, H.Y., 2006. Full-frame video stabilization with motion inpainting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(7), pp. 1150-1163.
- Mohamadabadi, B.B., Khaligh, A.B., Hassanpour, R., 2012. Digital video stabilization using Radon transform. In: *International Conference on Digital Image Computing Techniques and Applications*. Fremantle, Australia, pp. 1-8.
- Pandey, A., Pati, U.C., 2013. A novel technique for non-overlapping image mosaicing based on pyramid method. In: *The Annual IEEE India Conference*, Mumbai, India, pp. 1-6.
- Patwardhan, K.A., Sapir, G., Bertalmio, M., 2005. Video inpainting of occluding and occluded objects. In: *The IEEE Conference on Image Processing*, Genoa, Italy, vol. 2, pp. 69-72.
- Patwardhan, K.A., Sapiro, G., Bertalmio, M., 2007. Video inpainting under constrained camera motion. *IEEE Transactions on Image Processing*, 16(2), pp. 545-553.
- Peleg, S., Rousso, B., Rav-Acha, A., Zomet, A., 2000. Mosaicing on adaptive manifolds. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(10), pp. 1144-1154.
- Pérez, P., Gangnet, M., Blake, A., 2003. Poisson image editing. *ACM Transactions on Graphics*, 22(3), pp. 313-318.
- Sangeetha, K., Sengottuvelan, P., Balamurugan, E., 2011. Combined structure and texture image inpainting algorithm for natural scene image completion. *Journal of Information Engineering and Applications*, 1(1), pp. 7-12.
- Shih, T.K., Tang, N.C., Hwang, J.N., 2009. Exemplar-based video inpainting without ghost shadow artifacts by maintaining temporal continuity. *IEEE Transactions on Circuits and Systems for Video Technology*, 19(3), pp. 347-360.
- Shiratori, T., Matsushita, Y., Kang, S.B., Tang, X., 2006. Video completion by motion field transfer. In: *The IEEE Conference on Computer Vision and Pattern Recognition*, New York, NY, USA, vol. 1, pp. 411-418.
- Shukla, D., Ojha, A., Jha, R.K., 2017. A new composite multi-constrained differential-radon warping approach for digital video affine motion stabilization. *Computer Vision and Image Understanding*, 155, pp. 83-105.
- Szeliski, R., 2006. Image alignment and stitching: a tutorial. *Journal Foundations and Trends in Computer Graphics and Vision*, 2(1), pp. 1-104.
- Szeliski, R., Uyttendaele, M., Steedly, D., 2011. Fast Poisson blending using multi-splines. In: *The IEEE International Conference on Computational Photography*, Pittsburg, PA, USA, pp. 1-8.
- Veldandi, M., Ukil, S., Rao, K.G., 2013. Video stabilization by estimation of similarity transform from integral projection. In: *20th IEEE International Conference on Image Processing*. Melbourne, Australia, pp. 785-789.
- Wexler, Y., Shechtman, E., Irani, M., 2004. Space-time video completion. In: *The IEEE Conference on Computer Vision and Pattern Recognition*, Washington, DC, USA, vol. 1, pp. 120-127.
- Wexler, Y., Shechtman, E., Irani, M., 2007. Space-time completion of video. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(3), pp. 463-476.
- Xia, A., Gui, Y., Yao, L., Ma, L., Lin, X., 2011. Exemplar-based object removal in video using GMM. In: *The International Conference on Multimedia and Signal Processing*, Guilin, China vol. 1, pp. 366-370.
- Xiong, Y., Turkowski, K., 1998. Registration, calibration and blending in creating high quality panoramas. In: *The Fourth IEEE Workshop on Applications of Computer Vision*, Princeton, NJ, USA, pp. 69-74.
- Xiong, Y., 2009. Eliminating ghosting artifacts for panoramic images. In: *The 11th IEEE International Symposium on Multimedia*, San Diego, CA, USA, pp. 432-437.
- Yitzhaky, Y., Boshusha, G., Levy, Y., Kopeika, N.S., 2000. Restoration of an image degraded by vibrations using only a single frame. *Optical Engineering*, 39(8), pp. 2083-2091.