

# Chapter 4

## Panorama Construction from Multi-view Cameras in Outdoor Scenes

Lakhmi C. Jain, Margarita N. Favorskaya and Dmitry Novikov

**Abstract** The applications of panoramic images are wide spread in computer vision including navigation systems, object tracking, virtual environment creation, among others. In this chapter, the problems of multi-view shooting and the models of geometrical distortions are investigated under the panorama construction in the outdoor scenes. Our contribution are the development of procedure for selection of “good” frames from video sequences provided by several cameras, more accurate estimation of projective parameters in top, middle, and bottom regions in the overlapping area during frames stitching, and also the lighting improvement of the result panoramic image by a point-based blending in a stitching area. Most proposed algorithms have high computer cost because of mega-pixel sizes of initial frames. The reduction of frames sizes, the use of CUDA technique, or the hardware implementation will improve these results. The experiments show good visibility results with high stitching accuracy, if the initial frames were selected well.

**Keywords** Panorama construction · Image stitching · Projective transformation · Image selection · Robust detectors · Retinex algorithm · Texture blending

---

L.C. Jain (✉)  
University of Canberra, Canberra, ACT 2601, Australia  
e-mail: lakhmi.jain@unisa.edu.au

M.N. Favorskaya · D. Novikov  
Siberian State Aerospace University, Institute of Informatics and Telecommunications,  
31 Krasnoyarsky Rabochy, Krasnoyarsk 660014, Russian Federation  
e-mail: favorskaya@sibsau.ru

D. Novikov  
e-mail: novikov\_dms@sibsau.ru

## 4.1 Introduction

The panoramic representation of outdoor scenes is required in many applications involving computer vision systems in robotics [1, 2], industry, transport, surveillance systems [3], computer graphics, virtual reality systems [4], medical applications [5], etc. A digital panorama can be obtained by two ways: the use of a special panoramic camera or the analysis of many images received from a regular camera. On the one hand, the specified sensors, e. g. panoramic lens [6] or a fisheye [7] with a wide Field Of View (FOV) were applied at the first years of digital image processing. These sensors have high cost and provide images or videos with substantial distortions, which are not suitable for a user. On the other hand, a set of images can be provided by a single camera maintained on a tripod and rotated through its optical center, by using a single omni-directional camera, by multiple cameras pointing in different directions, or using a stereo panoramic camera. In the first case, the panoramic mapping and tracking is based on the assumption that only rotational movements of camera are available. This assumption allows the mapping of current frame onto a cylinder to create 2D panoramic image. A set of images or frames from video sequences ought to be composed to create a single panoramic image.

The process of panorama construction includes three main steps as mentioned below:

- The image acquisition, when a series of overlapping images are acquired [8].
- The images alignment, for which the parameters of geometrical transformations are required [9].
- The images stitching, when all aligned images are merged to create a composite panoramic image with the color correction [10].

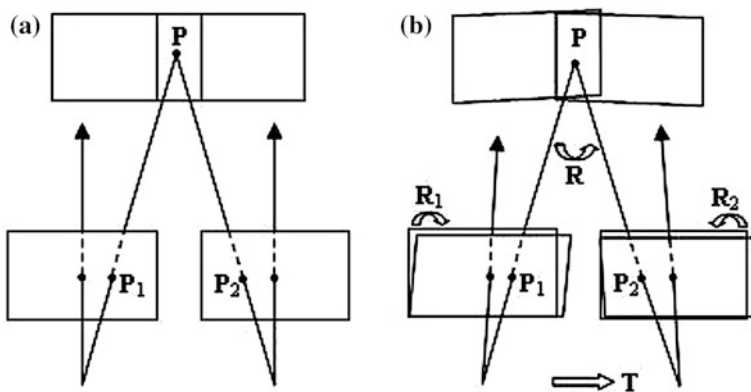
These three steps have different content according to mono-perspective and multi-perspective panoramic images. The mono-perspective shooting means that the images are produced with a tripod-mounted camera. A fixed focal point is situated in the center of projection. Only viewing direction is altered by camera rotations around vertical or vertical/horizontal axes. The multi-perspective shooting is taken from changing viewpoints and provides the patches, from which a panoramic image consists. This makes a seamless stitching procedure very difficult or even impossible because a scene changing according to various viewpoints cannot be aligned in a common case. Most of researches apply a conception of mosaics, which will be considered in Sect. 4.3. The smart approach connects with morphing or image metamorphosis application. This idea was proposed by Beier and Neely [11] and then was developed by Haenselmann et al. in the research [12]. The image in the middle of the metamorphosis is called the interpolated image. Then the color values are also interpolated and displayed in a panoramic image pixel by pixel. Such morphing requires a straightening of some lines in the panoramic image or reconstruction the textured non-visible regions. This issue is required in the following development to make a panoramic image more realistic.

Additionally the fourth step of panorama construction can be mentioned. This is a panorama improvement because the lighting and the color alignment are often required to compensate the shadows and the light-struck regions in separate images. Such step determines a final visibility of panoramic image.

The chapter is organized as follows. Section 4.2 explains a problem statement of multi-view shooting and the models of possible geometrical distortions. Related work is discussed in Sect. 4.3. Section 4.4 provides a procedure of the intelligent frames selection from multi-view video sequences. Section 4.5 derives the correspondence of reliable feature points for a seamless stitching in selected frames. The issues of panorama lighting improvement are located in Sect. 4.6. The experimental results and calculations for feature points detection and projective parameters are situated in Sect. 4.7. Section 4.8 summarizes the presented work and addresses the future development.

## 4.2 Problem Statement

Various types for receiving a set of initial images, discussing in Sect. 4.1, always have the geometrical or the geometrical/lighting distortions. The ideal case during the outdoor shooting is practically absent (Fig. 4.1a). A scheme of geometrical distortions appearing by a hand-held shooting is presented in Fig. 4.1b. The multi-cameras shooting, when cameras maintain on a moving platform, has similar distortions caused by vibrations and deviations under a movement in the real environment. If cameras are calibrated, then a panoramic image construction occurs faster and more accurate. However, the extended task—the use of non-calibrated cameras is more interesting for practice.



**Fig. 4.1** Scheme of a hand-held shooting with two points of shooting: **a** ideal vertical parallel image planes, **b** rotation and translation of image planes, which cause the geometrical distortions in images

One of the main tasks is to determine the parameters of such geometrical distortions. In literature, there are many models from the simple to the complicated ones including the perspective and the homographic transformations. If the possibility of stereo calibration exists, then a rotation matrix  $\mathbf{R}$  and a translation matrix  $\mathbf{T}$  are connected by Eq. (4.1), where  $\mathbf{P}_1$  and  $\mathbf{P}_2$  are points from two images in Euclidian coordinate system with coordinates  $(x_1, y_1)$  and  $(x_2, y_2)$ , respectively (Fig. 4.1).

$$\mathbf{P}_1 = \mathbf{R}^T(\mathbf{P}_2 - \mathbf{T}) \quad (4.1)$$

The mapping from coordinate  $\mathbf{p} = (x, y, f)$  to 2D cylindrical coordinates  $(\theta, h)$  are calculated by Eq. (4.2), where  $\theta$  is a panning angle,  $h$  is a scanning line,  $f$  is a focal length of a regular camera.

$$\theta = \arctan\left(\frac{x}{f}\right), \quad h = \frac{y}{\sqrt{x^2 + f^2}} \quad (4.2)$$

The cylindrical projection is easy to calculate, when the focal length is known and is not changed. However, this model does not consider the camera rotations around vertical axis. The spherical projection has enough degree of freedom and allows both vertical and horizontal rotations. An artifact of both approaches is a disturbing fish eye-like appearance in the synthesized image.

Usually an eight-parameter planar projection is used under the assumption that a camera has small rotations. The eight parameters are characterized a rotation within an image plane, a perspective turn in vertical and horizontal directions, a scaling factor, and the horizontal and vertical translations. For two points with coordinates  $(x_i, y_i)$  and  $(x_j, y_j)$  in two images, a perspective transformation has a view expressed by Eq. (4.3), where  $\mathbf{H}$  is a  $3 \times 3$  invertible non-singular homography matrix (homographies and points are defined as a non-zero scalar).

$$\begin{bmatrix} x_j \\ y_j \\ 1 \end{bmatrix} = \begin{bmatrix} h_1 & h_2 & h_3 \\ h_4 & h_5 & h_6 \\ h_7 & h_8 & h_9 \end{bmatrix} \begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} \quad (4.3)$$

In the case of projective transformation, the above Eq. (4.3) can be written as Eq. (4.4),  $h_9 = 1$ .

$$x_j = \frac{h_1x_i + h_2y_i + h_3}{h_7x_i + h_8y_i + 1}, \quad y_j = \frac{h_4x_i + h_5y_i + h_6}{h_7x_i + h_8y_i + 1} \quad (4.4)$$

A pair of corresponding points provides two following equations (Eq. 4.5).

$$\begin{aligned} h_1x_i + h_2y_i + h_3 - h_7x_ix_j - h_8x_jy_i &= x_j \\ h_4x_i + h_5y_i + h_6 - h_7x_iy_j - h_8y_iy_j &= y_j \end{aligned} \quad (4.5)$$

To compute eight parameters of  $\mathbf{H}$  matrix (excluding  $h_9$ ), the four-point correspondences are required. The contribution of this chapter includes the detection of accurate feature points correspondences, depending from an image content. Also the issue of lighting alignment after images stitching will be considered in details.

### 4.3 Related Work

Two main approaches to create the panoramic images are known. The first one is based on the patches [13–15] and the second one uses the feature points [1, 16]. However, the goal of these approaches is to find the “best” patches or feature points and their the “best” correspondence in the overlapping area of two neighbor images. Let us consider the related works according to both approaches.

The creation of an image mosaic is a popular method to increase a field of camera view, combining the several views of scene in the panoramic image. Such algorithms direct on the normalization of projective transformations between images called the homography. However, the creation of high quality mosaic requires the color and lighting corrections to avoid seams caused by the moving objects and the photometric variations across the blending boundaries. Kim and Hong formulated the blending problem as a labeling problem of Markov Random Fields (MRF) [17]. The authors used the image patches centered at the regular spatial grid. In each grid, a collection of image patches from the registered images are generated.

To find a smooth and consistent with the collected patches, Kim and Hong proposed to define an energy function  $E(f)$  using the MRF formulation given by Eq. (4.6), where  $N$  is a set of pairs of neighboring grids,  $P$  is a grid space representing a pixel location in the image plane,  $D_p(f_p)$  is a data penalty function in grid  $p$  (patch location) with label (image number)  $f_p$ , and  $V_{p,q}(f_p, f_q)$  is a pairwise smoothness constraint function of two neighboring grids  $p$  and  $q$  with labels  $f_p$  and  $f_q$ , respectively.

$$E(f) = \sum_{p \in P} D_p(f_p) + \sum_{pq \in N} V_{p,q}(f_p, f_q) \quad (4.6)$$

The energy  $E(f)$  can be minimized by using the graph-cuts algorithm. Kim and Hong used so called  $\alpha$ -expansion graph-cuts algorithm, which minimizes an energy function by the cyclically iterations through every possible  $\alpha$  label. During an

iteration, the algorithm determines a possibility of improvement for the current labeling by changing of some grids. The penalty function  $D_p(f_p)$  is determined by Eq. (4.7), where  $C$  is a constant,  $B_{\Omega}(i, j)$  represents a similarity measure (an average value) between two images  $I_i$  and  $I_j$  in the patch region  $\Omega$ .

$$D_p(f_p) = \sum_{k \neq f_p} \min(B_{\Omega_p}(k, f_p), C) \quad (4.7)$$

The smoothness constraint  $V_{p,q}(f_p, f_q)$  in the overlapping region  $\Omega_{pq}$  between neighboring patches is defined by empirical dependence expressed by Eq. (4.8), where  $S_{p,q}(f_p, f_q)$  is a general form of a smoothness cost (Eq. 4.9),  $R_{p,q}(f_p, f_q)$  encourages the labels to consist of a small number of uniform regions reducing the noise-like patterns and simplifies the shape of boundaries (Eq. 4.10),  $\lambda_s$  is a smoothness coefficient,  $\lambda_r$  is used to encourage the labels with a small number of the uniform regions. The recommendations of Kim and Hong are  $\lambda_s = 1$  and  $\lambda_r = 10$ .

$$V_{p,q}(f_p, f_q) = \lambda_s S_{p,q}(f_p, f_q) + \lambda_r R_{p,q}(f_p, f_q) \quad (4.8)$$

$$S_{p,q}(f_p, f_q) = B_{\Omega_{p,q}}(f_p, f_q) \quad (4.9)$$

$$R_{p,q}(f_p, f_q) = \begin{cases} 1 & f_p \neq f_q \\ 0 & o.w. \end{cases} \quad (4.10)$$

The MRF approach was applied to analyze the neighboring patches with moving objects. Also a simple exposure algorithm was proposed to reduce seams near the estimated boundaries by the correction of intensities.

Also a spectral analysis can be applied for patches detection and correspondence. The robust panoramic image mosaics in the complex wavelet domain were presented by Bao and Xu [18]. The panoramic mosaics are built by a set of alignment transformations between the image pixels and the viewing direction. The full planar projective panorama and the cylindrical panoramic image, which is warped in the cylindrical coordinates (assuming that a focal length is known and a camera has a horizontal position), are considered in this research. The multispectral mosaicing for enhancement of spectral information from the thermal infra-red band and visible band images, directly fused at the pixel level, was suggested by Bhosle et al. [19]. The authors developed a geometric relationship between the visible band of panoramic mosaic and the infrared one. A creation of panoramic sequences with video textures had been considered by Agarwala et al. [20] based on the smoothness constraints between patches with a low computational cost.

Let us consider the second approach, connecting with extraction of feature points. The matching of the corresponding points between two and more images is used successfully in many computer vision tasks particularly in panorama

construction [21]. The detection and description of feature points are the connected procedures. The distinctiveness and the robustness to photometric and geometrical transformations are two main criteria for feature point extraction [22]. Deng et al. considered the using of the panorama views instead of standard projection views in 3D mountain navigation system [23]. This approach is based on the features of interest. The authors specify the reference points located between the features of interest, trace them along the line of sight, and then determine the visibility of these features in 3D mountain scene. This permits to avoid occlusions of close objects by using a perspective projection for rendering of 3D spatial objects.

A modified Hough transform for the feature detection in panoramic images was proposed by Fiala and Basu [24]. The authors modified the well known Hough transform in such manner that only horizontal and vertical line segments were detected by the edge pixels mapping in a new 2D parameter space. The recognition of horizontal and vertical lines is very useful for mobile robots navigation in urban environment, where the majority of line edge features are either horizontal or vertical.

An efficient method to create the panoramic image mosaics with multiple images was proposed by Kim et al. [25]. This method calculates the parameters of a projective transformation in the overlapped area of two given images by using four seed points. The term “seed point” means the highly textured point in the overlapped area of the reference image, which is extracted by using a phase correlation. The authors assert that because a Region Of Interest (ROI) is restricted in the overlapped areas of two images, more accurate correspondences can be obtained. The proposed algorithm includes the following steps:

- The input a pair of images.
- The extraction of overlapped areas by using a phase correlation method.
- The histogram equalization and the selection of seed points.
- The detection of seed points' correspondence by using the weighted Block Matching Algorithm (BMA).
- The parameters calculation of projective transformation.
- The estimation of focal length.
- The mapping of image coordinates in the cylindrical coordinates.
- The calculation of displacement between two images.
- The blending of warped images using a bilinear weight function.

The parameters of a projective transformation are determined by using four seed points, which ought to exist in the overlapped area and not more than three of them should be collinear. The overlapped area is divided into four sub-areas, where the central pixel of the  $k$ th block with maximum variance is selected as the seed point  $q_i$  into the  $i$ th sub-area. Equation (4.11) provides a seed point  $q_i$  calculation, where  $\sigma_{k,i}^2$  is a variance,  $M_{k,i}$  is a mean value of the  $k$ th block into the  $i$ th sub-area,  $h_g$  is a histogram of a gray level  $g$  values, and  $G_{max}$  is a maximum of gray level value ( $G_{max} = 255$ ).

$$q_i = \operatorname{agr} \max_k \left| \sigma_{k,i}^2 \right|, \quad 0 \leq i \leq 3$$

$$\sigma_{k,i}^2 = \sum_g^{G_{\max}} g^2 \cdot h_g - M_{k,i}^2 \quad (4.11)$$

Such approach, representing by Kim et al. is enough simple. However, if the overlapped areas are the high textured regions, then Eq. (4.11) can provide several seed points with equal values of variance of the  $k$ th block into the  $i$ th sub-area, and the choice of single point will be non-determined. Also as any statistical model, the proposed decision is a noise dependent.

More common approach was proposed by Zhu et al. as a 2D/3D realistic panoramic representation from video sequence with dynamic and multi-resolution capacities [26]. The authors investigated the moving objects in the static scenes, which are taken by a hand-held camera undergoing 3D rotation (panning), zooming, and small translation. In this case, the motion parallax can be a non-zero value, but it is neglected due to small translation. The authors assumed that the camera covers the full  $360^\circ$  field of view around the camera, and the rotation is almost around its nodal point. The algorithm includes three steps such as the interframe motion estimation, the motion accumulation and classification, and a Dynamic and Multi-Resolution (DMR)  $360^\circ$  panoramic model generation.

The camera motion has six degrees of freedom: three translation and three rotation components. A coordinate system XYZ is attached to the moving camera with the origin O in the optical center of camera. An alternative interpretation is a scene movement with six degree of freedom. The authors represented three rotation angles (roll, tilt, and pan) through the interframe difference by  $(\alpha, \beta, \gamma)$  into a rotation matrix  $\mathbf{R}$  and three translation components  $\mathbf{T} = (T_x, T_y, T_z)^T$ . 3D point  $\mathbf{X} = (x, y, z)^T$  with the image coordinates  $\mathbf{u} = (u, v, 1)^T$  (UOV is a plane of image) at time instant  $t$  in the current frame moved from point  $\mathbf{X}' = (x', y', z')^T$  at the time instant  $t$  in the reference frame with the image coordinates  $\mathbf{u}' = (u', v', 1)^T$ . The relation between the 3D coordinates is expressed by Eq. (4.12).

$$\mathbf{X}' = \mathbf{R}\mathbf{X} + \mathbf{T} \quad (4.12)$$

These authors [26] used the simplified 2D rigid interframe motion model provided by Eq. (4.13), where  $s \approx ff'/f$  is a scale factor associated with zoom and Z-translation,  $f'$  and  $f$  are the camera focal lengths before and after the motion,  $(t_u, t_v) \approx (-\gamma_f, \beta_f)$  is a translation vector representing pan/X-translation and tilt/Y-translation, and  $\alpha$  is a roll angle.

$$\begin{cases} s \cdot u' = u + \alpha v + t_u \\ s \cdot v' = v - \alpha u + t_v \end{cases} \quad (4.13)$$



The motion model, representing by Eq. (4.13), is actual in far away scenes. The least square solution of motion parameters  $s$ ,  $t_u$ ,  $t_v$ , and  $\alpha$  can be obtained by given more than two pairs of corresponding points between two frames. The errors of approximation can be corrected by the mosaic algorithm [27, 28]. The color frames had the size  $384 \times 288$  pixels that permitted to process five frames per second.

Langlotz et al. proposed to use the Features from Accelerated Segment Test (FAST) keypoints for the mapping and tracking purposes [29]. The FAST keypoints are extracted at each frame to create the panoramic map. The authors try to extend the dynamic range of images by the choice of the anchor camera frame and the using of pixel intensities as a baseline for all further mappings. Then the differences of intensities are computed for all pairs of matching keypoints in the original frame and the panoramic map. The average difference of these point pairs is used to improve the current frame before its inclusion in the panorama image.

Sometimes the panoramic view morphing is used to generate a set of new images from different points of view, if two basic views of static scene uniquely determine a set of views on the line between the optical centers of cameras, and a visibility constraint is satisfied. Seitz and Dyer proved this statement in their research [30]. Often the panoramic images are required in medical applications for improvement the ultrasound images, in ophthalmology, etc. A fast and automatic mosaicing algorithm to construct the panoramic images for cystoscopic exploration was presented by Hernández-Mier et al. [31]. The algorithm provides the panoramic images without affecting the application protocol of cystoscopic examination.

Method of optimal stitching based on tensor analysis was proposed by Zhao et al. in the research [10]. This method has the advantage of causing less artifacts in the final panorama despite the presence of complex radiometric distortions such as vignetting by a new function of seam cost. An approach for image stitching based on Hu moments and Scale-Invariant Feature Transform (SIFT) features is presented by Fathima et al. [32]. The authors propose to determine the overlapping region of images using the combined gradient and invariant moments to reduce the processing area of features extraction. The selection of matching regions is realized by gradient operators based on the edge detection. After definition of the dominant edges, the images are partitioned in equal sized blocks and compared with each other to find the similarity. Then the SIFT features are extracted, and their correspondence are defined by RANDOM SAMPLE Consensus (RANSAC) algorithm [33]. The authors assert that their approach reduces 83 % of unreliable features.

Yan et al. combined the probability models of appearance similarity and keypoint correspondences in a maximum likelihood framework, which is named as Homography Estimation based on Appearance Similarity and Keypoint correspondences (HEASK) [34]. The probability model of the keypoint correspondences is based on the mixture of Laplacian distributions and the uniform distribution, which is proposed by the authors. Also the authors built a probability model of the appearance similarity. The authors named the method with only the term of a keypoint correspondence fitness as the HEASK-I, and named the method with only the term of an image appearance similarity as the HEASK-II. The experimental

results show that the HEASK achieves a good trade-off between the keypoint correspondences and the image similarity in comparison of the RANSAC, Maximum Likelihood Estimator SAC (MLE-SAC) [35], and the Logarithmical RANSAC (Lo-RANSAC) [36].

The analysis of related works shows that some complex issues of panorama creation especially through the multi-view cameras shooting require the additional research and development. This helps to achieve the reliable results in the construction of panoramic image acceptable for user surveillance in various tasks of computer vision.

## 4.4 Intelligent Selection and Overlapping of Representative Frames

Let several video sequences be entered from the multi-view cameras, which are maintained on a moving platform or vehicle. All six parameters including three angles and three coordinates in a camera 3D coordinate system relatively a platform 3D coordinate system are known. The main requirement connects with the overlapping regions between images, approximately 10 % of image area. The conditions of real shooting such as vibrations, deviations of lens systems, lighting change, etc. damage the ideal calculation.

The primary efforts are directed on the background analysis of image. This stage involves many methods and processing algorithms and has a high computational cost. The selection of representative frames provides the “good” frames from video sequences without artifacts. This procedure is represented in Sect. 4.4.1. An overlapping analysis of selected frames is discussed in Sect. 4.4.2.

### 4.4.1 Selection of Representative Frames

Let a moment of multi-cameras shooting is determined but with the ms error because cameras can be the non-synchronized, have different constructive types, and various access time for writing a video sequence in the inner device. Also a process of panorama creation has a large duration, and the sampling of forth-sixth panoramic images per s is a good result in a current experimental stage. The main acquisitions of representative frames selection are mentioned below:

- The selected frame is a sharp fame.
- The selected frame ought to have an equal median lighting.
- The selected frame ought to be contrast and has a color depth.
- An overlapping area between two frames from neighbor cameras ought to have a maximum value.

- It is desired that the vertical and horizontal lines in a frame satisfy to affine or perspective transformations.

For these purposes, a set of well known filters such as Laplacian [37], high dynamic range [38], and morphological [39] filters, also a Hough transformation to detect the vertical and horizontal lines [40] can be applied automatically in a parallel mode. The received estimations have a different contribution in the final decision. For example, a blurred frame ought to be rejected immediately. A frame with the non well-defined lines would like to be changed. The parameters of lighting, contrast, and color can be improvement during the following processing. The overlapping area strongly depends from the cameras positions and can be tuned preliminary. It is required to track the low threshold value of overlapping area. In spite of some computational cost of such preprocessing filtering, this step determines the accuracy of final panoramic result.

#### 4.4.2 *Overlapping Analysis of Selected Frames*

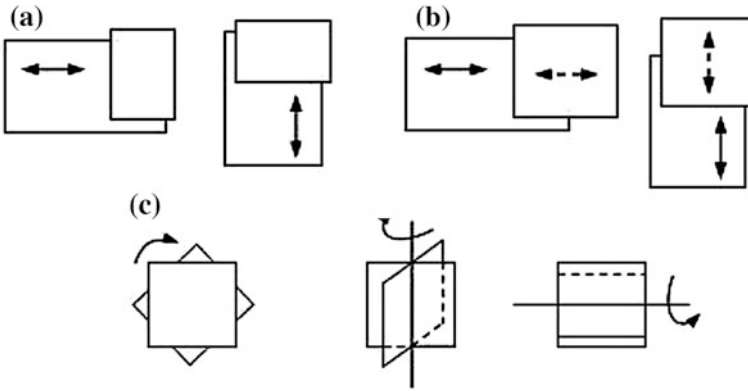
For small cameras rotations, the eight-planar perspective transformation is the most appropriate approach. If the translation is a non-linear, then the perspective transformation (Eq. 4.3) by homogeneous coordinates introduced by Maxwell [41, 42] and later applied to computer graphics by Roberts [43] is often used. Equation (4.14) provides such homogeneous transformation, where  $w$  is a warping parameter.

$$\begin{bmatrix} x_j \\ y_j \\ 1 \end{bmatrix} = \begin{bmatrix} h_1 & h_2 & h_3 \\ h_4 & h_5 & h_6 \\ h_7 & h_8 & 1 \end{bmatrix} \begin{bmatrix} x_i \\ y_i \\ w \end{bmatrix} \quad (4.14)$$

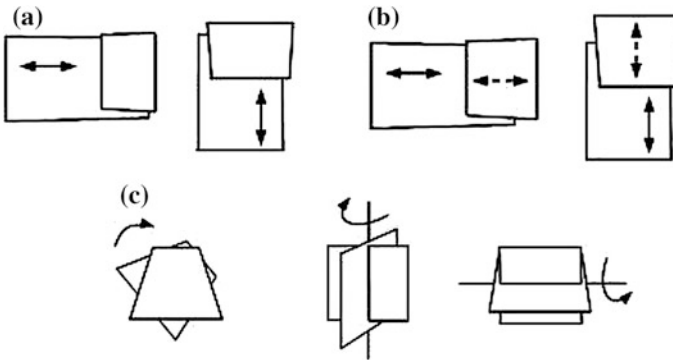
One can see the examples of frames overlapping in the case of the affine model (Fig. 4.2) and the perspective model (Fig. 4.3).

The homogeneous transformation is more complex in comparison of other procedures. The complexity of such transformations causes the necessity to find a way of frames matching with a high accuracy. The application of classical pattern recognition methods is non-useful here because of the unknown distortions of objects views. In this case, patches or feature points matching are recommended. In recent researches, the feature points correspondence is often used for the seamless stitching of selected frames. These issues will be discussed in Sect. 4.5.

The overlapping under the projective and the homogeneous transformations leads to the parallax effects. Let us notice that the affine transformation is the simplest case, which is rarely met in the panorama construction, and has not significant problems during frames stitching. Some examples of parallax effect are

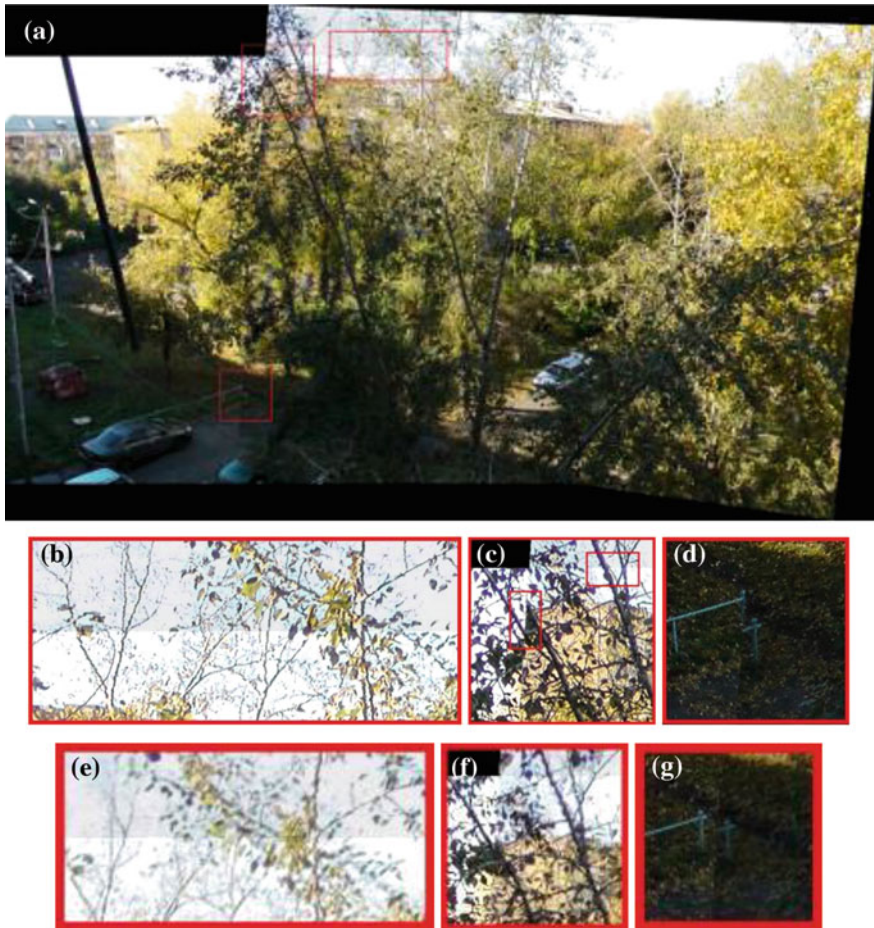


**Fig. 4.2** Variants of frames overlapping under the affine transformation: **a** horizontal and vertical translations, **b** horizontal and vertical scaling and translations, **c** rotations relatively three coordinate axes



**Fig. 4.3** Variants of frames overlapping under the projective transformation: **a** horizontal and vertical translations, **b** horizontal and vertical scaling and translations, **c** rotations relatively three coordinate axes

situated in Fig. 4.4. The experiments show the interesting property: if the initial frames will be reduced, for example, in two or four times, then some artifacts disappear (Figs. 4.4e, f, g). It may be explain that the scaled down panoramic images have more “rough” structure of feature points, and their number is significantly reduced in comparison with the initial image. However, the quality of such fragments has been lost. Let us discuss the problems and possible decisions for seamless stitching of the selected frames.



**Fig. 4.4** A panoramic creation with parallax effects, the item “Trees from Window”: **a** stitching panorama, **b** top good stitching (without lighting alignment), **c** top artifact, **d** bottom artifact, **e** top good stitching in reduced image, **f** top good stitching in reduced image, **g** bottom artifact in reduced image

### 4.5 Stitching of Selected Frames

A stitching of frames is the main core in the panorama construction, which includes the procedures of similar regions detection, similar regions matching, and some additional procedures determined by the content of selected frames. In Sect. 4.3, two main approaches were considered to determine the similar regions. More rough approach connects with so called “patches”, which are usually regions from  $3 \times 3$  pixels to  $11 \times 11$  pixels. More accurate approach is based on the technique of feature points detection. Great variety of feature detectors determines various

relations accuracy/computational cost. In this research, the feature points approach is developed. To accelerate a computational time, a pre-filtering was used before standard RANSAC application. In recent years, RANSAC or its modifications remain the basic algorithm to estimate the feature points correspondence.

The additional procedures are executed in the particular cases, for example, when the object with a high speed is situated in all selected frames, and it is required to understand its location in the panoramic image. Sometimes fragments in the selected frames have such big warping that the interpolation or the morphing procedures provide the better panoramic image. The feature points detection, matching, and correspondence will be discussed in Sects. 4.5.1–4.5.3, respectively. Section 4.5.4 provides the image projection and the geometrical improvement of panorama. A visualization of high speed objects in panoramic images is briefly considered in Sect. 4.5.5.

### 4.5.1 Feature Points Detection

To create an operator, which combines the feature points detection and the lighting alignment in stitching images simultaneously, is possibly in the future. At present, they are two original procedures. The analysis of the existing feature points detectors in panorama construction show that Scale-Invariant Feature Transform (SIFT), Speeded-Up Robust Feature (SURF), and corner detectors are the most popular among others. Our recommendations are to use the SURF detector (because it is faster than the SIFT detector while both detectors demonstrate the close results) in the landscape outdoor scenes and the corner detectors jointly with a Hough transform (to find lines and correct a parallax effect) in the indoor scenes.

The SURF is a fast modification of the SIFT, which detects feature points by using the Hessian matrix. The determinant of the Hessian matrix achieves the extremum in the point of maximum changing of intensity gradient. The SURF detects well spots, corners, and edges of lines. The Hessian is invariant to rotation and intensity, but not invariant to scale. The SURF uses a set of scalable filters for the Hessian matrix calculation. The gradient in a point is computed by the Haar filters. The feature points detection is based on a calculation for determinant of the Hessian matrix  $\det(\mathbf{H})$  by Eq. (4.15), where  $L_{xx}(P, \sigma)$ ,  $L_{xy}(P, \sigma)$ ,  $L_{yx}(P, \sigma)$ , and  $L_{yy}(P, \sigma)$  are convolutions the second derivative of Gaussian  $G(P)$  with a function describing a frame  $I_p$  in a point  $P$  along OX axis, diagonal in the first quadrant, OY axis, and diagonal in the second quadrant, respectively,  $\sigma$  is a mean square. Equation (4.16) provides an example of the convolution along OX axis.

$$\det(\mathbf{H}(P, \sigma)) = L_{xx}(P, \sigma) \cdot L_{yy}(P, \sigma) - (L_{xy}(P, \sigma))^2 \quad (4.15)$$

$$L_{xx}(P, \sigma) = \frac{\partial^2}{\partial x^2} G(P) * I_P \quad (4.16)$$

In practice, the SUFR uses a binary approximation of Gaussian Laplacian, which is called Fast Hessian, and Eq. (4.15) is replaced by Eq. (4.17), where coefficient 0.9 means an approximate character of calculations.

$$\det(\mathbf{H}(P, \sigma)) = L_{xx}(P, \sigma) \cdot L_{yy}(P, \sigma) - 0.9 \cdot (L_{xy}(P, \sigma))^2 \quad (4.17)$$

The invariance to scale is provided by partitioning a set of scales (9, 15, 21, 27, etc.) in the octaves. Each octave includes four filters with different scales: the first octave involves (9, 15, 21, 27) scales, the second octave involves (15, 27, 39, 51) scales, the third octave involves (27, 51, 75, 99) scales, and etc. The number of octaves is usually equal 5–6. Theoretically, it is enough to cover scales from 1 to 10 in an image with sizes  $1,024 \times 768$  pixels. The octave filters are not calculated for all pixels. The first octave uses the each secondary pixel, the second octave uses the each fourth pixel, the third octave uses the each eighth pixel, and etc. The true extremum cannot be agreed with the calculated extremum. For this purpose, a search is initiated (an interpolation by a quadratic function) until a derivative will not be much close to 0.

In such manner, a list of detected feature points is created. The following step is the calculation of orientations. The Haar filters permit to indicate and normalize a vector of orientation. A noise gives the additional gradients in directions, which are different from a direction of the main gradient. Thus, the additional filtrating is required.

After feature points detection, the SUFR creates its descriptors as a vector with 64 elements or with extended 128 elements. Such vectors characterize the gradient fluctuations in surrounding of feature point. For a descriptor creation, a square area  $20 \cdot s$ , where  $s$  is a scale value, is built. The first octave uses the area  $40 \times 40$  pixels around a feature point. This square orients along a priority vector direction and is divided into 16 sub-squares with sizes  $5 \times 5$  pixels. In each sub-square, a gradient is determined by using of the Haar filters. Four components in each sub-square have a view provided by Eq. (4.18), where  $\Sigma dX$  and  $\Sigma dY$  are the sum of gradients,  $\Sigma |dX|$  and  $\Sigma |dY|$  are the sum of gradients modules.

$$\sum dX, \quad \sum |dX|, \quad \sum dY, \quad \sum |dY| \quad (4.18)$$

Four components from Eq. (4.18), which are multiplied on 16 sub-squares, provide a 64-value descriptor in surrounding of feature point. Additionally, the received values are weighed by the Gauss filter with  $\sigma = 3.3 \cdot s$ . This is required for reliability of the descriptor to noises in the remote areas from a feature point.



### 4.5.2 Feature Points Matching

The feature points are detected in the whole frame because a prediction of overlapping area position is impossible, when many frames are stitched. A feature points matching is a step, when only “good” feature points correspondence are remained. The task of matching is realized as a multi-search of nearest neighbor for a selected current point in a feature space. Under a term “nearest neighbor”, such point is understood, which locates in a minimal distance to the selected current point. One can use Euclid metric or some others, however it is the difficult task to compare the 64-valued vectors. The good decision is a representation of all detected descriptors as the  $k$ -dimensional tree ( $k$ - $d$  tree). The  $k$ - $d$  tree is such data structure, which divides  $k$ -dimensional space to order the points in this space. The  $k$ - $d$  trees are a modification of binary trees, and usually used in a multi-dimensional space. The following operators are applied to the  $k$ - $d$  trees:

- The  $k$ - $d$  tree creation.
- The adding of element.
- The removal of element.
- The tree balancing.
- The search of nearest neighbor by using a key value.

For elements matching, it is enough only two operators: the  $k$ - $d$  tree creation and the search of a nearest neighbor for a couple of input frames. In this research, the following node structure was applied. Each node includes a single point in feature space called a descriptor. Additionally, the following data are stored:

- The index of frame, from which the descriptor was received.
- The index of separating hyperplane.
- The left sub-tree.
- The right sub-tree.

The index of separating hyperplane is calculated during a tree creation. The left and right sub-trees include the necessary references for the data structure. To detect the position of each element, the index of separating hyperplane is used. The simple procedure sorts the input array of descriptors in the increased or decreased manner relative the element having the index of separating hyperplane points. Then the input array of descriptors is divided into two sets—the left and the right. All elements of the left set have a component value, for which the index of separating hyperplane points, less than the element has, which is written in a current node of a tree, and vice versa for the right set (the nonstrict inequalities are used). The hyperplane is described by Eq. (4.19), where  $i$  is an index of separating hyperplane,  $a_i$  is a value of the  $i$ th component of descriptor, which is written in a current node,  $X_i$  is the  $i$ th component of basis.



$$X_i = a_i \quad (4.19)$$

The last step of algorithm is a recursive call of procedure to create the  $k$ - $d$  tree for left and right sub-sets, and a loading of descriptors values as the parameters of left and right sub-trees.

The  $k$ - $d$  trees have a possibility of fast search of nearest neighbor without considering all remaining points. A mean time of one nearest neighbor is estimated as  $O(\log n)$ , where  $n$  is a number of tree elements. A mean time of feature points matching can be estimated as  $k \cdot n \cdot O(\log n)$ , where  $k$  is a number of neighbors (against  $k \cdot n^2$  for a full search). The search by the  $k$ - $d$  trees is the directional search because a current node is chosen under the determined rule. The  $i$ th component of a target vector is compared the  $i$ th component in a current node, where  $i$  is an index of separating hyperplane. If the  $i$ th component of target vector less than the  $i$ th component in a current node, then a search is continue in a left sub-tree, otherwise in a right sub-tree.

If a nearest neighbor was not detected in a close sub-tree, then the search is continued in a far sub-tree. The goal of nearest neighbor algorithm is to find the nearest value, which cannot be the exact value. This specialty is very important for the SURF descriptor. In spite the SURF descriptor determines a small area, which is invariant to many transformations, including a noise up to 30 %, the exact matching between regions in various images is unlikely.

After feature points matching, the procedure of features points estimation is started. Usually the RANSAC or its modifications are used for these purposes.

### 4.5.3 Feature Points Correspondence

The quality of feature points matching is checked by the RANSAC algorithm. The concept of this algorithm is based on separation of initial data on outliers (noises, failure points, and random data) and inliers (points, which satisfy a model). The iterations of the RANSAC are divided logically in two stages. The first stage is a choice of points and a model creation, and includes the following steps:

- Step 1. The choice of  $n$  points randomly from a set of initial points  $\mathbf{X}$  or based on some criterion.
- Step 2. The calculation of  $\theta$  parameters of model using a function  $M$ . Such model is called a hypothesis.

The second stage is a hypothesis verification:

- Step 1. The point correspondence for the given hypothesis is checked by using estimator  $E$  and threshold  $T$ .
- Step 2. The points are marked as inliers or outliers.

- Step 3. After bypass of all points, the algorithm determines a quality of current hypothesis. If a hypothesis is the best, then it replaces the previous best hypothesis.

As a result, the last hypothesis is remained as the best hypothesis. One can read about the RANSAC algorithm in details in the researches [44–49]. As any stochastic algorithm, the RANSAC has some disadvantages. The upper boundary of computing is absent. Therefore, a suitable result cannot be achieved for the acceptable time. Also the RANSAC can determine only a single model for the initial data set with the acceptable probability. For panoramic images, the first disadvantage can be compensated by CUDA technology application or hardware decisions. The second disadvantage is removed by an algorithmic improvement for the projective parameters of panorama image (Sect. 4.5.4).

The research of Yan et al. [34] is the RANSAC development. The authors combined the probability models of appearance similarity and keypoint correspondences in a Maximum Likelihood framework. Such novel estimator is called as Homography Estimation based on Appearance Similarity and Keypoint correspondences (HEASK). The novelties of the HEASK are connected with a distribution of inlier location error, which is represented by a Laplacian distribution, and with the similarity between the reference and transformed image by the Enhanced Correlation Coefficient (ECC) feature. Yan et al. realized their algorithm in a RANSAC-based framework and consistently achieved an accurate homography estimation under different transformation degrees and different inlier ratios.

#### ***4.5.4 Image Projection and Geometrical Improvement of Panorama***

The RANSAC algorithm provides many feature points correspondences. In the task of panorama construction, it is important to determine the parameters of affine/projective/homography transformations for following recalculation image coordinates. For projective transformation, four corresponding feature points are required. They provide eight coefficients  $h_1 \dots h_8$  from Eq. (4.3). The full homography transformation was not investigated in this research.

The proposed algorithm for selection of corresponding feature points includes two scenarios. The first one selects points in the whole overlapping image area randomly. The second scenario divides the overlapping area in three regions—top, middle, and bottom. The algorithm calculates coefficients of transformation in each region separately. The last approach is required for images with different models of distortions. This is the case of image warping, non-well investigated issue in the theory of computer vision. The joint for different models of distortions in the result panoramic image may be provided by interpolation or warping.

The image projection consists in a multiplication of the coordinate vector for each pixel on the coefficients of transformation matrix. The new pixel coordinates are pointed in the coordinate system connecting with the second image (on which the first image is projected). During such pixels projection, the following problems appear:

- The new pixel coordinates have the fractional values.
- The projective image losses a rectangle shape. The regions appear, in which initial points overlap each other or where the new points appear without any color information.

Therefore, an interpolation task is required. The traditional interpolation methods such as bilinear, bicubic, or spline interpolation cannot be recommended because of high computational cost [50, 51]. In this research, a type of linear interpolation based on three predetermined points was realized as a simple and fast decision.

The affine transformation provides a non-equal uniform scale distribution, when a cell of image grid remains a rectangle. Under the projective and homography transformations, a scale distribution is becoming the non-equal and the non-uniform. Therefore, a cell of image grid transforms to a non-regular rectangle. The method of linear interpolation by using three values (vertexes of triangle) is fast algorithm. First, it is required to build such triangles. For this purpose, method of square comparison, vector method, beam tracing, among others, can be applied. In this research, vector method was used as the fastest algorithm. Second, the linear interpolation is executed by using three values. It includes three following steps:

- Step 1. An image is divided in non-overlapping triangles in such manner, that a sum of triangles squares would be equal to a square of total image projection. Then for all triangles Step 2 and Step 3 are executed.
- Step 2. A calculation of distances ( $d_1$ ,  $d_2$ ,  $d_3$ ) from the vertexes of triangle to the required point and a normalization of these distances so that their sum is equalled 1.
- Step 3. A calculation of color function values in the required point as a sum of the weighing function values in the vertexes of triangle (multiplied on the normalized distances).

A procedure of panorama stitching is very complex with different transformation models in top, middle, and bottom parts of overlapping area. It is necessary to use the compulsory lines straightening or the morphing in the final panoramic image. The morphing parameters can be fitted between two unmatched fragments in the initial frames. However, these issues require the following investigations.

#### ***4.5.5 Visualization of High Speed Objects in Panoramic Images***

In common case, a video sequence involves visual projections of objects moving with high speed. Due to such high speed, an object can appear in several frames selected for panorama construction. Two variants of representation exist:

- Show a moving object in a single from the selected frames. It will be mean that a moving object is a single in a scene.
- Show all views of a moving object in the selected frames. It will map a trajectory of moving object in a scene.

The choice depends from a goal of panorama application. This issue is outline of the current research. Previously, the scenes without objects moving with high speed were used for experiments.

## 4.6 Lighting Improvement of Panoramic Images

It is difficult to guess, that a shooting is executed under the ideal lighting conditions. Usually, this assumption is inversely. All digital processing methods and algorithms are based on the analysis of intensity or color functions, describing an image. Therefore, the issues of lighting are very important. Two tasks can be formulated relatively to a panorama construction—the enhancement of initial selected frames (Sects. 4.6.1 and 4.6.2) and the improvement of visibility of stitching areas (Sect. 4.6.3).

### 4.6.1 Application of Enhancement Multi-scale Retinex Algorithm

Between the known approaches for spectrum enhancement of color images such as histogram approach, homomorphic filtering, and the Retinex algorithm, the last one is the most suitable for panoramic images. This is explained that the Retinex (formed from the words “retina” and “cortex”) algorithm is the advanced method, which simulates the adaptation of human vision for dark and bright regions [52].

The Single-Scale Retinex (SSR) algorithm demonstrates the best results for a grey-scale images processing and has difficulties for color images. The Multi-Scale Retinex (MSR) algorithm provides the processing of color images. The 1D retinex function  $R_i(x, y, \sigma)$  according to the SSR-model calculates differences of logarithmic functions given by Eq. (4.20), where  $I_i(x,y)$  is an input image function in the  $i$ th spectral channel,  $c$  is a scale coefficient, sign “\*” represents a convolution of the input image function  $I_i(x,y)$ , and the surrounding function  $F(x, y, c)$ . Often the surrounding function  $F(x, y, c)$  has a view of Gaussian function including a scale vector  $\sigma$ .

$$R_i(x, y, \sigma) = \log\{I_i(x, y)\} - \log\{F(x, y, c) * I_i(x, y)\} \quad (4.20)$$

The MSR-model  $R_{M_i}(x, y, \mathbf{w}, \boldsymbol{\sigma})$  in the  $i$ th spectral channel is calculated by Eq. (4.21), where  $\mathbf{w} = (w_1, w_2, \dots, w_m)$ ,  $m = 1, 2, \dots, M$  is a weight vector of 1D Retinex functions in the  $i$ th spectral channel  $R_i(x, y, \boldsymbol{\sigma})$ ,  $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \dots, \sigma_n)$ ,  $n = 1, 2, \dots, N$  is a scale vector of 1D output Retinex function. A sum of weigh vector  $\mathbf{w}$  components is equaled 1.

$$R_{M_i}(x, y, \mathbf{w}, \boldsymbol{\sigma}) = \sum_{n=1}^N w_n R_i(x, y, \sigma_n) \quad (4.21)$$

The basic SSR and MSR algorithm improve the shadow regions well. However, they are non-useful for bright region processing. The Enhanced Multi-Scale Retinex (EMSR) algorithm based on an adaptive equalization of spectral ranges in dark bright regions simultaneously was developed by Favorskaya and Pakhirka [53]. The EMSR algorithm uses a special curve representing as a logarithmic dependence of image function for low values of intensity and a logarithmic dependence of inverse image function for upper values of intensity.

The application of such algorithms permits to balance the lighting in the initial images. The experiments show that a contour accuracy becomes higher especially in dark regions. However, a total visibility is worse because of too sharp edges of visual objects. For this purpose, the edge smoothing procedure can be recommended.

### 4.6.2 The Edges Smoothing Procedure

All Retinex similar algorithms increase a sharp of result image significantly. This is not essential for computer processing. However, a visibility of result image can not satisfy the user.

The image sharpness improves the details in a result image by the blurring application. Some known filters solve this problem, for example, the High pass filter, the Laplacian filter, or the Unsharp masking filter. All these filters increase the sharpness values by a contrast amplification of the tonal transitions. The main disadvantage of the High pass and the Laplacian filters consists in sharpening not only image details but also a noise. An unsharp masking filter blurs a copy of original image by Gauss function and determines a subtraction between the received image and input image, if their differences exceed some threshold value.

The Enhanced Unsharp Masking (EUM) filter proposed by Favorskaya and Pakhirka [53] improves the output image by joined compositing of contour performance and equalization performance based on empirical dependences. Let notice that the application of the EMSR algorithm and/or the edge smoothing procedure is not always required in a panoramic image.

### 4.6.3 Blending Algorithm for Stitching Area

The improvement of visibility in the stitching areas is often required in final panoramic images. A transparency-blending method, based on  $\alpha$ -composite transparent surface layers in a back-to-front order to generate the effect of transparency, is not suitable for panorama improvement. The Point-Based Rendering (PBR) approach attracts a high interest in geometric modeling and rendering primitives as an alternative to triangle meshes. The PBR-blending is used to interpolate between overlapping point splats within the same surface layer to achieve the smooth rendering results. Zhang and Pajarola [54] proposed a deferred blending concept, which enables the hardware accelerated transparent PBR with combined effects of multi-layer transparency, refraction, specular reflection, and per-fragment shading.

Recently, the region-based image algorithms were developed to improve the pixel-based algorithms. They include the Poisson fusion algorithm [55], the Laplacian pyramid transform, contrast pyramid, discrete or complex wavelet transform [56], curvelet transform [57], contourlet transform [58], among others. The interesting approach for combining a set of registered images in a composite mosaic with no visible seams and minimal texture distortion was proposed by Gracias et al. [59]. In this research, a modification of the PBR-blending was applied, which includes the steps as mentioned below:

- Step 1. The segmentation of blending area near a stitching line. First, the Weighed Coefficients Masks (WCMs) are built for each of input images. The size of the WCMs is equal to the size of input image. Each element has a back-proportional value of distance between the center of input image and a current point. Then the received WCMs are imposed each other, and their elements are compared to build the Binary Blending Masks (BBMs). If a value of current element in the WCM cannot be compared with any element from the WCM of other image, then the corresponding value of the BBM receives the minimal value. (It means that a current pixel is located in a final image.) If a value of current element in the WCM is less than a value of current element in the WCM of other image, then a maximum value is assigned to the corresponding element of the BBM (a current pixel is not moved in a final image), and vice verse.
- Step 2. After the BBMs forming, these masks are blended by Gauss filter in three sub-bands. Other pixels from the input images are moved in final panoramic image without any changing.

The example of lighting enhancement and seamless blending is situated in Fig. 4.5. Two initial images were hand-held received by using Nikon D5100 camera with the autoexposure.

The following improvement of blending is connected with the multi-scale or the multi-orientation sub-bands framework with a number of bands, not more than 4. The main idea is to blend the sub-bands of image with various blending degree (if a frequency is low, then a blending degree is high).



**Fig. 4.5** Example of lighting enhancement and the PBR-blending by using of Gauss filter in three sub-bands: **a** stitching panoramic image, **b** stitching panoramic image with lighting enhancement, **c** stitching panoramic image with lighting enhancement and blending, **d**, **e**, **f** fragments with artifacts, **g**, **h**, **i** fragments with lighting enhancement, **j**, **k**, **l** fragments with lighting enhancement and blending



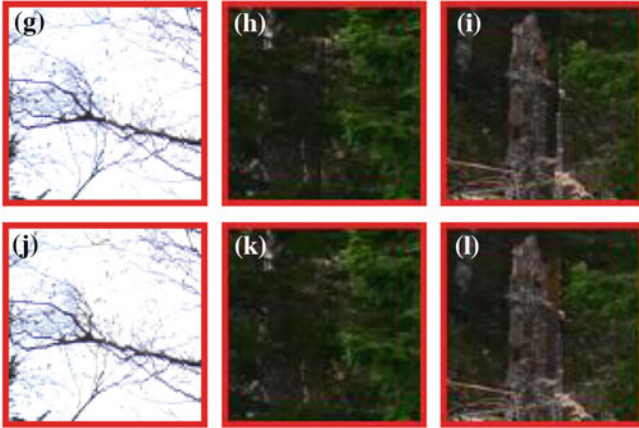


Fig. 4.5 (continued)

## 4.7 Discussion of Experimental Results

The software tool ‘‘Panorama Builder’’, v. 1.07 supports the main functions for a panorama construction such as frames or images loading, execution of automatic stitching procedure, save of results, and tuning of algorithm parameters. The designed software tool is written on C# language and uses the open libraries OpenCV and EmguCV. The calculations are realized by the central processor card and the graphic card, if it supports NVIDIA/CUDA technology. All discussed and proposed algorithms were realized in this program.

Figures 4.6, 4.7, 4.8, and 4.9 demonstrate a visual tracking of basic algorithmic operations to receive a final panoramic image. The initial images represented in Figs. 4.6, 4.7, 4.8, and 4.9 were received during the hand-held shooting by using Panasonic HDC-SD800 camera with the weighted auto exposure. Figures 4.6 and 4.7 show the close example with bad and good stitching. The bad stitching in Fig. 4.6 can be explained by the unsuccessful selection of frames, when a building tower crane was in motion with different directions of a jib. The final visual results in Figs. 4.7, 4.8 and 4.9 are enough appropriate.

The number of feature point correspondences in left image, right image, and overlapping area as well as the calculation time for different sizes of initial images  $640 \times 480$  (VGA),  $800 \times 600$  (SVGA),  $1,024 \times 768$  (XGA), and  $1,280 \times 960$  from Figs. 4.6, 4.7, 4.8 and 4.9 are presented in Table 4.1. During experiments, the different values of the SURF parameter and a brightness threshold, which influences on a number of feature points correspondences, was applied. The time of homography processing directly depends from a quantity of common feature point correspondences. All results are given for PC configuration: Intel Pentium Dual-Core T4300 @ 2.10 GHz, 3 Gb RAM.

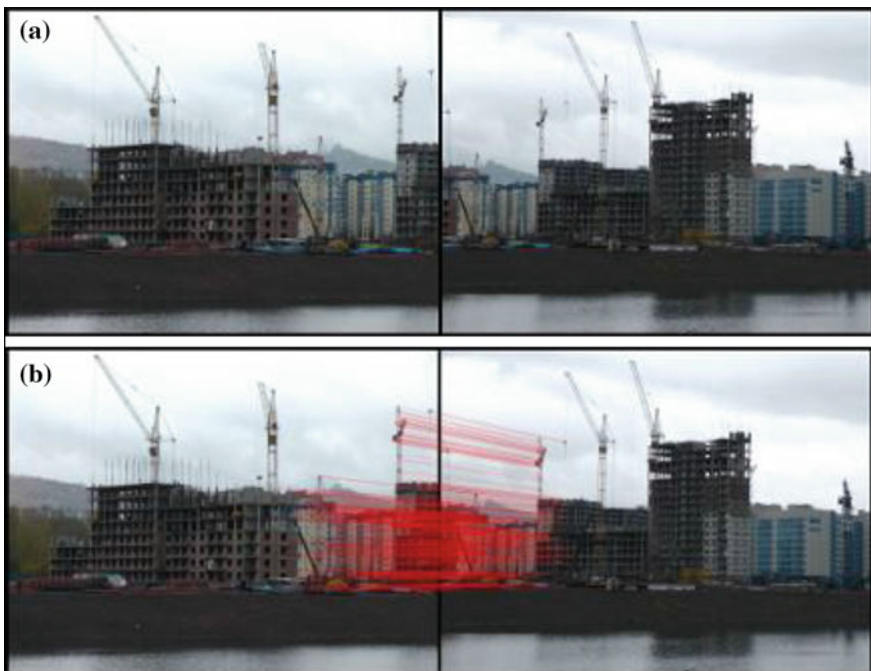


The analysis of data from Table 4.1 shows that the increment of resolution leads to the increased number of feature points detected in the input images. If the number of feature points increases, then the computational cost becomes higher. Therefore, a calculation time for the homography parameters increases. The stitching results are different by a mutual location of fragments for various values of the SURF brightness threshold. However, the lines of stitching save their visibility. The item “Road” is characterized by the increased number of feature points. This is explained by a wide overlapping area of the initial images.

Also the coefficients of homography matrixes for such examples are shown in Table 4.2.

The following investigations were connected with the parameters behavior during the perspective transformation. The results of top, middle, and bottom regions in an overlapping area are situated in Fig. 4.10. In Table 4.3, the coefficients of homography matrix of top, middle, and bottom regions are represented.

As it is seen from Fig. 4.10, the middle and bottom models are close, the top model is differed. This is explained by shooting from the Earth surface. However, such differences have not a significant value in landscape images.



**Fig. 4.6** The item “Houses failure”: **a** input images, **b** feature points matching, **c** projection result, **d** final panoramic image, **e** *top* fragment with an artifact stitching, **f** good *middle* fragment, **g** *bottom* fragment with an artifact stitching

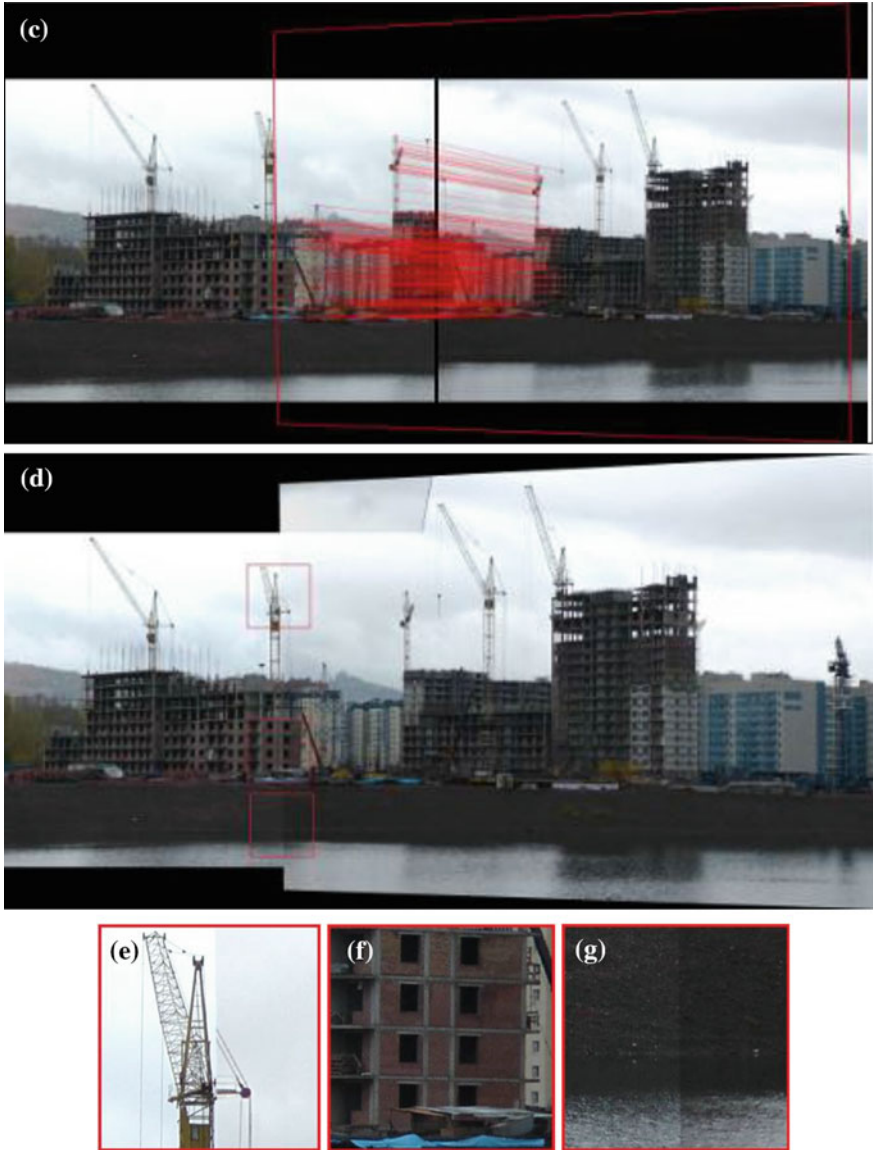
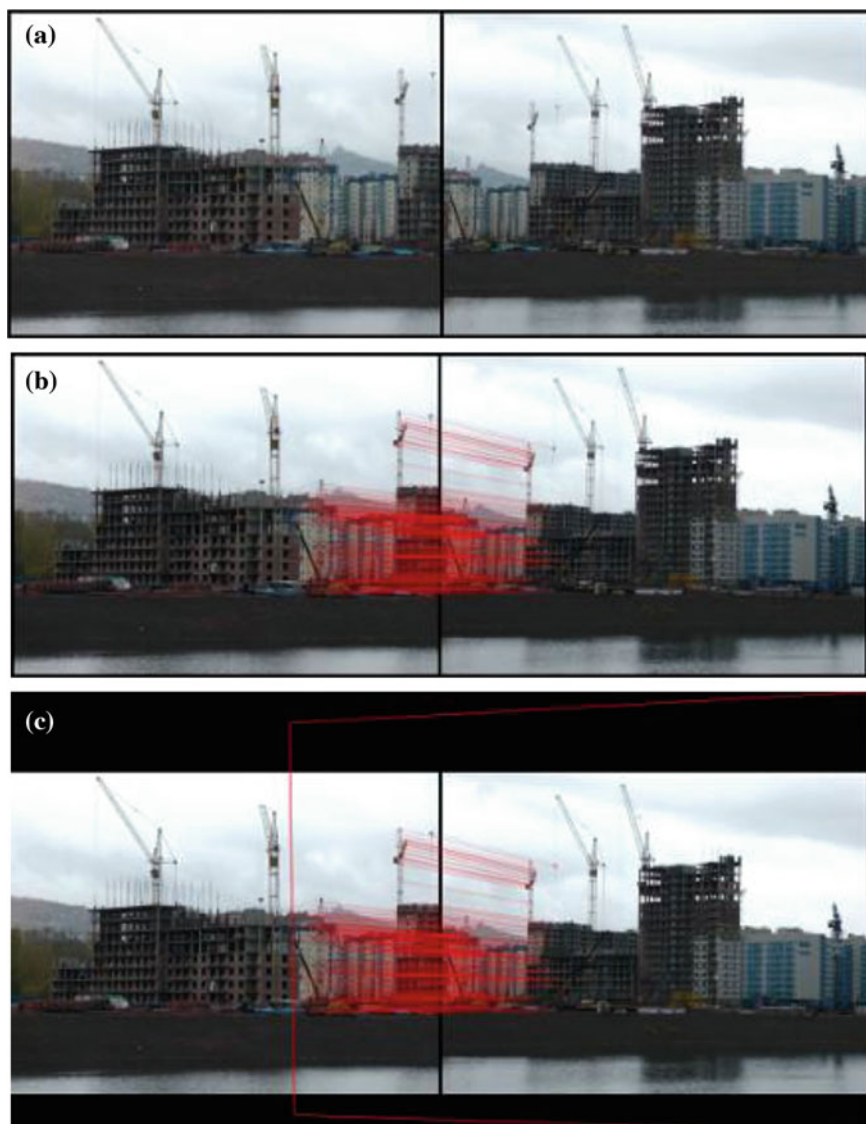


Fig. 4.6 (continued)



**Fig. 4.7** The item “Houses”: **a** input images, **b** feature points matching, **c** projection result, **d** final panoramic image, **e** good *top* fragment, **f** good *middle* fragment, **g** *bottom* fragment with an artifact stitching

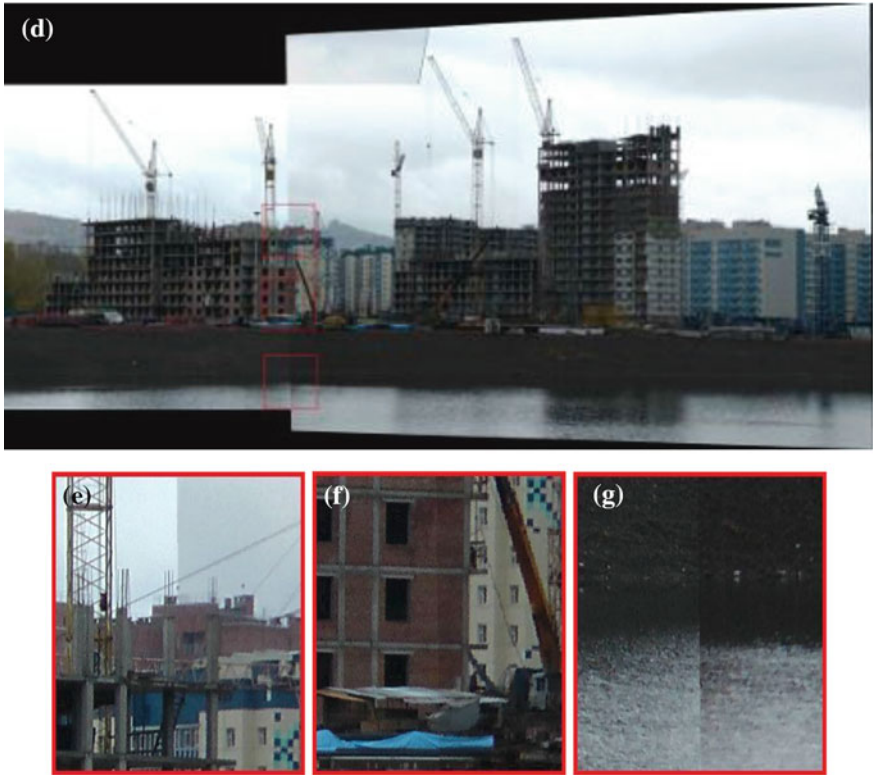


Fig. 4.7 (continued)



Fig. 4.8 The item “Road”: **a** input images, **b** feature points matching, **c** projection result, **d** final panoramic image, **e** good *top* fragment, **f** *middle* fragment with an artifact stitching, **g** *bottom* fragment with an artifact stitching



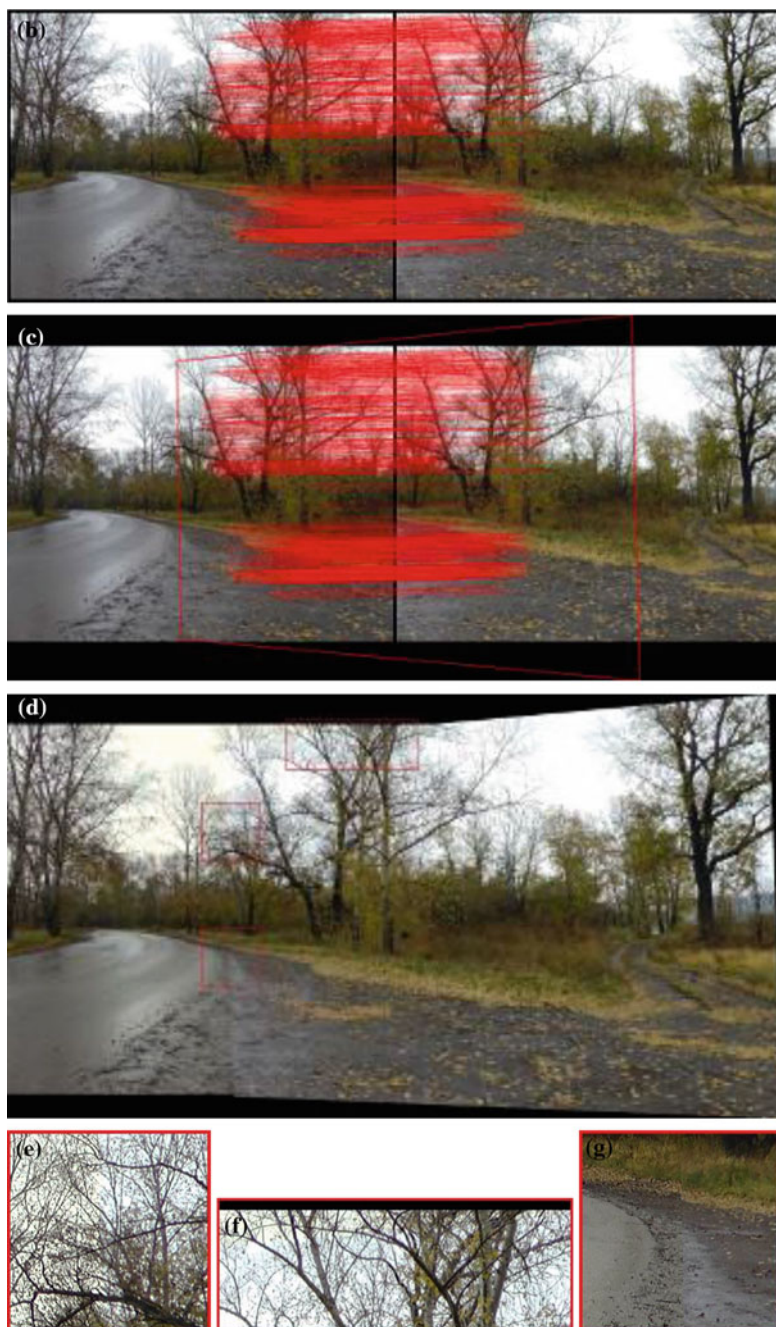
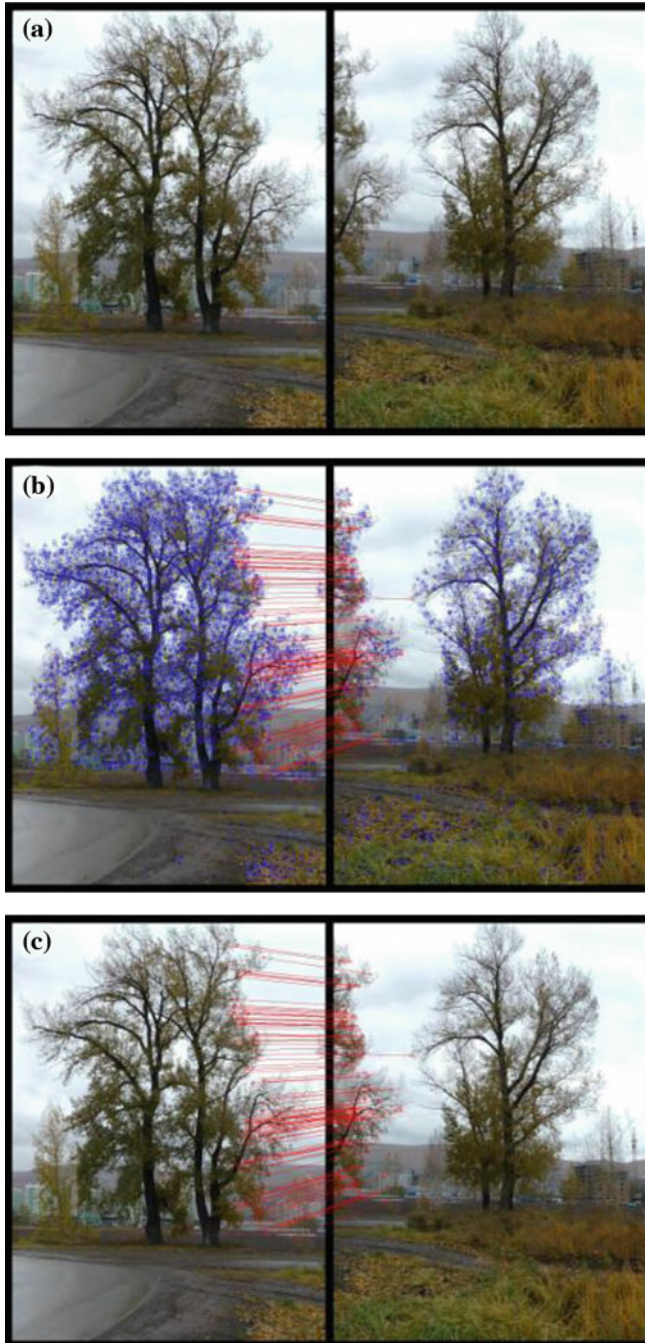


Fig. 4.8 (continued)



**Fig. 4.9** The item “Trees”: **a** input images, **b** all detected feature points, **c** feature points matching, **d** projection result, **e** final panoramic image, **f** good *top* fragment, **g** good *middle* fragment, **h** good *bottom* fragment

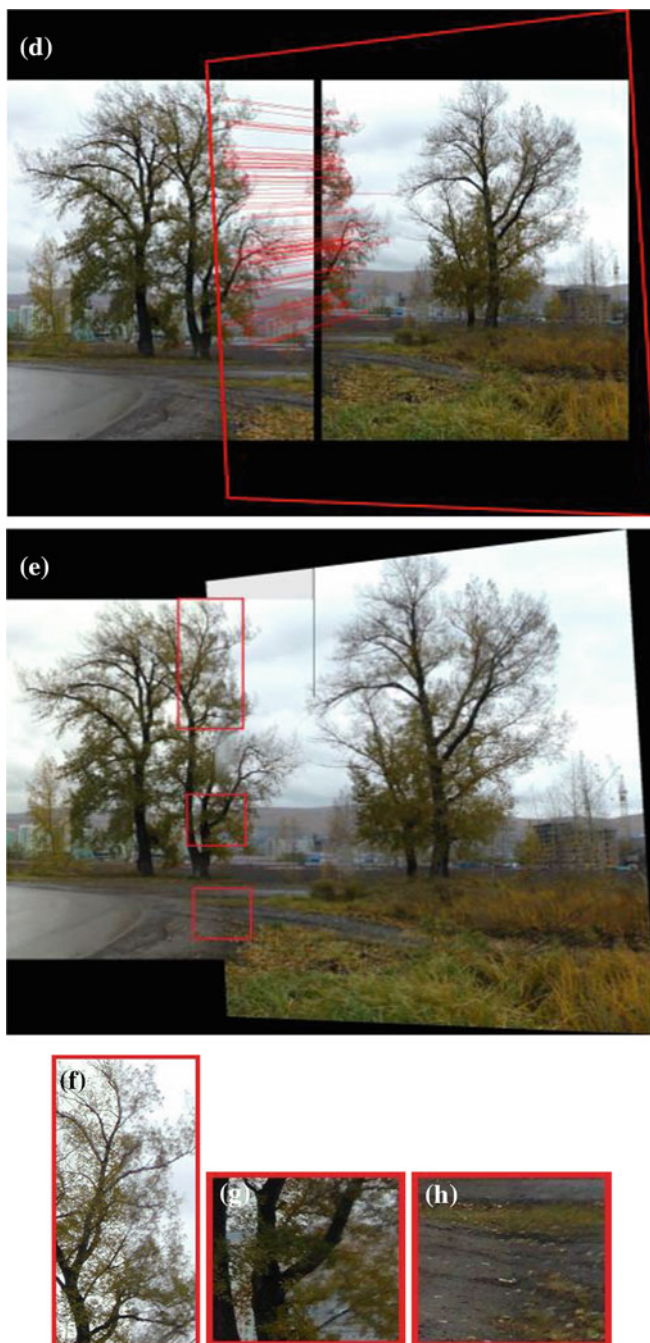


Fig. 4.9 (continued)

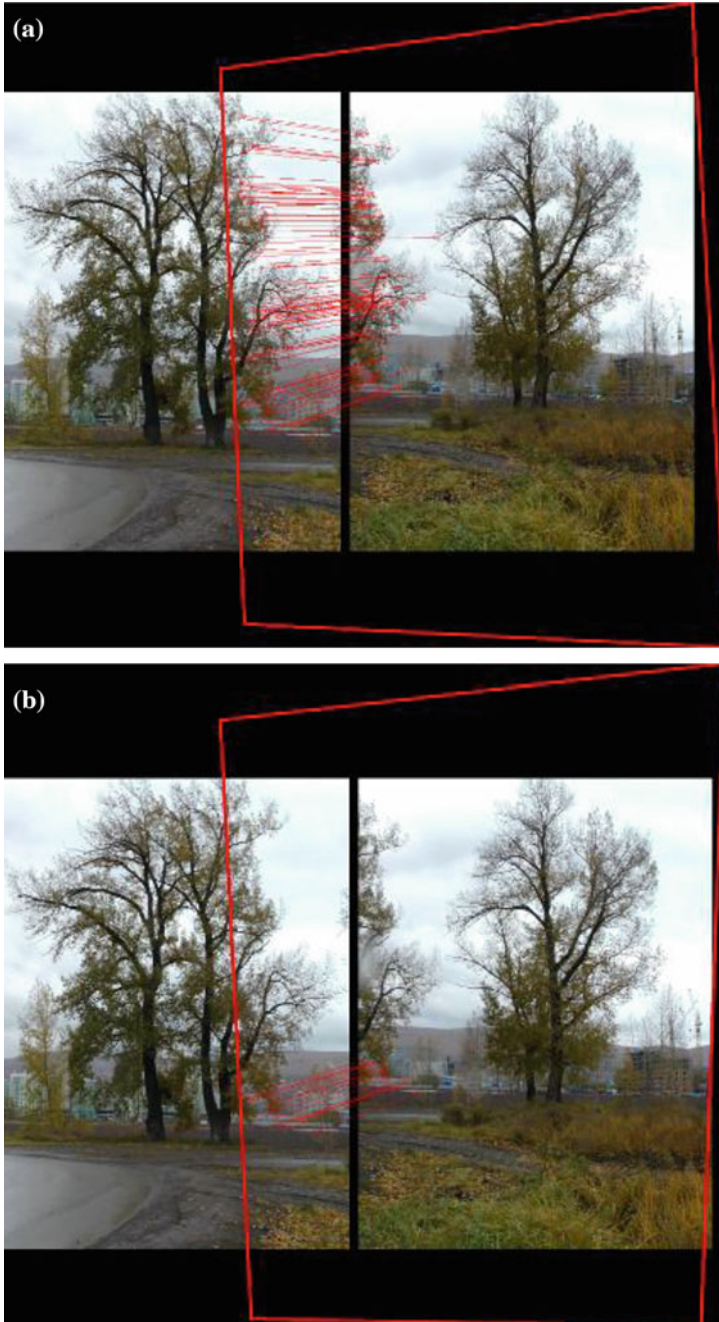
**Table 4.1** The parameters of panoramic images creation

Items	Frame sizes, pixels	SUFR brightness threshold	Numbers of feature points correspondence			Time, s
			Left image	Right image	Common	Homography processing
Houses failure and Houses	640 × 480	500	592	830	72	0.784
	800 × 600	500	1,285	1,235	126	1.235
	1,024 × 768	500	1,828	1,747	162	2.141
		1,250	744	716	101	1.223
	1,280 × 960	500	2,631	2,643	241	3.350
		1,250	1,100	1,082	105	1.868
1,500		859	875	87	1.679	
Road	640 × 480	500	1,595	1,322	44	1.286
	800 × 600	500	2,636	2,626	279	2.686
	1,024 × 768	500	4,255	4,372	622	5.152
		1,250	2,438	2,459	284	2.649
	1,280 × 960	500	7,453	7,411	819	11.847
		1,250	4,322	4,175	575	5.360
1,500		3,739	3,608	459	4.480	
Trees	640 × 480	500	1,595	1,322	44	1.286
	800 × 600	500	2,527	2,069	62	2.232
	1,024 × 768	500	4,043	3,515	99	4.597
		1,250	2,603	1,828	65	2.436
	1,280 × 960	500	6,508	6,002	112	9.747
		1,250	4,134	3,168	60	4.704
1,500		3,669	2,695	60	4.000	

**Table 4.2** Values of coefficients in homography matrixes

Parameter	Items			
	Houses failure	Houses	Road	Trees
h <sub>1</sub>	1.12064947	1.12458137	0.77233217	1.09293649
h <sub>2</sub>	0.01455373	0.02066196	0.01209813	0.04614185
h <sub>3</sub>	1597.91839937	1666.59485125	1150.95532389	498.58782054
h <sub>4</sub>	-0.04899116	-0.05201935	-0.09843960	-0.15573117
h <sub>5</sub>	1.21681933	1.22309076	0.93694815	1.19738618
h <sub>6</sub>	-280.89284415	-294.01541213	96.86158032	-52.85478987
h <sub>7</sub>	-0.00004299	-0.00004421	-0.00009345	-0.00017982
h <sub>8</sub>	0.00000109	0.00000252	-0.00000203	-0.00001127





**Fig. 4.10** The item “Trees” with various projective models: **a** total projection result, **b** *bottom* projection result, **c** *middle* projection result, **d** *top* projection result

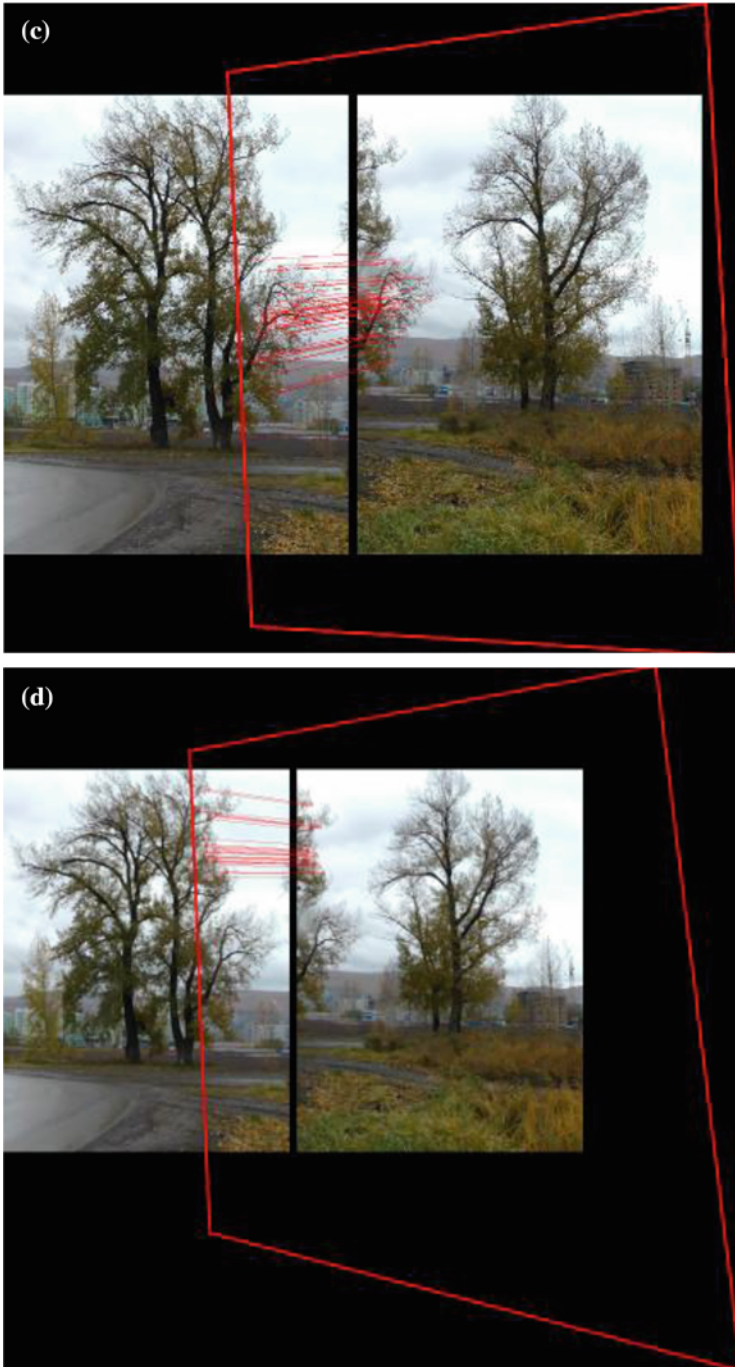


Fig. 4.10 (continued)

**Table 4.3** Coefficients in homography matrix with various projective models for item “Trees”

Parameter	Region in overlapping area			
	Total	Bottom	Middle	Top
Total number of feature points	101	18	48	35
$h_1$	1.09293649	1.22312982	1.07997240	0.94064686
$h_2$	0.04614185	0.12969314	0.04410856	0.01186247
$h_3$	498.58782054	487.78575565	499.32996392	499.70470638
$h_4$	-0.15573117	-0.12683291	-0.16033099	-0.18758969
$h_5$	1.19738618	1.40435098	1.18993275	1.16906946
$h_6$	-52.85478987	-126.8548956	-50.06574907	-49.89502673
$h_7$	-0.00017982	-0.00012848	-0.00019690	-0.00039302
$h_8$	-0.00001127	0.0001145	-0.00001319	-0.00007330

## 4.8 Conclusion

In this chapter, some methods and algorithms were investigated for panorama construction from frames, which are received from multi-view cameras in the outdoor scenes. All steps of panorama construction were described with a special attention for the main issues—the geometrical and lighting alignment during the images stitching. Some novel procedures were proposed for selection “good” frames from video sequences, more accurate estimation of projective parameters in an overlapping area, and the lighting improvement of panoramic image by a point-based blending in the stitching area. The illustrations show well the visible intermediate results of images stitching. The calculated parameters of homography matrixes indicate on the necessity of following investigations in interpolation, morphing, and warping transformation for improvement of result panoramic images. Also it is important to accelerate an automatic panorama construction for practical implementation.

## References

1. Briggs AJ, Detweiler C, Li Y, Mullen PC, Scharstein D (2006) Matching scale-space features in 1D panoramas. *Comput Vis Image Underst* 103(3):184–195
2. Dang TK, Worring M, Bui TD (2011) A semi-interactive panorama based 3D reconstruction framework for indoor scenes. *Comput Vis Image Underst* 115(11):1516–1524
3. Zhang W, Cham WK (2012) Reference-guided exposure fusion in dynamic scenes. *J Vis Commun Image Represent* 23(3):467–475
4. Chen H (2008) Focal length and registration correction for building panorama from photographs. *Comput Vis Image Underst* 112(2):225–230

5. Ni D, Chui YP, Qu Y, Yang X, Qin J, Wong TT, Ho SSH, Heng PA (2009) Reconstruction of volumetric ultrasound panorama based on improved 3D SIFT. *Comput Med Imaging Graph* 33(7):559–566
6. Powell I (1994) Panoramic lens. *Appl Opt* 33(31):7356–7361
7. Xiong Y, Turkowski K (1997) Creating image-based VR using a self-calibrating fisheye lens. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp 237–243
8. Luong HQ, Goossens B, Philips W (2011) Joint photometric and geometric image registration in the total least square sense. *Pattern Recogn Lett* 32(15):2061–2067
9. Fan BJ, Du YK, Zhu LL, Tang YD (2011) A robust template tracking algorithm with weighted active drift correction. *Pattern Recogn Lett* 32(9):1317–1327
10. Zhao G, Lin L, Tang Y (2013) A new optimal seam finding method based on tensor analysis for automatic panorama construction. *Pattern Recogn Lett* 34(3):308–314
11. Beier T, Neely S (1992) Feature-based image metamorphosis. In: *Proceedings of the 19th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH'92*, vol 26, no 2. pp 35–42
12. Haenselmann T, Busse M, Kopf S, King T, Effelsberg W (2009) Multi perspective panoramic imaging. *Image Vis Comput* 27(4):391–401
13. Pazzi RW, Boukerche A, Feng J, Huang Y (2010) A novel image mosaicking technique for enlarging the field of view of images transmitted over wireless image sensor networks. *J Mob Netw Appl* 15(4):589–606
14. Jain DK, Saxena G, Singh VK (2012) Image mosaicking using corner techniques, *Int Conf on Communication Systems and Network Technologies* 79–84
15. Yang J, Wei L, Zhang Z, Tang H (2012) Image mosaic based on phase correlation and Harris operator. *J Comput Inf Syst* 8(6):2647–2655
16. Kwon OS, Ha YH (2010) Panoramic video using scale invariant feature transform with embedded color-Invariant values. *IEEE Trans Consum Electron* 56(2):792–798
17. Kim D, Hong KS (2008) Practical background estimation for mosaic blending with patch-based Markov random fields. *Pattern Recogn* 41(7):2145–2155
18. Bao P, Xu D (1999) Complex wavelet-based image mosaics using edge-preserving visual perception modeling. *Comput Graph* 23(3):309–321
19. Bhosle U, Roy SD, Chaudhuri S (2005) Multispectral panoramic mosaicing. *Pattern Recogn Lett* 26(4):471–482
20. Agarwala A, Zheng C, Pal C, Agrawala M, Cohen M, Curless B, Salesin D, Szeliski R (2005) Panoramic video textures. *ACM Trans Graph* 24(3):821–827
21. Brown M, Lowe DG (2007) Automatic panoramic image stitching using invariant features. *Int J Comput Vis* 74(1):59–73
22. Li C, Ma L (2009) A new framework for feature descriptor based on SIFT. *Pattern Recogn Lett* 30(5):544–557
23. Deng H, Zhang L, Ma J, Kang Z (2011) Interactive panoramic map-like views for 3D mountain navigation. *Comput Geosci* 37(11):1816–1824
24. Fiala M, Basu A (2002) Hough transform for feature detection in panoramic images. *Pattern Recogn Lett* 23(14):1863–1874
25. Kim DH, Yoon YI, Choi JS (2003) An efficient method to build panoramic image mosaics. *Pattern Recogn Lett* 24(14):2421–2429
26. Zhu Z, Xu G, Riseman EM, Hanson AR (2006) Fast construction of dynamic and multi-resolution 360° panoramas from video sequences. *Image Vis Comput* 24(1):13–26
27. Zhu Z, Riseman EM, Hanson AR (2004) Generalized parallel-perspective stereo mosaics from airborne videos. *IEEE Trans Pattern Anal Mach Intell* 26(2):226–237
28. Steedly D, Pal C, Szeliski R (2005) Efficiently registering video into panoramic mosaics. In: *IEEE International Conference on Computer Vision (ICCV'2005)*, vol 2. pp 15–21
29. Langlotz T, Degendorfer C, Mulloni A, Schall G, Reitmayr G, Schmalstieg D (2011) Robust detection and tracking of annotations for outdoor augmented reality browsing. *Comput Graph* 35(4):831–840

30. Seitz SM, Dyer CR (1997) Viewing morphing: uniquely predicting scene appearance from basis images. DARPA Image Understanding Workshop, pp 881–887
31. Hernández-Mier Y, Blondel WCPM, Daula C, Wolf D, Guillemin F (2010) Fast construction of panoramic images for cystoscopic exploration. *Comput Med Imaging Graph* 34(7):579–592
32. Fathima AA, Karthik R, Vaidehi V (2013) Image stitching with combined moment invariants and SIFT features. *Procedia Comput Sci* 19:420–427
33. Fischler MA, Bolles RC (1981) Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Commun ACM* 24(6):381–395
34. Yan Q, Xu Y, Yang X, Nguyen T (2014) HEASK: Robust homography estimation based on appearance similarity and keypoint correspondences. *Pattern Recognit* 47(1):368–387
35. Torr PHS, Zisserman A (2000) MLESAC: a new robust estimator with application to estimating image geometry. *Comput Vis Image Underst* 78(1):138–156
36. Chum O, Matas J, Obdrzalek S (2004) Enhancing RANSAC by generalized model optimization. In: *Asian Conference on Computer Vision, ACCV*. pp 812–817
37. Nashat S, Abdullah A, Abdullah MZ (2012) Unimodal thresholding for Laplacian-based Canny-Deriché filter. *Pattern Recogn Lett* 33(10):1269–1286
38. Shen F, Zhao Y, Jiang X, Suwa M (2009) Recovering high dynamic range by multi-exposure retinex. *J Vis Commun Image Represent* 20(8):521–531
39. Dokládál P, Dokládálová E (2011) Computationally efficient, one-pass algorithm for morphological filters. *J Vis Commun Image Represent* 22(5):411–420
40. Fernandes LAF, Oliveira MM (2008) Real-time line detection through an improved Hough transform voting scheme. *Pattern Recogn* 41(1):299–314
41. Maxwell EA (1946) *Methods of plane projective geometry based on the use of general homogeneous coordinates*. Cambridge University Press, Cambridge
42. Maxwell EA (1951) *General homogeneous coordinates in space of three dimensions*. Cambridge University Press, Cambridge
43. Roberts LG (1965) *Homogeneous matrix representations and manipulations of n-dimensional constructs*, Technical Representation Document MS 1405, Lincoln Laboratory, MIT, Cambridge
44. Chum O, Matas J (2002) Randomized ransack with t(d,d) test. *British Machine Vision Conference (BMVC'2002)*, vol 2. pp 448–457
45. Capel D (2005) An effective bail-out test for ransack consensus scoring. *British Machine Vision Conference (BMVC'2005)*. pp 629–638
46. Matas J, Chum O (2005) Randomized RANSAC with sequential probability ratio test. In: *10th IEEE International Conference on Computer Vision*, vol 2, pp 1727–1732
47. Chum O, Matas J (2008) Optimal randomized ransac. *IEEE Trans Pattern Anal Image Underst* 30(8):1472–1482
48. Raguram R, Frahm JM, Pollefeys M (2008) A comparative analysis of RANSAC techniques leading to adaptive real-time random sample consensus. In: *10th European Conference on Computer Vision*, vol 2. pp 500–513
49. Cheng CM, Lai SH (2009) A consensus sampling technique for fast and robust model fitting. *Pattern Recogn* 42(7):1318–1329
50. Kiciak P (2011) Bicubic B-spline blending patches with optimized shape. *Comput Aided Des* 43(2):133–144
51. Kineri Y, Wang M, Lin H, Maekawa T (2012) B-spline surface fitting by iterative geometric interpolation/approximation algorithms. *Comput Aided Des* 44(7):697–708
52. Meylan L, Alleysson D, Süssstrunk S (2007) Model of retinal local adaptation for the tone mapping of color filter array images. *J Opt Soc Am A*: 24(9):2807–2816
53. Favorskaya M, Pakhirka A (2012) A way for color image enhancement under complex luminance conditions. In: Watanabe T, Watada J, Takahashi N, Howlett RJ, Jain LC (eds) *Intelligent interactive multimedia: systems and services*. Springer, Berlin
54. Zhang Y, Pajarola R (2007) Deferred blending: Image composition for single-pass point rendering. *Comput and Graph* 31(2):175–189

55. Sun J, Zhu H, Xu Z, Han C (2013) Poisson image fusion based on Markov random field fusion model. *Inf Fusion* 14(3):241–254
56. Mills A, Dudek G (2009) Image stitching with dynamic elements. *Image Vis Comput* 27(10):1593–1602
57. Gómez F, Romero E (2011) Rotation invariant texture characterization using a curvelet based descriptor. *Pattern Recogn Lett* 32(16):2178–2186
58. Yang S, Wang M, Jiao L, Wua R, Wang Z (2010) Image fusion based on a new contourlet packet. *Inf Fusion* 11(2):78–84
59. Gracias N, Mahoor M, Negahdaripour S, Gleason A (2009) Fast image blending using watersheds and graph cuts. *Image Vis Comput* 27(5):597–607