ISSN 1064-2269, Journal of Communications Technology and Electronics, 2007, Vol. 52, No. 3, pp. 332–338. © Pleiades Publishing, Inc., 2007.
Original Russian Text © Yu.V. Slyn'ko, V.N. Lagutkin, A.P. Luk'yanov, 2007, published in Radiotekhnika i Elektronika, 2007, Vol. 52, No. 3, pp. 351–357.

THEORY AND METHODS OF SIGNAL PROCESSING

Development and Investigation of Real-Time Robust Algorithms for Estimating the Parameters of Geometric Transformations of Video-Sequence Frames

Yu. V. Slyn'ko, V. N. Lagutkin, and A. P. Luk'yanov

Received December 1, 2005

Abstract—The problem of estimating the parameters of geometric transformations of the frames in a video sequence is considered. The solution to this problem is found through a combination of three basic approaches: the optical-flow feature-point methods and the direct correlation methods. A procedure for the detailed analysis of the behavior of the correlation function is used to ensure stable real-time operation of the proposed algorithms on modern (even unspecialized) computing systems for a wide range of shooting conditions.

PACS numbers: 07.05.Pj, 42.30.Va

DOI: 10.1134/S1064226907030072

INTRODUCTION

One of the main problems of video-sequence processing consists in determining the parameters of geometric transformations of the video frames. Solution of this problem consolidates the solution of problems related to image stabilization, tracking of moving objects, estimating the parameters of motion of these objects, and mosaic construction.

Furthermore, many image-processing algorithms, including the foreground separation and pattern recognition algorithms, require prestabilization of images.

There exist different approaches to estimation of the frame deformation parameters: the optical flow method [1] based on computation of the shift vector for each pixel of an image, the feature-point method [2, 3] based on determination of the characteristic singularities of an image and the correlation between these singularities in successive frames, and the direct method using the brightness of input-image points [4].

However, these algorithms are not designed for real-time processing of large frames and/or operation on unspecialized computing systems. These algorithms operate stably when input images satisfy the selected mathematical model; however, operation often becomes unstable when this model changes or input images are inconsistent with it.

In this paper, we describe algorithms that determine the geometric transformation of frames and satisfy the following conditions:

(i) These algorithms must operate stably (i.e., robustly) with video sequences of different types, including low-quality video sequences obtained in an unfavorable shooting environment.

(ii) These algorithms must be capable of performing real-time processing of video sequences on modern unspecialized computing systems.

General techniques for improving the reliability and accuracy of algorithms and their application to the algorithms based on the synthesis of the above three approaches are described.

1. IMAGE-DISTORTION MODEL AND FORMULATION OF THE PROBLEM

The problem under study is formulated as follows. For each frame of a video sequence, it is necessary to estimate the parameters of its geometric transformation relative to the preceding frame. In addition, video-sequence frames are assumed to be images of the same stationary scene. Geometric transformation (spatial warping) can be interpreted as an arbitrary continuously differentiable one-to-one transformation of coordinates. The commonly used warping models are affine, projective, and quadratic transformations of coordinates that are determined by the following respective relationships:

$$\begin{pmatrix} x(t+1) \\ y(t+1) \end{pmatrix} = \begin{pmatrix} a_0 & a_1 \\ a_2 & a_3 \end{pmatrix} \begin{pmatrix} x(t) \\ y(t) \end{pmatrix} + \begin{pmatrix} a_4 \\ a_5 \end{pmatrix}, \quad (1)$$

$$\begin{pmatrix} x(t+1) \\ y(t+1) \end{pmatrix} = \frac{1}{a_6 x(t) + a_7 y(t)} \begin{pmatrix} a_0 & a_1 \\ a_2 & a_3 \end{pmatrix} \begin{pmatrix} x(t) \\ y(t) \end{pmatrix} + \begin{pmatrix} a_4 \\ a_5 \end{pmatrix}, \quad (2)$$

$$\begin{pmatrix} x(t+1) \\ y(t+1) \end{pmatrix} = \begin{pmatrix} a_0 + a_1 x(t) + a_2 y(t) + a_6 x(t)^2 + a_7 x(t) y(t) \\ a_3 + a_4 x(t) + a_5 y(t) + a_6 x(t) y(t) + a_7 y(t)^2 \end{pmatrix}, \quad (3)$$

where $x(t)$ and $y(t)$ are the coordinates of a scene point in a frame, $x(t+1)$ and $y(t+1)$ are the coordinates of this point in the next frame, and a_i are the desired unknown parameters of spatial warpings.

When real images are estimated, the problem is solved in the presence of various deviations from an ideal mathematical model: blurred areas, nonlinear distortions in the camera lenses, information loss due to encoding of a video sequence, etc.

An algorithm for estimating the parameters of frame-to-frame transformation can always be reduced to the following optimization problem:

$$\hat{Q} = \arg\min_{Q} X(\|F(t-1)_{ij}\|, \ \|T_Q(F(t))_{ij}\|), \quad (4)$$

where $\hat{Q}$ is the estimate of the vector of the parameters of frame-to-frame image warping, $Q = (a_1, a_2 \ldots)$, X is the "discrepancy" function that must approach the smallest value when its first and second arguments come closer to each other, $F(t)_{ij}$ is the frame corresponding to instant t, and T_Q is the frame transformation according to the specified vector of warping parameters.

The obtainment of the required reliability and speed of processing necessitates both minimization of function $X(F_1, F_2)$ and estimation of its behavior characterizing the quality of decision making. This circumstance is of primary importance when several processing steps are used to compute function $X(F_1, F_2)$. Depending on the circumstances, the quality of decision making can correspond to various parameters, for example, the accuracy of estimation of the optical-flow vector for a pixel in the optical-flow method and the accuracy of determination of the point coordinates in the feature-point method.

Information on the behavior of function $X(F_1, F_2)$ allows implementation of the following processing principles.

(i) The amount of computations is limited without loss in the quality and reliability of estimation. If the error of determination of the required parameters is known, computations can be terminated when this error reaches an acceptable value.

(ii) The computation accuracy is improved by weighting of the information obtained from various points. In the limiting case, the certainly erroneous information can be excluded. The image of a frame often contains small areas or objects that introduce nothing but an additional error (a noise component) into the computed parameters.

(iii) The difficulty or even impossibility of correct determination of image motion is predictable. In such a situation, it is possible either to regularize the solution and obtain an approximate acceptable estimate or to exclude a questionable frame from processing.

Application of the optimization procedure substantially improves the accuracy and reliability of algorithms and simultaneously diminishes the amount of computations, thereby speeding up algorithmic processing.

The basic algorithm described in this paper is the algorithm using the brightness of input-image points. However, sequential implementation of the above principles leads to creation of an algorithm based on the synthesis of three main approaches.

Here, function $X(F_1, F_2)$ is meant to be the rms residual function or any of its generalizations. This function is computed sufficiently fast; if necessary, computations can be performed with modern vector processors connected in parallel. Function $T_Q(F)$ is chosen to be a linear integer-shift function. The duration of its computation is much less than the duration of remained computations.

2. CORRELATION METHOD

In the simplest case, function $X(F_1, F_2)$ can be represented as

$$X(F(t-1), T_Q(F(t))) = \frac{1}{S(M)} \sum_{(i,j) \in M(dx, dy)} (F(t-1)_{ij} - F(t)_{i+dx, j+dy})^2, \quad (5)$$

and spatial warping has the form

$$\hat{Q} = (d\hat{x}, d\hat{y}) = \arg \min_{(dx, dy) \in D} X(F(t-1), T_Q F(t)), \quad (6)$$

where $\hat{Q} = (d\hat{x}, d\hat{y})$ are the shifts along two axes, D is the region of possible shifts in which the search is performed, $M(dx, dy)$ is the region used to compute the residual between frames (for example, the region with points (i, j) and $(i + dx, j + dy)$ belonging to $F(t-1)$ and $F(t)$, respectively, i.e., the region of osculation of these frames), and $S(M)$ is the area of region M (the number of points).

Function $X(F_1, F_2)$ will be referred to as a two-frame correlation function, which depends on arguments dx and dy corresponding to the shifts along two axes. Below, this function is designated as $X_{(F_1F_2)}(dx, dy)$.

This method is characterized by the multimodality of function $X_{(F_1F_2)}(dx, dy)$. Hence, the shift cannot be found with the use of the descent methods. It is neces-

sary to search for all possible variants. Owing to the high speed of computation of the correlation function, an exhaustive search does not diminish the speed of algorithm operation; however, the reliability of determination of the shift value is improved substantially.

Being applied to the entire frame, the correlation method requires too much computational time (see the results presented in Section 7).

To diminish the amount of computations, the proposed algorithm uses (i) a frame compression method, (ii) a method for selection of correlation windows, and (iii) a method of reduction of the region of possible shifts.

In addition, according to principles described in the preceding section, the quality of estimates is controlled at each processing step so as to diminish the number of unnecessary computations and perform only the computations required to attain an acceptable quality.

3. DETERMINING THE ESTIMATION ACCURACY

To implement the principles formulated in Section 1, a procedure for estimating the quality of decision making is required.

The following procedure is proposed. Let the values of function $X_{(F_1F_2)}(dx, dy)$ be computed in certain region D. The values of desired parameters correspond to the minimum of this function. Hence, to estimate the quality of these parameters, it is necessary to investigate the behavior of function $X_{(F_1F_2)}(dx, dy)$ near its minimum.

It will be assumed that the pixel noise is normally distributed and uncorrelated. In this case, the values of function $X_{(F_1F_2)}(dx, dy)$ obey a chi-square distribution, and its minimum value corresponds to the mean.

Thus, at each pixel, the noise variance is

$$\sigma_p = \sqrt{\frac{X_m}{N_p}}, \tag{7}$$

where X_m is the minimum value of function $X_{(F_1F_2)}(dx, dy)$ and N_p is the number of points in region $M(dx, dy)$ corresponding to the minimum.

The mean and variance of the values of function $X_{(F_1F_2)}(dx_i, dy_i)$ are determined from the following respective relationships:

$$m_i = \frac{X_m}{N_p}N_i, \quad \sigma_i = \frac{X_m}{N_p}\sqrt{2N_i}, \tag{8}$$

where m_i is the mean, σ_i is the variance, and N_i is the number of points in region $M(dx_i, dy_i)$ used to calculate function $X_{(F_1F_2)}(dx_i, dy_i)$.

The confidence region that can contain values of function $X_{(F_1F_2)}(dx_i, dy_i)$ that are distorted by noise is determined as follows:

$$\tilde{D} = \{dx_i, dy_i: X_{(F_1F_2)}(dx_i, dy_i) < m_i + B\sigma_i\}, \tag{9}$$

where B is the threshold characterizing the confidence probability of error.

Thus, for the point corresponding to the minimum value of function $X_{(F_1F_2)}(dx_i, dy_i)$, we obtained the region containing a true solution that can be found with a specified probability. This region sufficiently completely determines the quality of determination of the shift value.

However, since this estimate is essentially nonlinear, its application in the analysis is not always suitable. With the use of the second moments of the distribution, it is possible to replace this estimate with a simpler estimate, i.e., to apply a Gaussian model of errors with the covariance matrix of estimates of the unknown shift. The 2×2 inverse covariance matrix of the errors of the shift estimates is expressed as

$$\mathbf{C}^{-1} = \begin{Vmatrix} A & B \\ B & E \end{Vmatrix}. \tag{10}$$

Its elements are the estimated parameters of the ellipse approximating region $\tilde{D}$. According to the least square method, the equation of this ellipse is written as

$$\begin{aligned} A(x - x_m)^2 + 2B(x - x_m)(y - y_m) \\ + E(y - y_m)^2 = 0, \end{aligned} \tag{11}$$

where (x_m, y_m) is the point corresponding to the minimum of function $X_{(F_1F_2)}(dx, dy)$.

4. FRAME COMPRESSION METHOD

To ensure the reliability of operation of the proposed algorithm, the shift must be determined from the entire frame. However, processing of the entire frame via the correlation method requires a huge amount of computations. To ensure an acceptable computational speed, the shift can be determined with compressed frames, i.e., scaled-down frames containing a smaller number of points.

In this case, the shift can be successively estimated from frames of smaller scales as follows. An initial frame is successively compressed by a factor of 2 in order to obtain several frames of different scales. The shift is first determined for the frame of the largest scale and, then, with the use of the subsequent frames of smaller scales. In this study, a frame compressed 2^p times will be designated as F^p.

The shift of frames is calculated according to the following procedure.

(i) At the first step, region D_1 is chosen as a rectangle having specified dimensions (half-frame shifts are determined rather accurately). In this region, function $X_{(F_1^P F_2^P)}(dx, dy)$ representing the correlation function of two compressed frames is computed. The value of parameter P is chosen such that each side of the frames compressed 2^P times contains 15–30 pixels.

(ii) Region D_k is determined according to formulas (7), (8), and (9), where function $X_{(F_1F_2)}(dx, dy)$ is replaced with $X_{(F_1^{P-k}F_2^{P-k})}(dx, dy)$.

(iii) At the kth step, function $X_{(F_1^{P-k}F_2^{P-k})}(dx, dy)$ is computed for frames F_1^{P-k} and F_2^{P-k} (initial frames compressed $P-k$ times).

(iv) If the kth step is not final, we pass to point (ii) and perform the $(k+1)$th step. Otherwise, we pass to point (v). Note that the computation process must be terminated long before the step in which frames of the initial scale are used. First, comparison of large frames is a time-consuming operation. Second, when the frame size is large, computation of not only the main parameter of spatial warping (shift) but also other parameters (rotation, scaling, and nonlinear components) entails substantial difficulties. It is suitable to perform two or three steps and then pass to determination of the warping parameters with the use of several correlation windows. This procedure is described below.

(v) At the final step, a fractional shift is determined (for example, with the use of the methods described in [5]) and all shifts are reduced to the scale of the initial frame.

5. METHOD OF SELECTION OF CORRELATION WINDOWS

It is necessary to refine the rough estimate of the shift obtained from the compressed frames. In addition, the rotation angle, scaling factor, and other parameters of spatial warping must be estimated.

It is necessary to solve this problem with the use of the minimum possible number of computations. The solution can be obtained via a sufficiently effective approach based on correlation windows, which are suitably selected within a video frame.

Windows must be selected with allowance for the quality (information value) of different areas of an image. Here, the information value is interpreted as the mean error of the shift determined for a given area of a frame. As the correlation windows applied in the subsequent analysis, the areas with the highest information value must be selected. At the same time, it is necessary to take into account that the error of determination of the rotation angle will diminish with an increase in the distance between windows.

The kth window is selected according to the criterion

$$\hat{\vec{r}}_k = \arg\max_{\vec{r}_k}\left(I(\vec{r}_k)\left(\min_{i=1\ldots k-1}\left|\vec{r}_k - \hat{\vec{r}}_i\right| + C\right)\right), \qquad (12)$$

where $I(\vec{r}_k)$ is the information value of point $\vec{r}_k$ and C is a constant.

The information value of a window is computed via minimization of the autocorrelation function along several directions:

$$I(\vec{r}_k) = \min_{l=1\ldots L}\frac{1}{\sqrt{dx^2(l)+dy^2(l)}} \times \sum_{(i,j)\in W(\vec{r}_k)}\left(F_{ij} - F_{i+dx(l),\,j+dy(l)}\right)^2, \qquad (13)$$

where $(dx(l), dy(l))$ is the set of directions, $W(\vec{r}_k)$ is the correlation window centered at point $\vec{r}_k$, and L is the number of directions.

The autocorrelation function is selected as a criterion of the information value because the residual function plays the role of an objective function. Therefore, its slope is proportional to the accuracy of the shift estimate.

After correlation windows are selected, it is necessary to determine the shift of each window and combine data on all windows for computing the warping parameters of the entire frame.

The shift of each window is found via minimization of correlation function $X_{(F_1F_2)}(dx, dy)$. To eliminate subordinate minima, an exhaustive search for all variants is performed. In this case, if parameters (dx, dy) are changed with some step S, the number of computational operations can be diminished by a factor of S^2.

However, this approach can be used only when the decision-making quality is continuously controlled, because the probability of finding a subordinate minimum is very high. Hence, to achieve an acceptable level of probability upon terminating the rough search, it is necessary to determine region $\tilde{D}$ according to formulas (7), (8), and (9) and compute function $X_{(F_1F_2)}(dx, dy)$ in this region with a step of 1. As a result, coordinates of the integer shift of an image are found. The fractional part of this shift is estimated by

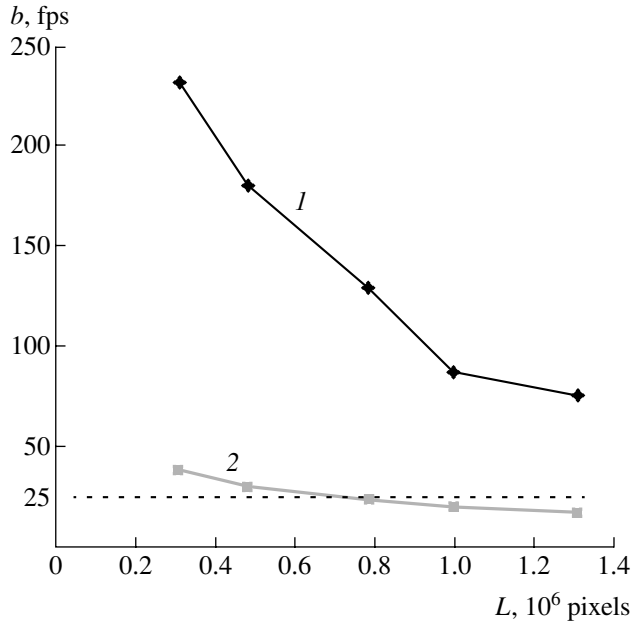

Fig. 1. Processing speed b vs. frame size L: (*1*) the operating speed of the proposed algorithms and (*2*) the speed of complete processing of the video sequence.

analogy with the algorithm used in the frame compression method.

6. COMPUTATION OF THE TOTAL VECTOR OF WARPING PARAMETERS

All warping parameters must be computed with the use of the estimates of the shifts of all correlation windows. It is reasonable to relate the computed estimates of local shifts to the centers of the corresponding windows. The warping vector is found according to the criterion

$$\hat{Q} = \arg\min_{Q} \sum_{k=1}^{N} (\vec{\rho}_k - \tau_Q(\vec{r}_k))\mathbf{C}_k^{-1}(\vec{\rho}_k - \tau_Q(\vec{r}_k))^T, \quad (14)$$

where $\vec{r}_k$ are the coordinates of the centers of windows in frame $F(t-1)$, $\vec{\rho}_k$ are the coordinates (computed from formulas (5) and (6)) of the same points in frame $F(t)$, N is the number of rectangles, $\mathbf{C}^{-1}$ is the covariance matrix of errors, $\tau_Q(\vec{r}_k)$ is the coordinate transformation function, and T designates the transposition operation.

Function $\tau_Q(\vec{r}_k)$ determines the frame-warping model: affine, projective, quadratic, or other transformations of coordinates. This circumstance implies that the minimization problem is solved in an n-dimensional space. For affine, projective, and other types of transformations, n equals, respectively, 6, 8, etc.

In the case of affine transformations, the problem involves solution of a set of linear equations. With introduced regularization [6], the corresponding set of equations yields an acceptable approximate solution even when the errors in the estimated window shifts are large, i.e., in the cases of low-quality input images.

Furthermore, if this problem is solved with a Kalman filter, it is possible to achieve a specified accuracy via computation of the shift for an appropriate number of windows. When a frame contains objects moving relative to a background scene, these objects can be eliminated from computations through a standard exclusion processing method [7].

Results of investigation and comparison between one of the proposed algorithms and the main algorithm using correlation of the entire frame

Characteristics	Algorithm	
	correlation of the entire frame	proposed algorithm
Shift determination error, pixels	0.07	0.03
Maximum shift, % of frame size	25	50
Angle determination error, deg	–	0.034
Maximum angle, deg	–	10
Scaling-factor determination error, %	–	0.2
Number of frames (320 × 240) per second	0.9	715
Probability of failure, %	0.3	<0.1
Maximum frame size processed in real time	–	1600 × 900

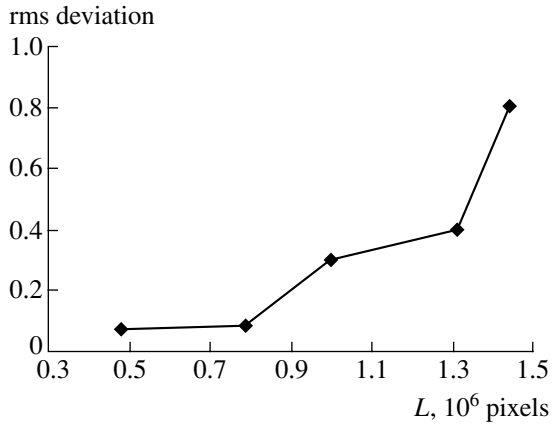

Fig. 2. Root-mean-square deviation of the error in the determined shift vs. frame size L in the case of a limited processing time for each frame.

The above problem can also be solved with the nonlinear method:

$$\hat{Q} = \arg\min_{Q} \sum_{k=1}^{N} X_k(\tau_Q(\vec{r}_k) - \vec{\rho}_k), \qquad (15)$$

where $X_k(\tau_Q(\vec{r}_k) - \vec{\rho}_k)$ are the correlation functions that are precomputed for each window.

7. ANALYSIS OF THE PROPOSED ALGORITHM

The described algorithms were investigated on a PC with a 2.4-GHz Intel® Pentium® 4 central processor, 512 MB RAM, an NVIDIA GeForce4® MX 440 video adapter, and the Microsoft® Windows™ 2000 operating system.

The accuracy characteristics and speed of the algorithms were tested for synthetic sequences with a frame size of 320×240 pixels. Analysis of the accuracy characteristics revealed that the spreads in shifts and in rotation angles were from –30 to +30 pixels and from –2.5° to +2.5°, respectively. The maximum values of these parameters corresponded to a twofold lowering of accuracy.

Efficiency of one of the proposed algorithms was investigated. The results of investigation and comparison with the basic-shift-determination algorithm that uses correlation of the entire frame are presented in the table. It is seen that the proposed algorithm has a noticeable advantage in the computation speed and in the maximum determinable shift. The processing speed of the algorithms is shown in Fig. 1 as a function of the frame size. It is seen that the algorithm operates in super-real time (75 frames per second) even when the frame size is 1.3×10^6 pixels.

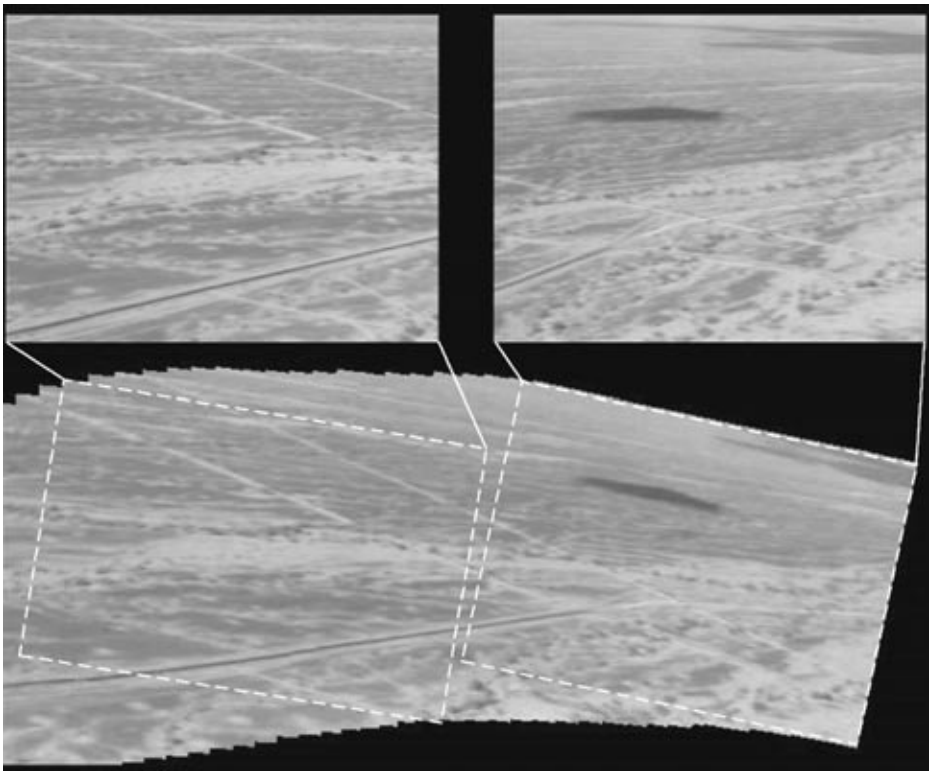

Fig. 3. An example of construction of a mosaic: (top) two frames of the initial video sequence and (bottom) the mosaic image constructed from two frames with the use of the computed warping parameters.

The proposed algorithm has the capability to process real video sequences of different types and characters (more than a hundred various sequences were tested) and yields good results. In this case, it is not easy to establish stringent numerical performance criteria, because an exact value of the shift of the frames of real video sequences is unknown. Hence, the quality was tested visually and the probability of error perceptible by eye was selected as the main criterion.

In addition, the processing speed was investigated in the case when the real time was limited, taking into account the duration of video-sequence recording and displaying.

To ensure real-time operation, an algorithm must be capable of interrupting computations if a specified time is exceeded. In this case, it is possible to obtain intermediate results representing a rough estimate of the desired parameters. The rms deviation of the error in the determined shift is shown in Fig. 2 as a function of the frame size. It follows from the results of investigations that the proposed algorithms ensure an acceptable quality of real-time processing (25 frames per second) of video sequences with a frame size of up to 1.4×10^6 pixels.

To illustrate the capabilities of these algorithms, the mosaic constructed from an input video sequence of extremely low quality (a low-information-density initial scene, a blurred input image with a noticeable level of noise, etc.) is shown in Fig. 3. The mosaic was constructed as follows. For each frame, the parameters of warpings occurring relative to a reference frame were computed. Next, with the use of these parameters, all frames were combined into a single large frame.

CONCLUSIONS

The algorithms based on the correlation approach have been created for estimating the parameters of spatial warping of video frames. Owing to a detailed analysis of the behavior of the correlation function used in these algorithms, not only the efficiency of the correlation methods but also their reliability and speed were improved.

The developed algorithms demonstrate robust operation under a wide range of shooting conditions: different models of spatial warping, the influence of moving objects, different noise models, etc.

It is shown that the proposed algorithms provide high-accuracy estimates of the warping parameters (similarly to other known algorithms). At the same time, some of their characteristics, such as the maximum estimated shift and the processing speed, are superior to the corresponding parameters of the known algorithms.

REFERENCES

1. S. Negahdaripour and S. Lee, in *IEEE Workshop Visual Motion, Princeton, N. J., October 1991* (IEEE, Piscataway, N.J., 1991), p. 132.
2. R. Chipolla, Y. Okamoto, and Y. Kuno, in *Proc. 4th Int. Conf. Computer Vision, Berlin, May 1993* (IEEE, New York, 1993), p. 374.
3. F. Lustman, O. D. Faugeras, and G. Toscani, in *Proc. 1st Int. Conf. Computer Vision, London, 1987* (IEEE, New York, 1987), p. 25.
4. B. K. P. Horn and E. J. Weldon, Jr., Int. J. Computer Vision **2** (1), 51 (1988).
5. A. K. Kim, A. E. Kolessa, V. N. Lagutkin, et al., Radiotekhnika, No. 12, 39 (1998).
6. J. E. Dennis and R. Schnabel, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations* (Prentice Hall, Englewood Cliffs, N. J., 1983; Mir, Moscow, 1988).
7. D. A. Forsyth, and J. Ponce, *Computer Vision: A Modern Approach* (Upper Saddle River, N.J., 2003; Vil'yams, Moscow, 2004).