New Motion Prediction Algorithm Invariant to Rotation and Occlusion

José Tomás Tello[1], Oleg Starostenko[1], Gennadiy N. Burlak[2]

[1] Computer Science Department, University de las Américas, Cholula, Puebla, 72820, Mexico
oldwall@mail.udlap.mx
[2] Center for Research on Engineering and Applied Sciences, Autonomous State University of Morelos, Cuernavaca, Mor. Mexico
gburlak@uaem.mx

Abstract. The novel algorithm for motion estimation and prediction of objects in dynamic scene by processing a limited sequence of images is presented in this paper. The proposed hybrid algorithm is based on computing differences between consecutive frames of video stream for fast detection of objects in motion by processing of their time-varying edges. For quantitative prediction of motion and computing of its characteristics, the interpolation technique has been used in combination with the motion stabilization technique useful for analysis of irregular displacement of objects. The invariance to occlusion and rotation of object in motion is achieved by proposed double envelope approach. The principal goal of proposed algorithm is the development of simple and efficient facilities for motion prediction when the possible routes may be computed even the objects in motion has occlusions and rotation. The introduced algorithm has been tested and evaluated.

1 Introduction

An artificial vision system for motion detection and prediction consists of three basic subsystems: the preprocessing of digital images (segmentation and description), the patterns (objects) recognition and persecution, and prediction (computing the next possible state of the traced object within the scene). A key module for analysis of translation of rigid objects in sequences of images is pursuance unit which principal problem is the quantity of the processed data due to necessity of bit by bit image inspection for detecting the changes in the position of objects. The input of a system for analysis of dynamic scene is a frame sequence where each frame represents the scene at time *t*. It assumes that image is obtained using camera located at the origin of a three-dimensional coordinate system. The projection used in this observer-centered system may be either perspective or orthogonal. It is also assumed that since the frames are usually taken at regular intervals, *t* represents the *t-th* frame of the sequence, rather than the frame taken at absolute time *t*. Any perceptible by eyes motion in a scene consists of some changes in the frame sequence within video stream. Motion characteristics can be analyzed if such changes are detected. A good quantitative

© A. Gelbukh, H. Calvo (Eds.)
Avances en la ciencia de la computación en México
Research on Computing Science 13, 2005, pp. 23-33

estimation of motion components of an object may be obtained if it is restricted to a parallel motion to the image plane.

The sequence of images contains necessary information about dynamic scene and usually it is defined by optical flow or motion field, which can be estimated by well-known methods [1], [2]. The obtained information about motion can be used as input of different subsequent processes including motion detection, motion compensation, motion-based data compression, 3-D scene reconstruction, autonomous navigation, analysis of dynamic processes in scientific applications [3], [4]. Normally the optic flow and the motion field (object displacement vector) are different, but they are based on the similar motion characteristics and with certain approximation their quantitative properties can be equal [5]. Thus, the optic flow equation is used for computing the motion field, and the motion characteristics are used for optic flow construction. Usually, the computing the motion characteristics can be obtained more quickly on base of processing the object edges or principal corners instead of analysis of intensity variations or complete object correlation in consecutive frames [6].

Sometimes the quantitative estimation of motion characteristics is provided by block correspondence techniques where estimation of the best similarity of features or regions within consecutive frames at high level (configuration, regions analysis) or low level (corners, patterns, color changing analysis) is obtained. In these approaches the coordinates (x', y') of the center of analyzed pattern are found after pattern translation from coordinates (x, y) during the time interval Δt [7]. Thus, the same pattern is used within consecutive images as reference one. It allows overcoming the problem of progressive increment of compared patterns but aggregates accumulative error proportional to time function. This process can be modeled by the brightness function of analyzed pattern calculated according the equation:

$$C(x,y,x',y') = \sum_{-M/2}^{M/2} \sum_{-L/2}^{L/2} \{E(x+m, y+l, t-\Delta t) - E(x'+m, y'+l, t-\Delta t) \tag{1}$$

where $E(x, y, t)$ is the brightness value of the pattern at the time t, M and L are two dimensions of the pattern toward the axis x and y respectively. Computing the motion characteristics is provided by estimation of the displacement vector F (motion field vector) during comparison of the similar $C(x, y, x', y')$ functions for analyzed and reference patterns as it shown in Fig. 1.

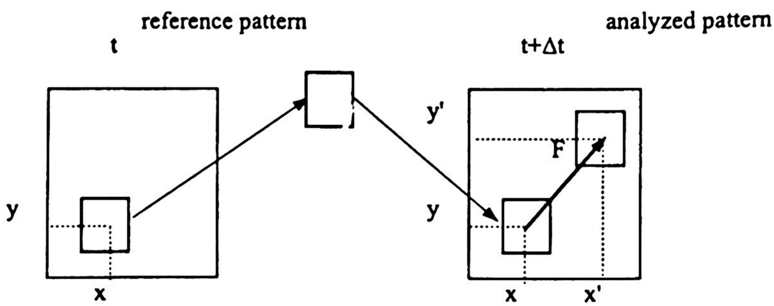

Fig. 1. Blocks correspondence technique for dynamic scene analysis.

On the base of absolute value of F and time interval Δt between two consecutive images the speed of object translation is calculated. The disadvantage of this technique is dependence on the dimensions of pattern in motion [8].

To predict the motion of dynamic objects the additional techniques must be used. These techniques are based on computing the optic flow or motion fields where the tracing and prediction of objects in the consecutive frames may be described by mathematical equations. The powerful technique is interpolating approach which provides high efficient and fast description of well-defined trajectories. In case of irregular motion frequently occurred in dynamic scenes the motion compensation or stabilization technique must be applied. It is important to mention that proposed motion prediction approach must operate efficiently when occlusion or rotation of objects in motion is appeared.

2 Motion Prediction Techniques

There are a lot of reports about well-known methods for border estimation which can be classified as it follows: approaches based of computing the probability of intensity function of objects in motion [9], gradient-based methods [1], block correspondence techniques [7], spatial-temporal volumetric filtering methods of the 3D images [4], phase-based approach applying the speed sintonization filters [3], USAN circular mask area manipulation technique [10], neural networks based approach [11]. But the border detection techniques require a long process of pixel by pixel analysis for each frame and then recognition of the detected borders, forms, etc.

The fastest and most obvious method for detecting changes between two frames consists in comparison the corresponding pixels of two frames to determine whether they are similar. In this approach, a binary difference picture $DPjk(x, y)$ between frames $F(x, y, j)$ and $F(x, y, k)$ is obtained by applying the following equation

$$DPjk(x,y) = \begin{cases} 1, & \text{if} \quad |F(x,y,j) - F(x,y,k)| \rangle \tau \\ 0, & \text{in other cases} \end{cases} \tag{2}$$

where τ is a threshold. The difference and accumulative difference presents the well-interpreted regions with changes usually due to motion of objects. This approach is very sensitive and permits to detect very small displacement of object. The disadvantage is dependence on correct selection of threshold τ.

There are some efficient motion prediction approaches which may be mentioned as

a) In the interpolation approach the motion predictor tracks the screen space location of points on an object. Each object in the scene has associated transformation matrix, which transforms object from model space to screen space in frame [4]. During the initial rendering of the image the matrix transformation number of the object visible in each pixel is stored - each object has a unique transformation number. If more than one object is visible in the pixel, for example in the case of transparent objects, then it would be possible to store multiple transformation numbers with the relative contributions of each object to the total pixel intensity. However, the actual implementation of the algorithm stores only the transformation number of the front most object visible in the pixel.

b) The Brownian motion model is applied in the statistical approach where the Bayesian filters for the tracking of object with easy prediction of motion are useful. The Brownian model is extremely conservative and does not attempt to model the dynamics of irregular motion [12]. This is often a poor estimate of the actual distribution over a possible position because objects with easy predicted displacement do not move randomly.

c) The Kalman's filter is the standard technique using prediction to improve state estimation over the time. They have been successfully used to track, for example, the boundaries of hands in image sequences. However, these filters need preprocessing, or data association, to determine which measurements in the image should be used to update the motion model. The filter is distinguished for its ability to predict the state of a model in the past, present and future, still when the precise nature of the modeling system is ignored.

d) Multi-hypothesis motion compensated prediction predicts a block from a weighted sum of multiple reference blocks in the frame buffer. By efficiently combining these reference blocks, it can provide less prediction errors so as to reduce the coding bit rate [13]. Usually long-term memory motion compensation generates only one motion vector for each macro-blocks, so this approach allows more than one motion vector.

After analysis of efficient well-known methods the hybrid approach for motion prediction is introduced, particularly the proposed algorithm is developed taking into account advantages of the interpolation technique, block correspondence approach for tracing of objects in motion invariant to occlusion and rotation, and motion compensation method for irregular motion stabilization.

3 Proposed Algorithm for Motion Prediction

The classical but enough efficient approach of motion prediction is based on using the interpolation technique when expected values are computed by applying the function derived on base of known values. Conceptually, the interpolation process has two stages: fit an interpolating function to known data and evaluate the errors of prediction using interpolating function at new target. The proposed algorithm for fast motion prediction based on application of subtraction operator and Lagrange polynomial interpolation may be described as it follows:

a) The input images of normalized size are the frames of video streams to be converted to gray scale images.

b) The detection of objects in motion using limited sequence of images is obtained by analysis of changes $DPjk(x,y)$ between consecutive images $F(x,y,j)$ and $F(x,y,k)$ according the equation *(2)* where threshold τ is selected with respect to necessary quantity of segments representing the object in motion.

c) The objects in motion detected by subtraction operator on step b) are wrapped up by circular envelope. It is used as description of the same object in the following frame without taking into account the small variations of detected objects in motion. Using the envelope instead of complete borders of object reduces the errors of subtraction operator and simplifies the manipulation with objects with occlusions.

d) The center of circle (envelope) for object in motion is computed. The circle center represents the gravity center of the object and it is used as coordinates of object within the image. If there is more than one object in motion, each one is represented by center of its envelope and the general overall envelope for all these object is defined. The overall envelope is used on step g) if the occlusions of the objects with other ones in motion are detected. The occlusion is automatically detected when the size of overall envelope is less than the double size of envelopes representing the objects in motion.

e) The generation of interpolating polynomial is provided on base of position of envelopes corresponding to the same object in video stream. The interpolating polynomial of degree *N-1* for the *N* points (envelope coordinates) $y1 = f(x1)$, $y2 = f(x2), \ldots, yN = f(xN)$ is given by Lagrange classical equation:

$$P(x) = y_1 Ln,l(x) + y_n Ln,n(x) = \sum_{k=0}^{n} y_k Ln,k(x) \tag{3}$$

f) The extrapolated predicted object position $P(x+1)$ is computed on base of results of previous step. In this way the accuracy of predicted and real translation is evaluated computing their absolute and relative errors.

g) In case of occlusions between envelopes the additional analysis of the object in motion is applied. This analysis is based on detection of the object borders and their continuous tracking using the Sobel spatial gradient operator [14]. Obtained borders in motion are traced for detection of the displacement vector according block correspondence approach as it shown in Fig.1 and the Segment and Neighbors Matching method proposed by authors in [15]. This method is based on the concept of fuzzy sets and operates with membership grade as principal criteria for including the analyzed element to the set of fuzzy segments which compose the pattern. This step permits to reduce the number of similarly oriented segments, increment the speed of tracking the objects in motion, and predict their next position in case of occlusion with other ones. For evaluation of the proposed algorithm some applications have been designed and tested.

4 Results and Discussion

According to proposed algorithm the objects which manifest motion are detected by clustering the areas tracked them in the sequence of frames in video stream. Some experiments for estimation of efficiency of proposed algorithm have been done taking into account that the video stream is captured by a single fixed camera and computing the characteristics of relative motion between objects is not provided. The Fig. 2a) shows the displacement of two objects and the detecting them by computing the difference of two superimposed frames. Also an envelopes of two objects and overall envelope as the result of clustering the area with motion are shown.

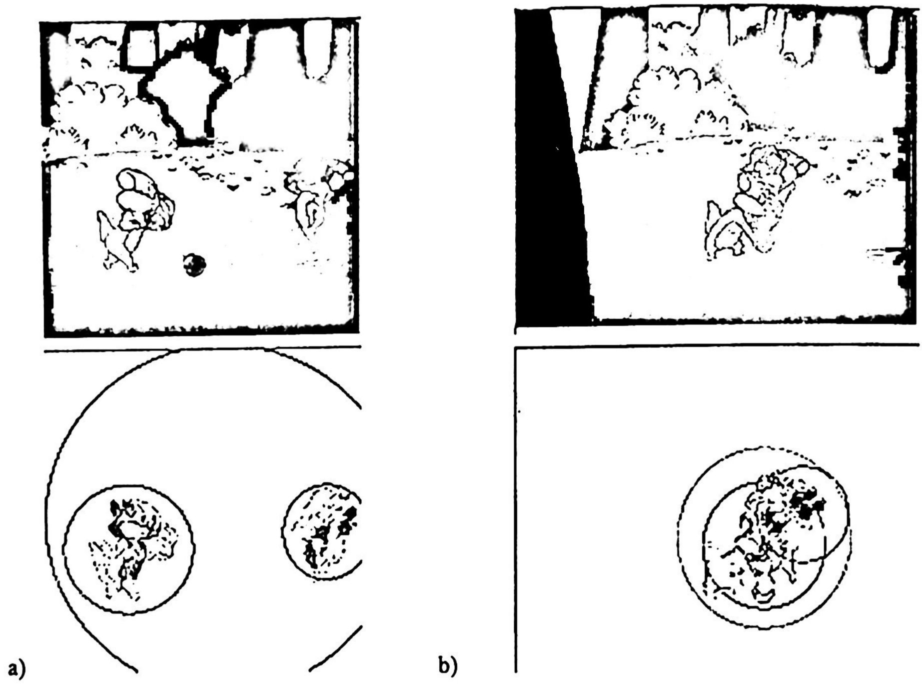

Fig. 2. Detection of object in motion a) the objects in motion without occlusion are presented by separate envelope. b) the objects in motion with occlusion are detected by intersection of envelopes. The overall envelope shows the area with objects in motion.

A small circle in Fig.2 represents the envelope and rounds the area where there is an object in motion. Their center is defined by the coordinates computed as average value of coordinates of the segments in motion. In Fig. 2b) the overall envelope has dimensions about the size of envelopes of single objects. In this case the step g) of the proposed algorithm is applied.

If the periodical motion is detected the prediction becomes effective when at least one period of oscillation is described by interpolation function. The irregular motion can not be predicted with high accuracy because it is impossible to generate exact interpolation function. But anywhere, the greater number of processed frames permits to reduce the number of predicted positions of object in next step. In this case it is possible to talk about so-called motion stabilization function. The stabilization heuristic function is obtained as a set of possible routes of the object in motion when the error between the predicted and described by interpolating function displacement vector goes to zero. Taking into account interpolating function the probability of each route is computed and more probable one is selected. This algorithm implements the dynamic prediction because the next object position depends on particular instant of time and the result of application of interpolation and stabilization functions at this moment.

In the Fig. 3 the stabilization of the interpolation process is shown. The white line presents the direction of object displacement described by interpolating polynomial, the black one shows the direction of predicted motion computed on base of analysis of real position of the object in previous frames. The white circle defines the position of the object in predicted step.

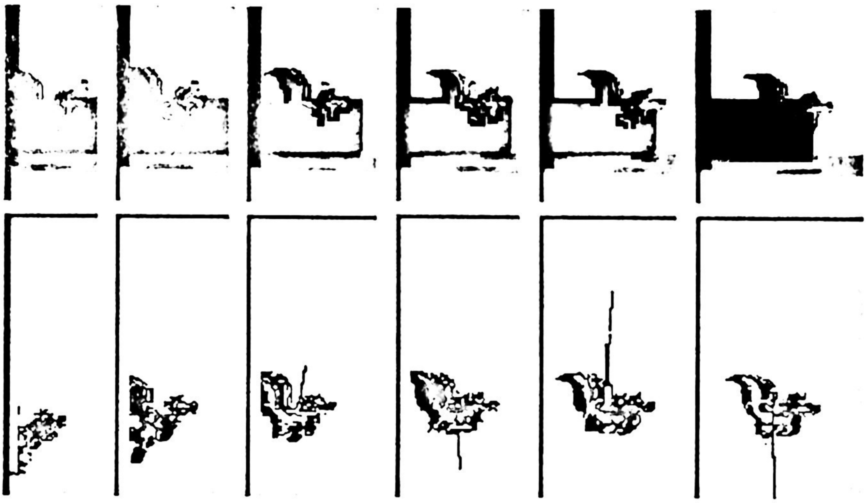

Fig. 3. Oscillation of the interpolating polynomial until reaching the stabilization.

The stabilization is obtained when the predicted and computed by interpolating polynomial displacement vectors have the similar direction with the certain error, for example, the displacement is less than 45°. The last frame shows the fact that predicted and computed displacement vectors have the similar direction. The error between them is reduced by processing more frames taken for analysis and more terms in interpolating polynomial. That minimizes the number of possible routes on the next step.

In the dynamic scenes with quasi-lineal object displacement as it shown in Fig. 4, the motion description is predictable with smaller number of terms in interpolating polynomial and motion stabilization is obtained after first two frames.

The development of proposed hybrid algorithm for motion prediction has the principal objective to apply the advantages of some high-performance methods. From the experimental data it is clear that the interpolating approach provides high efficient and fast description of well-described trajectories. The motion compensation or stabilization technique is applied in case of irregular motion frequently occurred in dynamic scenes that reduce the error of the predicted route. It has been proved that high accuracy object tracing is provided fast enough by using the block correspondence method that permits the prediction of motion in case of occlusion of objects and the motion invariant to their rotation. This is the principal contribution of our report.

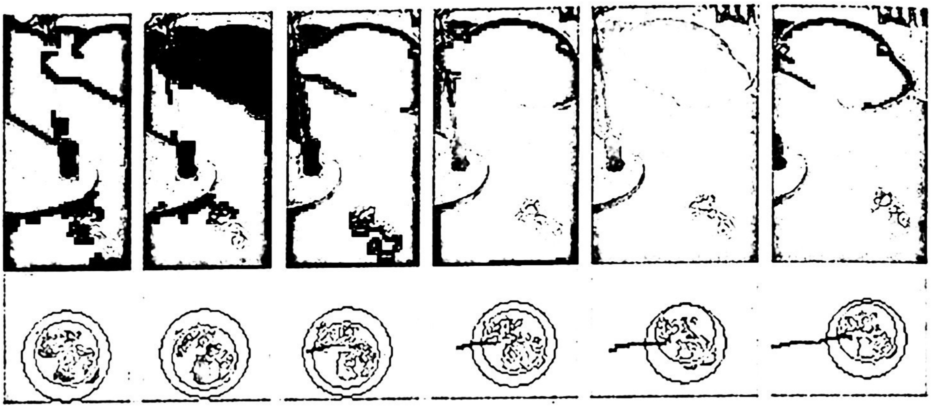

Fig. 4. Description of quasi-linear motion. The real and predicted motion is presented by black and white lines. The black line on the left part of processed image shows real accumulative translation from the initial point. The white line shows the real translation with respect to previous frame, and the black line on the right of the white line presents the direction and the distance of the predicted translation.

In order to evaluate the efficiency of the proposed approach the analysis of predicted by Lagrange extrapolation and real translations of object are compared as it shown in table 1.

Table 1 Analysis of errors for real and predicted motion. The row in bold letter with real coordinates $x=100$, $y=128$ shows the lowest error due to the lowest speed of object in motion.

x_real	y_rea	x_pred	y_pred	Δx	Δy	δy
20	82	20	61	0	21	0.3442
40	108	40	92	0	16	0.1739
60	125	60	113	0	12	0.1061
80	132	80	125	0	7	0.056
100	**128**	**100**	**127**	**0**	**1**	**0.0078**
120	116	120	119	0	3	0.0252
140	93	140	101	0	8	0.0792

The real x and y coordinates of object within each frame are presented in columns x-real and y-real. The relative predicted positions of object are computed and then adjusted to absolute coordinates as it shown in columns x_pred and y_pred. These values now may be compared, therefore the absolute additive errors Δx = $|x\_real$ —

$x\_pred|$, $\Delta y = |y\_real - y\_pred|$ (column Δx and Δy) and relative error $\delta y = (y\_real - y\_pred) / y\_real$ (column δy) are computed. The image corresponding to motion characteristics presented in table 1 is shown in Fig. 5 where the parabolic motion of the shot ball with initial acceleration has been analyzed.

It is important to mention that the absolute and relative errors of object's translation along x axis are equal to zero because the x component of motion is uniform due to constant speed. But the y component of motion has acceleration and the predicting model produces the errors of value proportional to magnitude of the acceleration. This problem may be solved by more frequent sampling of frames during intervals with acceleration of high magnitude. In the Fig. 5 the positions of object corresponding to its real positions in two consecutive frames are shown on one processed frame as two white balls. The overall black circle shows the area where there are objects on motion in two consecutive frames and small one shows the possible position of object in next the frame. It may be mentioned that the algorithm provide exact prediction of position (the direction of the real and predicted displacement practically are matched) due to the route is well-defined function.

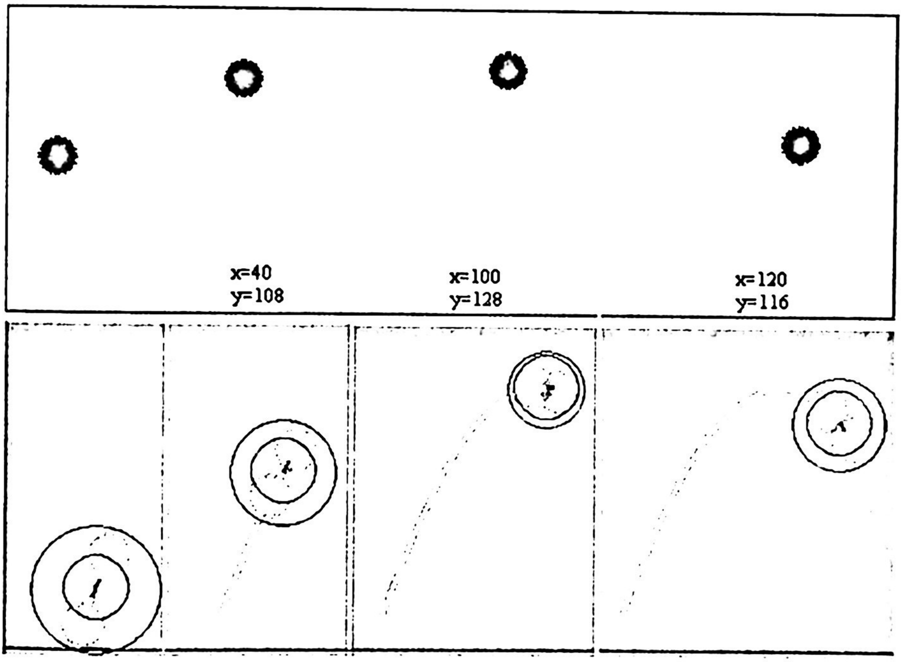

Fig. 5. Example of the parabolic motion of object and its some predicted and real position which have been presented in the Table 1. The thin white line on the left of the object shows accumulative real displacement. The thick white line presents the direction of real translation with respect to previous frame and the black thick line is direction of predicted displacement.

5 Conclusions

This paper presents the early research results in development of motion analysis and prediction facilities. There are a lot of techniques for motion prediction but they are not so efficient and fast enough for real-time applications which require the great quantity of data preprocessing of video steams. The proposed algorithm is one of possible solution for estimating and predicting the motion of objects within dynamic scenes based on interpolation of their routes.

For reduction of processed data the small quantity of the segments represented the object in motion are processed by tracking them using double envelopes useful for route construction and prediction. For exact prediction of irregular motion the motion stabilization for interpolated displacement vectors are proposed. The advantage of manipulation with motion stabilization function permits to apply same approach to irregular motion as it is provided for periodical of well-described translation. The disadvantage of approach for motion stabilization is a long time period of computing the new possible routes until the difference in direction between predicted and described by interpolating function displacement vectors reaches the minimum value.

Moreover, the proposed algorithm solves the problem of object tracking and their motion prediction when the occlusion between objects is presented. The processing objects borders is applied using the Block correspondence and Segment and Neighbors Matching method which permit exactly to trace the segments on base of fuzzy sets. It takes additional time but this method will be used only during the occlusions. The motion without occlusion is processed by analysis of envelope translation proposed for representation of object without taking into account small variations of it form or color. The use of envelopes provides simple and fast detection and tracing of object in motion with rotation. The principal restriction of proposed approach is that the computing the characteristics of relative motion of objects is not provided.

Preliminary experiments with proposed algorithm show that the faster motion prediction is achieved due to the lower quantity of processed data taking into account irregular motion, possible objects rotation and occlusion.

In the future works a generalization of the interpolation procedure will be made. It is important task to evaluate and compare the obtained results with some efficient approaches such as the Kalman's filter and the probabilistic correspondence of blocks. The proposed algorithm is the one of possible solutions of still open problem of motion estimation and prediction

References

1. Fleet, D.: Measurement of Image Velocity, Kluwer Academic Publishers, Massachusetts, (1992)
2. Kollios, G., Tsotras, V.: Indexing animated objects using spatiotemporal access methods, IEEE Trans. on Knowledge and Data Engineering, Vol. 13. 5 (2001) 758-777
3. Laplante, F.: Real-time imaging. Theory, Techniques, and Applications. IEEE Press, USA, (1996)

4. Chung H.Y. et al.: Fast Motion Estimation with Search Center Prediction, University of Hong Kong. www.Eee.hku.hk (2004)
5. Cheung, C.: Fast motion estimation techniques for video compression, Hong-Kong University, www.cityu.edu.hk (1998)
6. Jahne, B. Haußecker, H.: Computer Vision and Applications, Academic Press, USA (2000)
7. Lagendijk, I.: Motion estimation. information and communication theory group, Delft University of Technology, www-ict.its.tudelft.nl/ (1999)
8. Hong, T. Hongche, L.: Motion-model-based boundary extraction and a real-time implementation, Comp. Vision and Image Understanding Vol. 70. 1 (1998) 87-100
9. Hongche, L. Hong, T.: Accuracy vs efficiency trade-offs in optical flow algorithms, Comp. Vision and Image Understanding, Vol. 72. 3 (1998) 271-286
10. Smith, S. Brady, J.: SUSAN - a New Approach to Low Level Image Processing, Int. J. of Comp. Vision, Vol. 23. 1 (1997) 45-78
11. Lau, C.: Neural Networks, Theoretical Foundations and Analysis, IEEE Press, USA (1988)
12. Kevin, P. M.: Dynamic Bayesian Networks: Representation, Inference and Learning, Ph.D. Thesis. Computer Science. University of California, Berkeley (2002)
13. Brian, G.: Motion Compensated Noise Reduction. Graphics Research Group Microsoft Research, WA., USA (2004)
14. Sobel, J.: Machine vision for three-dimensional scenes, Academic Press, New York (1990)
15. Starostenko, O. Neme, J.A. Automatic Complex Glyphs Recognition and Interpretation, International Journal IEICE Transaction, Japan, Vol. E82-A. 10 (1999) 2154-2160