International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-8 Issue-4, April 2019

Adaptation of the Machine Vision System to Environmental Conditions

Marija Olegovna Korljakova, Vladislav Olegovich Miloserdov

***Abstract*: The procedure for the formation of the frame processing algorithm training model in a machine vision system designed to solve the visual odometry issue has been considered. The main difficulties arising in matching the stereopair frames have been described. The scheme and the set of frame processing features have been proposed. The work of the teacher algorithm based on the neural network has been considered. The options of the algorithm adaptation to an applied task have been analyzed. The options to implement the system and the results of computational and field experiments have been presented.**

***Index Terms*: learning from examples, machine vision, neural networks.**

I. INTRODUCTION

Modern autonomous mobile systems should be able to navigate in a changing external environment. The vision is the most natural means of orientation in an environment intended for the man, and in the case of artificial systems, they include visual odometry, navigation, and pattern recognition. The stages of the video stream processing when solving the odometry issue in the camera stereopair are as follows:

- receiving the stereopair frames at the moment of time $t(i)$,
- preprocessing and selection of conjugate areas or points belonging to images of the same object of the scene,
- obtaining the XYZ coordinates of the scene objects from the stereo reconstruction procedure [1], [2], and
- comparing these coordinates with the frame analysis results at the moment $t(i-1)$ and calculating the displacement of the machine vision system (MVS) associated with the mobile system.

The video stream processing depends on external conditions: time of day, displacement speed, scene composition, etc. The basic algorithm involves the search for identical objects in the left-right frame pair to determine the coordinates of the scene points and in successive pairs of frames to determine the MVS displacement and speed. The difficulty of finding conjugate areas for the left and right frames of a stereopair is due to the fact that real sensors react differently to the same scene. For example, one of the stereopair frames may be brighter or darker, and the MVS shift can change this ratio of frames in any direction. The image processing algorithm should be changed in accordance with the changes in the external environment.

Manuscript published on 30 April 2019.
***** Correspondence Author (s)
Marija Olegovna Korljakova, Bauman Moscow State Technical University (Kaluga branch), Kaluga, Russia.
Vladislav Olegovich Miloserdov, Scientific Production Company "Turbocon", Kaluga, Russia.

II. PROPOSED METHODOLOGY

A. Comparison of The MVS Frames

The article considers in more detail the features of matching the MVS frames in time and in the stereopair itself. The main computational complexity of the frame processing falls on matching between the left and right frames and matching the frame areas in time. The MVS frame matching scheme is shown in Figure 1, where points 1, 2 are the stereo-conjugate points at the moment $t(i-1)$, 3, 4 **are** their pairs at the moment $t(i)$, $M(t(i))$ **is**the mathematical expectation and $D(t(i))$ is the dispersion of the possible position of the paired points at an unknown speed, $M'(t(i))$ is the mathematical expectation, $D'(t(i))$ is the dispersion of the possible position of paired points at a known speed; gray lines connect the episode frame points conjugated by epipolar geometry (1 – 2 and 3 – 4) and video stream (1 – 3 and 2 – 4).

The matching of the left and right frames of a stereopair has limitations associated with the MVS geometric model, which allows any pixel of one camera to match a relatively small subset of pixels of the second camera's screen plane. These pixels are placed along the epipolar line of the second camera [2], [3], and using image correlation methods or descriptors comparison, it is possible to find among them the best match for any point [4],[5]. The use of epipolar geometry greatly simplifies the task of finding conjugate points but does not eliminate the need for each stereopair to find the minimum allowable number of such points. The adjustable parameter of such a search is the classification threshold (p_{kor}), which sets the rule for rejecting or accepting the found pair of points on the left and right frames.

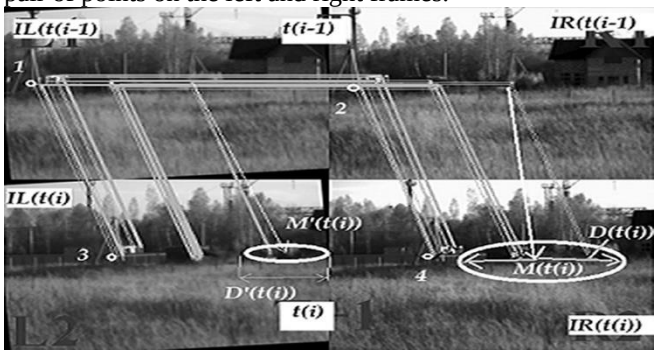


Fig. 1. An example of a video stream episode processing

Matching the frames for the $t(i)$ and $t(i-1)$ moments depends on the change of the MVS position, and in addition to the illumination change, the scene objects are shifted over the frame. With constant parameters of the frame processing algorithm, a different number of singular points is formed for successive frames, resulting in the loss of the video stream processing continuity. In addition, the matching of the frame point features at the moment $t(i)$ and $t(i-1)$ subject to an unknown direction of movement significantly increases the algorithm operation time. Knowledge of the direction of movement reduces the processing complexity. The known shift direction defines the area of the most likely shift in which to look for matching areas of conformity. The parameters (mathematical expectation $M(t(i))$ and variance $D(t(i))$) of this area should be adapted for the current speed and direction of motion. Thus, the list of custom parameters of the frame processing algorithm will be formed as follows:

$$Q =< p_{kor}, C_{por}, M(t(i)), D(t(i)) >.$$

The general scheme of the frame processing algorithm is as follows:

- search for special points based on the C_{por} threshold,
- C_{por} adjustment due to iterative selection,
- determination of conjugate points based on the correlation areas' comparison method using an iterative parameter selection p_{kor}, and
- search point shift by frame for the $t(i)$ moment relative to the $t(i-1)$ moment, with the use of $M(t(i))$beingequal to the coordinates in the previous frame and $D(t(i))$beingequal to 0.5 W, where W is the frame width in pixels.

A large variety of situations and changing environmental conditions are an important argument for the use of problem-solving training models in MVS, for example, searching for specific scene objects [6]-[8]. Neural networks will be used as a training model since this type of solvers has a large variety of architectural solutions and well-developed formation mechanisms for different types of tasks [8],[9]. This implies a description of the input and output feature set for the examples that will describe the relationship between the input and output characteristics.

B. Feature Space Formation

The features describing the video stream frames will be defined as follows. The special points should be detected by the joint analysis of the image gradient by image coordinates, and for the temporal analysis of frames, the time derivative of the video stream continuity condition is added [1]. Therefore, for the analysis, the signs built on the gradient for $I = \{I_L(t(i-1)),\ I_R(t(i-1)),\ I_L(t(i)),\ I_R(t(i))\}$ frames will be used, where I_L, $I_R$are the left and right frames, and $t(i-1)$, $t(i)$ are the consecutive moments of time of the video stream. It should be noted that all the frames in the search for specific points are subjected to the calculation of the image gradient, which allows not increasing the computational complexity of the algorithm.

As the initial input feature space, the following parameters will be used: the intensity dispersion, average frame intensity, the average value of the frame gradient, frame variance, etc. The length of the X input vector is 18 signs, and the Y output vector contains 4 adaptable parameters of the processing algorithm. Thus, a sample of 100 – 200 examples should be formed to train the model on real video sequences. Consider the examples generating algorithm.

C. Formation of The Solver Training Examples

The process of forming many examples is the teacher algorithm built on the basis of evaluation of the brute-force algorithm operation results. The initial sample capture was built for reference trajectories of several types such as indoor scenes, street scenes with different content. Since the $Y(X)$ is formed only for suitable video stream processing options, the selection of examples is reduced to the allocation of the most successful episodes in the video sequence. The selection criterion was determined on the basis of experiments and is a rule of the following type:

- if the calculated MVS displacement and the coordinates of the XYZ scene points correspond to the allowable range for at least half of the total number of points and the number of such points is $N > 100$,
- then the current fragment is taken as an example,
- otherwise, the fragment is rejected by the teacher algorithm.

The initial sample capture included 130 examples, for which a multilayer perceptron had been formed (a hidden layer with the number of neurons from 2 to 15). The analysis of the problem solution quality for test examples allowed determining the solver architecture in the amount of 12 neurons. The training results for the C_{por} parameter on the right (param 2) and left (param 1) stereopair frames are shown in Fig.2.

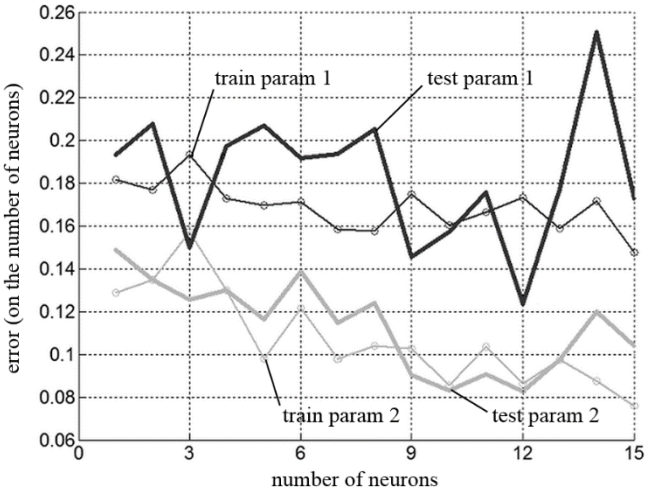


Figure 2. Evaluation of the training error and testing the neural network for the C_{por} parameter on the right (param 2) and left (param 1) frames of the stereopair.

The use of a network trained on the initial sample of examples reduces the frame processing time by 1 %. In addition, it is possible to improve the accuracy in scenes equivalent to the training sample capture for bench examples from 25 % to 11 % slip. The testing was carried out on the video stream sections in the amount of 7 – 20 frames in the context of the indoor stand.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication (BEIESP)

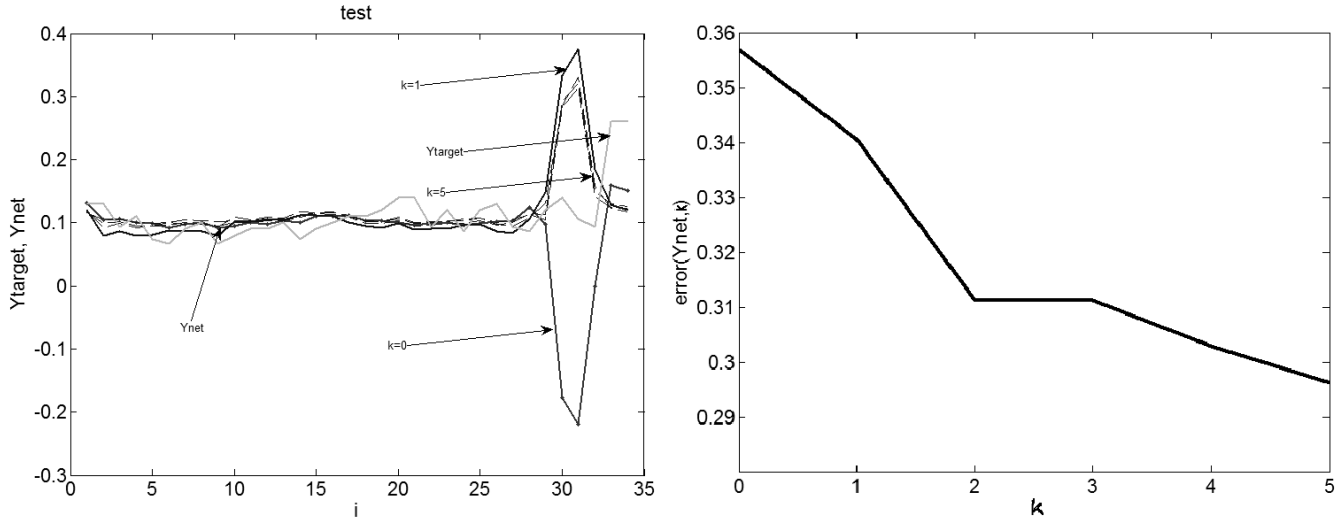
The generalization of the experimental results has shown that in the context of a stable scene, the stereopair movement demonstrates errors in determining displacement at 10 mm level for 100 mm of movement at a scene depth of 3 – 4 m. The given error in depth for scenes in the room is no more than 1 %. In the motionless observation of the scene, the displacement is estimated as zero, which is true.

However, the scene change leads to a change in the "good" levels of the parameters of the MVS frame processing algorithm. The sample capture extension and the inclusion of other scenes in it allow processing exactly those scenes that had been used in training. New scenes lead to a change in the values of the "good" levels of the algorithm adapted parameters. For example, for the C_{por} parameter (param 1 in Figure 2) the relative error changes from 0.16 to 0.3 in the network of 12 neurons.

III. RESULTS ANALYSIS

A. Adaptive Selection of The Frame Processing Algorithm Parameters

Adaptation to a specific problem is implemented by building a solver with an architecture that expands as it works (for example, ART networks, the associations of solvers, such as AdaBoost [9], [10], etc.) and by changing the solver with changing the current sample capture.



a) b)

Fig. 3. Dynamics of change in the solver output (a) and the error of the solver output (b) when additionally training on a new k-size sample capture

The first option leads to an increase in the size of the solver but allows solving complex problems without losing information. In the case of the MVS movement in the natural environment, the likelihood of the occurrence of the same conditions is not great, and therefore it is not necessary to store the previously obtained data. The plasticity property of the perceptron will be used and the solver will be additionally taught in case of loss of the video stream processing continuity on a sample capture formed by the teacher algorithm in the current scene. The dynamics of the solver's output change with the appearance of new examples are shown in Figure 3, where *k* is the sample capture size, where the additional training is performed, and *i* is the number of the test episode of new scenes.

The second element of the processing adaptation refers to the definition of the $M(t(i))$ and $D(t(i))$ parameters. It was necessary to determine the direction of the frame shift, for which an algorithm based on the Lucas-Canada method had been used, which, without training, allowed making a good prediction for $M(t(i))$ [4], and the variance size was determined by $D(t(i)) = max(2|M(t(i)) - M(t(i-1))|, 20)$.

The general processing scheme for video stream episodes upon the solver's additional training takes the following form:

- to take frames at the $t(i)$ moment and determine the images intensity gradients,
- to use the $t(i-1)$ moment information and form X, $Y(X)$,
- to calculate $M(t(i))$ and $D(t(i))$ for the $t(i)$ moment,
- to form, based on $M(t(i))$ and $D(t(i))$, the list of the conjugate points for the $t(i-1)$ and $t(i)$ moments, and to check its N length,
- if $N \leq 10$, to consider the video stream as lost and initiate additional training of the solver,
- if the continuity of the video stream is lost, the current episode is subjected to an iterative search for processing parameters and checked by the teacher algorithm that builds a new pair $\langle X, Y \rangle$, and
- if $N > 10$, to calculate the subsequent coordinates of the scene, determine the MVS displacement and proceed to the next episode.

The model testing has shown that the processing speed dropped for the first 50 – 100 episodes (the solver's additional training stage) and practically did not affect the speed of solving the problem for the following frames.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication (BEIESP)

The displacement estimation error for an unknown scene (street) under the condition that the frame processing algorithm is adapted is given in Table 1. Two modes were considered: driving at a speed of 20 km/h and stopping. The measured speed of the car's speedometer was 15 – 25 km/h, and the speed measured by the MVS was 21.6 km/h. In terms of quality, the MVS results corresponded to the indications of alternative sources of measurement. The error in estimating the displacement was 8 m. For comparison, the initial error of the visual odometry system in this area was 23 m.

Table 1. Model testing results

Parameters: scene depth: 100 – 300 m	Speed 20 km/h	Stop mode
Measured displacement, m	317	0.04
Evaluation of displacement, m	308.3	0
Absolute error, m	8.7	0.04
Relative error in depth, %	2.8 %	-

The second example – a stationary MVS demonstrates an almost accurate assessment of the state of the system (length of the section was one min., with 80 pairs of frames).

IV. CONCLUSION

The visual odometry system with an adaptive frame processing algorithm has made it possible to conduct a series of field experiments, which showed that the state of rest had been monitored almost perfectly with a natural change in illumination and movement of individual objects in the scene. Positioning error was no more than 1 % of the range.

The movement along the road is more difficult to estimate since there is no reference path. However, a comparison with GPS data, maps and ranging data has allowed to estimate the potential error of the model at the level of 5 – 10 % of the average range of objects in the scene in all areas of motion. Subsequent work will be aimed at the formation of the training model for comparing areas in the search for conjugate points of the scene.

REFERENCES

1. B. Yane, *Tsifrovayaobrabotkaizobrazheniy* [Digital image processing]. Moscow: Tekhnosfera, 2007.
2. M.O. Korljakova, and A.Yu. Pilipenko, Realizatsiyaneyrosetevoysistemytekhnicheskogozreniyadlyaresheniyazadachiopredeleniyaperemeshcheniya v sredemnogopotochnoyobrabotki [Implementation of the neural network vision system for solving the problem of determining the movement in a multithreaded processing environment]. *Abstracts of the NTK "Tekhnicheskoyezreniye v sistemakhupravleniya"*. Moscow, 2015.
3. E.A. Devyaterikov, and B.B. Mikhailov, Upravleniyedvizheniyemmobilnogorobota s ispolzovaniyemdannykhvizualnogoodometra [Motion control of a mobile robot using visual odometer data]. *Robotics and technical cybernetics, vol. 1*, 2013, pp. 22-26.
4. R. Woods, and R. Gonzalez, *Tsifrovayaobrabotkaizobrazheniy* [Digital image processing]. Moscow: Technosfera, 2005.
5. V.I. Syryamkin, and V.S. Shdidlovskiy, *Korrelyatsionno-ekstremalniyeradionavigatsionnyyesistemy* [Correlation-extreme radio navigation systems]. Tomsk: Tomsk Un-ty Publishing House, 2010.
6. J. J. Y. Torres, L. M. Bergasa, R. Arroyo, and A. Lazaro, Supervised learning and evaluation of kitti's cars detector with DPM. *2014 IEEE Intelligent Vehicles Symposium Proceedings*. Dearborn, MI, 2014, pp. 768-773.
7. S. Ozbay and E. Ercelebi, Automatic Vehicle Identification by Plate Recognition A Neural Network. *World Academy of Science, Engineering and Technology, vol. 9*, 2005, pp. 222-225.
8. P. Sermanet, andY. LeCun, Traffic sign recognition with multi-scale convolutional networks. *The 2011 International Joint Conference on Neural Network.*San Jose, CA, 2011, pp. 2809-2813.
9. S. Khaikin. *Neural networks: a full course*. 2nd ed.: Transl. from English. Moscow: LLC I.D. Williams, 2006.
10. M.Yu. Khomyakov, and N.L. Schegoleva, Sokrashcheniyevychislitelnoyslozhnostiklassifitsiruyushchikhalgoritmovsemeystva ADABOOST [Reducing the computational complexity of classifying algorithms of the ADABOOST family]. *News of higher educational institutions of Russia. Radioelectronika*, *vol. 4*, 2010, pp. 32-39.