Time Invariant Hand Gesture Recognition For Human-Computer Interaction

Dmitry Kostyrev[1], Sergey Anishchenko[2], and Mikhail Petrushan[2]

Southern Federal University Faculty of Physics, Zorge str. 5, Rostov-on-Don, Russia
dmitry.kostyrev@gmail.com
A.B. Kogan Research Institute for Neurocybernetics, Stachki av. 194/1, Rostov-on-Don, Russia

Abstract. Hand motion driven human-computer interface based on novel time-invariant gesture description is proposed. Description is represented as a sequence of overthreshold motion distribution histograms. Such description utilizes information about gesture spatial configuration and motion dynamics. K-nearest-neighbour classifier was trained on six gesture types. Application for remote slideshow control was developed based on the proposed algorithm.

Keywords: human-computer interfaces, hand motion tracking, dynamic pattern recognition

1 Introduction

Popularity of natural interfaces for desktop and mobile computer control has been rapidly growing within last decade. Nowadays common human-computer interfaces (like keyboard) are gradually replaced by natural control interfaces based on gesture-driven, voice-driven, finger or full body motion driven control. These new methods are widely used in entertainment applications or in such fields, where a contact between human and input device is impossible or unwanted because of sterility requirement or in case, when a device have to be controlled by a group of people simultaneously.

Hand motion recognition task is concerned with several fundamental computer vision problems, in particular with the problems of dynamic patterns detection and recognition. The standard pipeline for single image analysis is represented as the sequence of procedures: preprocessing – segmentation – classification. This pipeline is admissible for video analysis only if the task is to detect and classify the objects, which movements are not of value. Otherwise, additional information about object movement or transformation has to be considered. In contrast to single image, video contains such additional information, that have to be utilized for object detection and recognition.

The main goal of our project is to develop the robust descriptor for dynamic objects, invariant to object deformations and perspective transformations during a movement. In order to develop such descriptor the modifications of the standard single image analysis pipeline are proposed. The new dynamic gesture

recognition method is described below. It utilizes information about duration, direction and amplitude of a motion along with spatial and intensity-based feature descriptions of images in video sequence. This method was used to develop the human-computer interface and application for presentation remote control.

2 Research background

Gesture-based human-computer interfaces can utilize hand stationary configuration (configuration is relevant), like "open palm" or "thumb up", hand motion (dynamics is relevant), like "from palm to fist" motion, "hands up" motion, etc. A variety of gesture recognition algorithms was developed, that can be divided into two groups according to configuration or motion relevance:

1. single image analysis algorithms, that detect and recognize hand configuration in each frame of video;
2. image sequences analysis algorithms, that detects hand configuration changing pattern for whole gesture video sequence.

Single image analysis pipeline for gesture recognition is similar to commonly used analysis procedures: preprocessing, segmentation, classification. Following steps for gesture analysis were proposed [1]: hand contour extraction, tracking and recognition based on selected features. The image sequence analysis pipeline differs from single image approach and contains following steps: background subtraction, description of a gesture and classification. In spite of the fact that every gesture detection and recognition method is unique, combining different approaches and algorithms, there are some common steps used in most methods, such as background subtraction, features extraction and classification of a gesture.

Gesture detection and recognition methods use different background subtraction algorithms from very simple like frame difference [2], [3], [4] to more complex methods such as Adaptive Mixture of Gaussians [5][6] and frame difference enhanced with Gaussian filter [7].

Various feature extraction methods for gesture recognition have been described in [8], [5], [3], [7]. Methods based on calculation of histograms of oriented gradients (HOG) are used in [5] and [3]. In [3] it is used for features extraction from a motion history image which is created from several frames in sequence. Fourier analysis based methods are used for different kind of movements in [7]. Hand shapes within single image analysis approach can be described by shape context descriptor [8].

Gesture classification is performed by different algorithms based on fuzzy logic [8], SVM for large feature vector (3780 elements) [5], Euclidian distances [3], and spectrum analysis [7].

Each above-mentioned gesture analysis method has its own advantages and disadvantages. For instance, single image based analysis provides precise estimation of hand location in image but it is limited with fixed hand configuration because of non-rigid nature of a human hand. On the other hand image sequence

based methods are invariant to hand configuration in common but are dependend on gesture duration and completeness. Thus, gesture analysis methods have its limitations, for example, most of the methods based on skin color segmentation [9], [10], [11] or background subtraction methods [2], [3], [4], [5], [7] are highly dependend on scene light conditions, quality of camera sensor, etc. Along with recognition quality the method processing speed is one of the key factors in a sense of end user experience, so it has to be taken into account when comparing gesture recognition methods.

3 Gesture detection and recognition

The new method of a hand gestures detection and recognition is presented. Following requirements were set during problem formulation. According to them the method must be invariant to gesture duration, hand initial position and be able to detect and recognize transformable hand configurations. According to the requirements and research background this method should describe a gesture in terms of integral motion characteristics. Algorithm workflow scheme is presented in Fig. 1.

According to this scheme two main components can be highlighted in the workflow of this algorithm: background subtraction and features extraction.

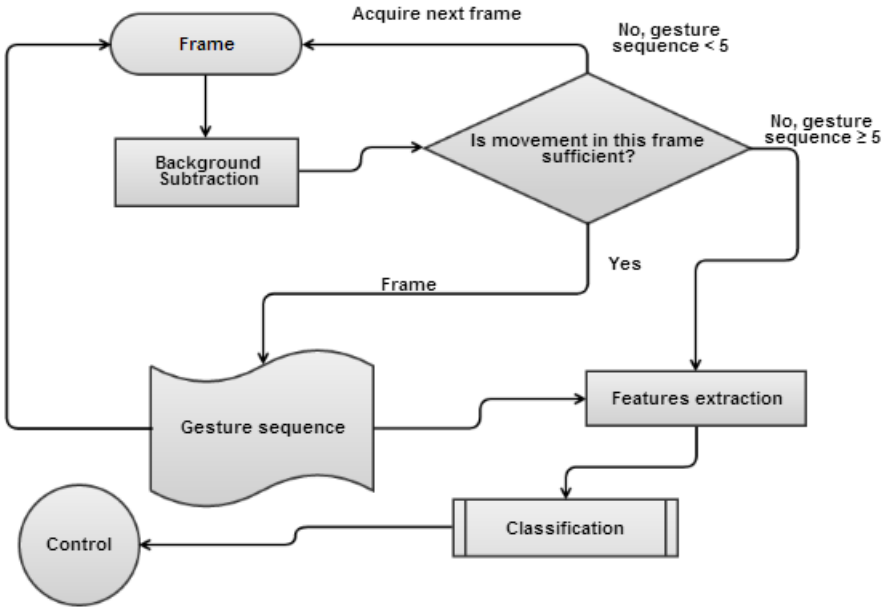


Fig. 1. The workflow of the algorithm

3.1 Background subtraction

Background subtraction is used for gesture duration estimation and as a preprocessing step for gesture description in our approach. Folowing requirements for background subtraction method were formulated:
- object contours estimation;
- no contour traces;
- realtime performance.

Several popular background subtraction methods were reviewed within the research. Each of them was evaluated according to the following parameters: performance, contours continuity, length of contour traces. Background subtraction quality was evaluated qualitatively and algorithm performance was estimated according to the video processing frame rate. Algorithm performance is considered "realtime" with framerate ≥ 30. All algorithms (except ViBe) were evaluated according to the current implementation in BGSLibrary [12] on 3th generation Intel Core i5 processor powered PC. Results of performance and quality estimations are presented in the Table 1.

Table 1. Overview of background subtraction methods

Algrorithm	Performance	Contours continuity	Length of traces
Frame difference	realtime	low	no
Moving mean	realtime	low	no
Adaptive Mixture of Gaussians	realtime	low	short
Adaptive Background Learning	offline	low	short
Gaussian Average	realtime	high	short
Multi Layer BGS	offline	low	no
Fuzzy Gaussian	realtime	high	long
Fussy Adaptive Som	offline	high	long
ViBe (serial implementation)	offline	high	no

Reviewed background subtraction methods are disbalanced in a sense of sufficient quality, performance and traces absence. Thus, a new background subtraction algorithm is presented which should combine performance and contours continuity. It is based on the computation of time-dependent intensity variance map between N frames as follows:

$$D(x,y) = \frac{\sum_i^N p_i(x,y)^2}{N} - \left(\frac{\sum_i^N p_i(x,y)}{N}\right)^2 \tag{1}$$

$$B(x,y) = \begin{cases} 255, if D(x,y) \geq Threshold^2 \\ 0, otherwise \end{cases} \tag{2}$$

where $D(x,y)$ - resulting variance map, p_i – i_{th} frame, N – number of frames in sequence, x, y – pixel coordinates, $Threshold$ – binarization threshold.

Presented background subtraction algorithm demonstrates realtime performance, sufficient quality of contour separation without long traces. However, this method is depended on light conditions. Parameter N regulates background model update speed. Two adjacent frames binarized variance maps with different N is presented in the Fig. 2.

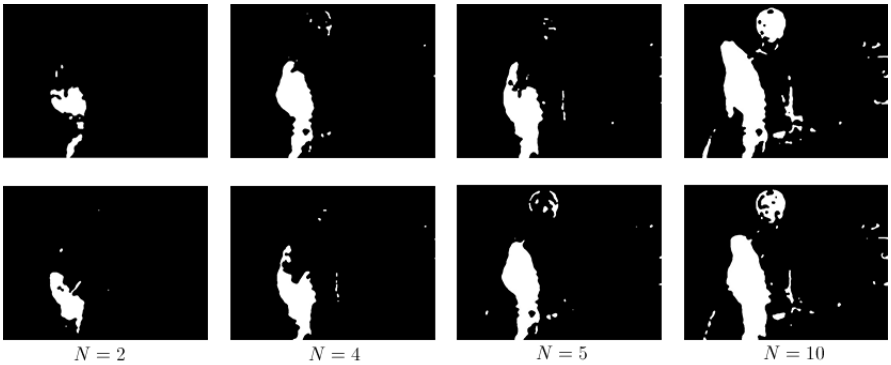


Fig. 2. Two adjacent frames binarized variance maps with different N

Binarized variance map B is prepared by median filtering for noise reduction and nearby contours merging.

Current frame motion rate is estimated by counting all non-zero elements in the binarized variance map. Motion is considered to be a gesture candidate when the following condition is met:

$$\sum_{x}^{w}\sum_{y}^{h} B(x,y) \geq w * h * MotionThreshold \tag{3}$$

where x, y — pixel coordinates in binarized variance map B, w, h – width and height of variance map accordingly, D – variance map, $MotionThreshold$ – threshold for estimating a gesture candidate in the frame.

If the current frame satisfies the above-mentioned condition it is being added to the current gesture sequence. Current gesture sequence is considered complete when number of frames in sequence $\geq Nmin$ and the condition (Eq. 3) was not met in the $Nmin + 1$ frame. As soon as the gesture sequence is considered complete it is being described and classified.

3.2 Gesture description and classification

Any gesture sequence, containing more than $Nmin$ frames, can be described by a feature vector. Motion Distribution Histogram is proposed to be the integral

description of the frame in gesture sequence. It describes a distribution of over-threshold variance (Eq. 3) over frame. Each motion distribution histogram is calculated according to the mean center of masses for all variance maps in the sequence, thus, we can achieve initial hand position invariance. Variance maps are being divided into 16 sectors in which all non-zero elements are counted and stored in corresponding element of motion distribution histogram. Sectors are numbered clockwise from horizontal axis orientation. Each element of motion distribution histogram is normalized according to the area of binarized variance map of the frame. Examples of variance maps and corresponding motion distribution histograms are shown in the Fig. 3.

Variance map and motion distribution histogram calculated for all frames in the sequence. Motion distribution histograms sequence containing N histograms are being divided into 4 subsequences according to the following intervals I:

$$I_1 = \left[1..\frac{N}{4}\right] I_2 = \left[\frac{N}{4}..\frac{N}{2}\right] I_3 = \left[\frac{N}{2}..\frac{3N}{4}\right] I_4 = \left[\frac{3N}{4}..N\right] \tag{4}$$

where N is the number of histograms in the sequence.

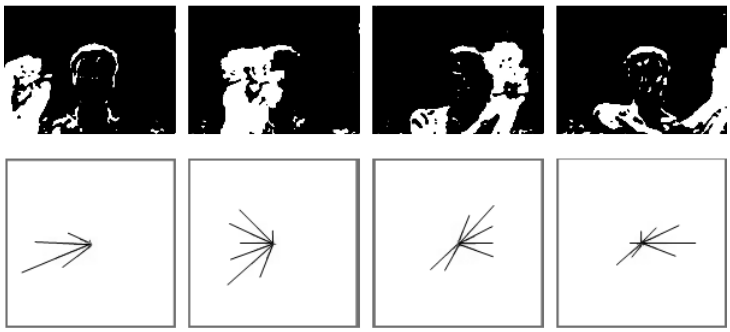


Fig. 3. Binarized variance maps for different frames of "from left to right" gesture sequence and corresponding motion distribution histograms

4 Experimental results

Proposed method has advantages and disadvantages comparing to the methods that are based on object detection and recognition in each frame of video. The main advantages of this approach are invariance to hand configuration transformations during the movement, invariance to the speed of the gesture and hand initial position in the camera field of view. However, this method is highly depended on background movement (which can be eliminated with depth map), distance between human and camera and lighting conditions.

In our approach we used $Nmin = 5$, $MotionThreshold$ (Eq. 3) equal to 0.08 for motion detection in close range from the camera (approximately 50 cm), $N = 2$ (Eq. 2) and $Threshold = 2$ (Eq. 2).

Six gesture types were selected for detection and recognition: hand movement from left to right, right to left, hand up, hand down, both hands from the center of the screen and both hands to the center of the screen. Training set containing samples of each gesture type was collected. The set containing seven examples of 2 classes (hands down and hands up) are displayed in the Fig. 4, where hand down movement is visualized with black colour, and hands up movement is visualized with gray colour.

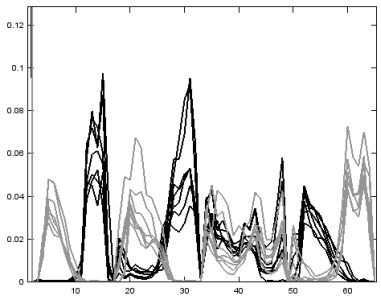
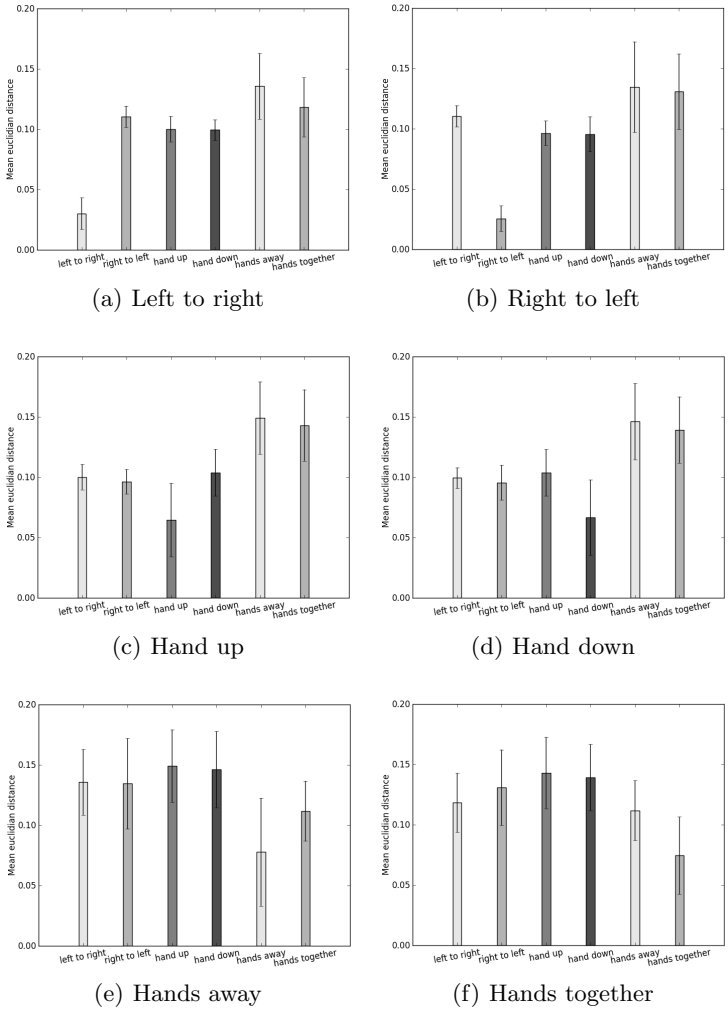


Fig. 4. Plot of 14 histograms containing samples from two classes: hand up and hand down. Black colour corresponds to hands down movement, gray colour corresponds to hans up movement. Feature vector element number are on horizontal axis and values on vertical axis.

Classification is based on k-nearest neighbours method with Euclidian distance. Cross-class similarity estimation was performed by calculating mean Euclidian distances and standard deviation between descriptions of each class compared to another classes in the Fig. 5.



(a) Left to right (b) Right to left

(c) Hand up (d) Hand down

(e) Hands away (f) Hands together

Fig. 5. Mean Euclidian distances and standard deviation (error bars) between feature vectors of each class compared to another classes

5 Application

The new human-computer interface was implemented based on the gesture recognition method described above. The application for hand driven slideshow control was developed as an example of this interface. The client–server application architecture was proposed to achieve remote slideshow control. The architecture of the demonstration application is presented in the Fig. 6.

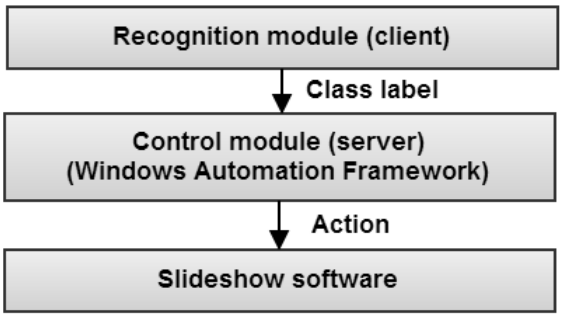


Fig. 6. The architecture of the application

6 Conclusion

Hand motion guided human-computer interface based on the new dynamic patterns descriptor is presented. The distinctiveness of proposed gesture description was demonstrated by cross—class Euclidian distance measurement of training samples. Hand motion is described by the sequence of motion distribution histograms. This method demonstrates sufficient processing speed in terms of end user experience and classification accuracy for gesture sequences to be used for remote slideshow control. Further research within proposed approach aims to support different gestures types and non-relevant objects motion filtering using skin color map, depth map and motion map.

Acknowledgments

The work is supported by the Russian Foundation for Basic Research, grants 12-01-31226 mol_a and 12-01-31266 mol_a.

References

1. Rautaray, S.S., Agrawal, A.: A real time hand tracking system for interactive applications. International Journal of Computer Applications (0975 8887) Volume 18 No. 6 (2011).
2. Shan, C., Tan, T., Wei, Y.: Real time hand tracking using a mean shift embedded particle filter. Pattern Recognition 40 (2007).
3. Davis, J. W.: Recognizing Movement using Motion Histograms. MIT media laboratory. M.I.T. Media Laboratory Perceptual Computing Section Technical Report No. 487 (1998)
4. Torres, G.: Gesture recognition using motion detection. University of Kansas (2009)
5. Banerjee, P., Sengupta, S.: Human motion detection and tracking for video surveillance. Department of Electronics and Electrical Communication Engineering Indian Institute of Technology. National Conference on Communication (2008)

6. Stauffer, C., Grimson, W.E.L.: Adaptive background mixture models for real-time tracking. The Artificial Intelligence Laboratory Massachusetts Institute of Technology Cambridge, MA 02139 (1998)
7. Cutler, R., Davis, L.: Robust Real time periodic motion detection, analysis, and applications. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 22, No. 8 (2000)
8. Belongie, S., Mori, G., Malik, J.: Matching with Shape Context. Department of Electrical Engineering and Computer Sciences, University of California at Berkley (2000)
9. Fogelton, A.: Real-time Hand Tracking using Modicated Flocks of Features Algorithm. Information Sciences and Technologies Bulletin of the ACM Slovakia, Special Section on Student Research in Informatics and Information Technologies, Vol. 3, No. 2 37-41 (2011)
10. Manresa, C., Varona, J., Mas, R. Perales, F. J.: Real time hand tracking and gesture recognition for human computer interaction. Electronic Letters on Computer Vision and Image Analysis 0(0):2-7 (2000)
11. Deng, L.Y., Hung, J.C., Keh, H., Lin, K., Liu, Y., Huang, N., Real time hand gesture recognition by shape context based matching and cost matrix. Journal of networks, vol. 6, no. 5 (2011)
12. Sobral, Andrews, BGSLibrary: An OpenCV C++ Background Subtraction Library. IX Workshop de Viso Computacional (WVC'2013) (2013)