P.G.Demidov Yaroslavl State University

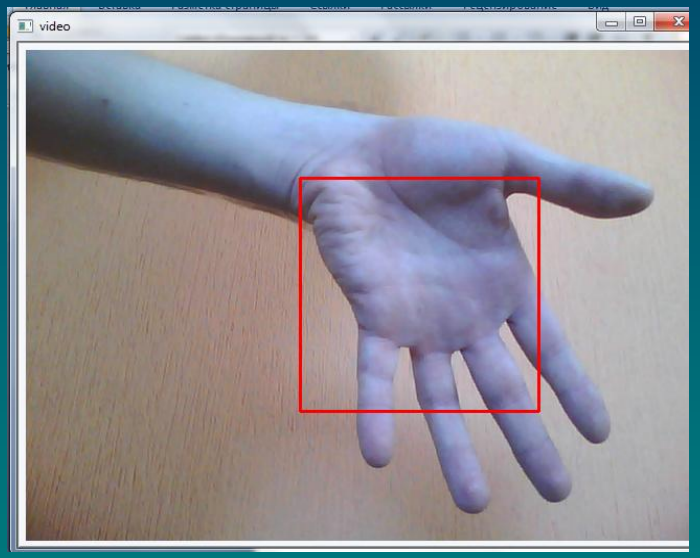
Recognition of Hand Gestures on the Video Stream Based on a Statistical Algorithm with Pre-treatment

Authors: Palochkin V.V.,
Maksimovskiy A.S.,
Priorov A.L.

FRUCT 15th, 2014

Introduction

Object recognition is a classical problem of computer vision, image processing and machine learning. This is important for solving the problems of detection, assessment of the objects position, digital image processing and for remote systems management.



Goal and tasks

The goal is to develop an algorithm for recognizing hand gestures on the video stream in real time. This is achieved by combining the algorithms of texture allocation, color allocation and other machine learning algorithms.

To achieve this goal it is necessary to solve the following tasks:

- *Preliminary processing*
- *Statistical algorithm*
- *Gesture classification*

PRELIMINARY PROCESSING

During pre-processing color allocation algorithm, texture allocation algorithm and Lucas–Kanade method were used.

For the object detection the pixel coordinates are determined both by the spatial coordinates and by the coordinates in RGB. On each frame of video stream objects, which pixels fall within ε - neighborhood of the selected pixel in RGB color system, are segmented. On Fig. the results of color detection algorithm are illustrated for test object.

Texture allocation method

Texture can be described by some attributes. Under the textures attributes characteristic properties are considered. They are common to all textures of a certain class. Texture attributes play a crucial role for their classification and the separation of images into separate areas.

As such attributes you can use the statistical moments of the spatial distributions :

• k –initial moment

$$T_1^k = n^{-2} \sum_{i=1}^{n} \sum_{j=1}^{n} [f(i,j)]^k$$

• energy:

$$T_3 = \sum_{g=0}^{N-1} [F(g)]^2$$

• entropy:

$$T_2 = -\sum_{g=0}^{N-1} F(g) log_{10} F(g)$$

• variation:

$$T_4 = \sum_{g=0}^{N-1} (g-\mu)^2 F(g)$$

Texture attributes

There is an approach for the formation of textural features that take into account the mutual arrangement of the pixels within the sliding window. It is based on the use of adjacency matrix.

By calculating the adjacency matrices it is possible to calculate directly the numerical estimates of a number of textural features:

- average:

$$T_5 = \mu_i = \mu_j = \sum_{i=0}^{N-1}\left[i \sum_{j=0}^{N-1} P(i,j) \right]$$

- variation:

$$T_7 = \sigma_i^2 = \sum_{i=0}^{N-1}\left[(i-\mu_2)^2 \sum_{j=0}^{N-1} P(i,j) \right]$$

- energy:

$$T_6 = \sum_{i=0}^{N-1}\sum_{j=0}^{N-1} [P(i,j)]^2$$

- uniformity:

$$T_8 = \sum_{i=0}^{n}\sum_{j=0}^{n} P(i/j) \Big/ (1+|i-j|)$$

Lucas–Kanade method

It is necessary for practical application of object detection algorithm that not all objects will be allocated on each frame of a video stream if their pixels fall within the ε – neighborhood. But only that object will be allocated if the pixel is within the chosen part of image. The pixel coordinates of the user object are defined in the spatial coordinates, and the tracking under this pixel is implemented. An Lucas–Kanade method is used as such algorithm.

The results of detection algorithm based on the Lucas-Kanade method

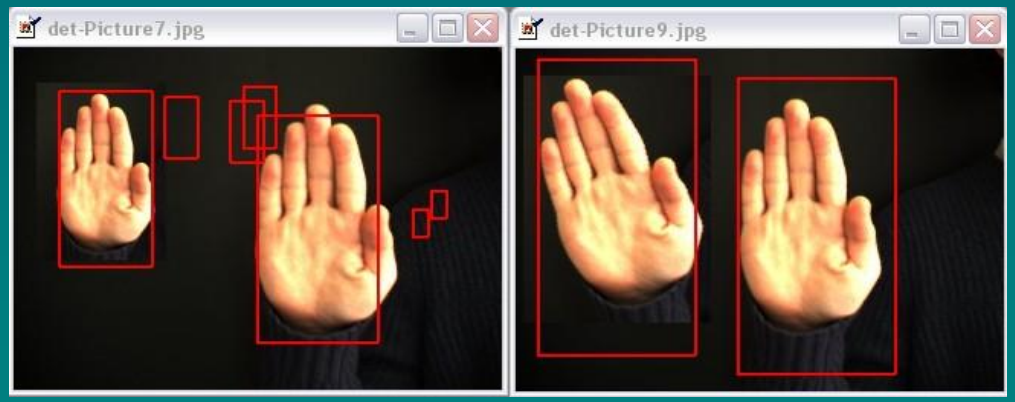
During the processing of a pyramidal iterative algorithm is used and the point is searched that minimizes the weighted quadratic model of restrictions on first-order derivatives in the search window.

Before application of the algorithm it is necessary to smooth the input image to remove random noise, because the algorithm is local and it depends strongly from any noise.

Object detection methods

There are various methods for detecting hands in the image. These methods can be divided into two groups: vision-based approach and 3D hand model based approach. Hand detection method is implemented on the video stream with using Haar classifier. It is based on the image pre-processing algorithm and background subtraction algorithm.

Nowadays Viola-Jones method with using Haar features and AdaBoost-classifier is one of the best algorithms for solving problems of finding objects on the video stream in real time. In this paper, Viola-Jones method is used to solve the problem of hand detecting in real time.

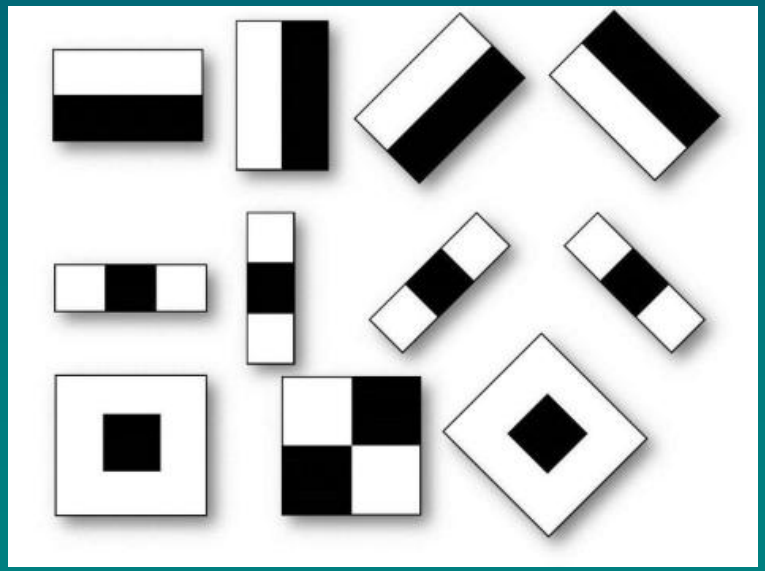


Haar-like features

Haar-like features are calculated by the formula:

$$f(x) = \sum_{S_1} I - \sum_{S_2} I ,$$

where I – pixels intensity; S_1 – all pixels in the area of dark rectangle; S_2 – all pixels in the area of white rectangle.



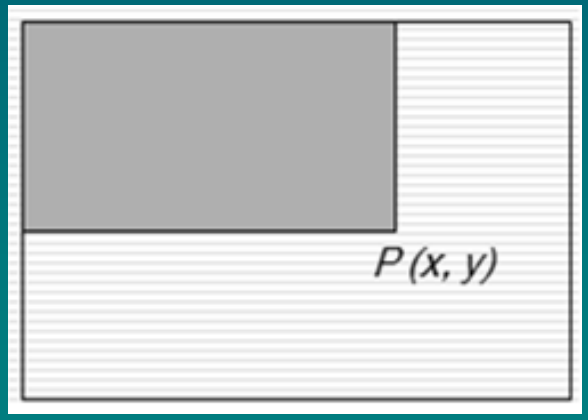
Integral image

Integral image. Integral representation of the image is a matrix with the same size as the original image. The sum of the intensities of all pixels located above and to the left of this element is stored in each its element. Matrix elements are calculated using the following formula:

$$L(x,y) = \sum_{i=0,j=0}^{i \le x, j \le y} I(i,j)$$

where I(i,j) — pixel brightness of the original image.

Using the integral image allows to calculate the Haar-like features more quickly. And it does not depend on the size of image.



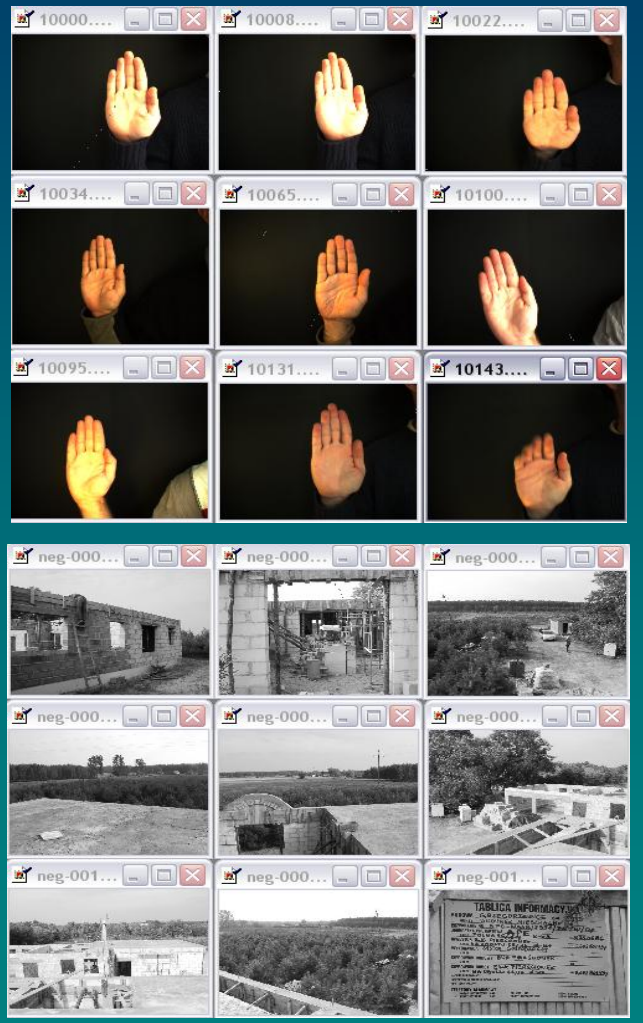
Boosting

Boosting is the set of methods that improve the accuracy of the analytical models. Effective model that admits little misclassification is called "strong". "Weak" model does not allow to provide a significant separation between classes or to give accurate predictions. It makes a lot of errors. Therefore, boosting means "strengthening" of "weak" models. It is a procedure for constructing the serial composition of machine learning algorithms, where each following algorithm tries to compensate the shortcomings of the composition of all previous algorithms.

Classification problems

In classification problems set of real numbers is commonly used as a guest space into two disjoint classes. Decision rules may have customized settings. Thus, in the algorithm of Viola-Jones a threshold decision rule is used. At the beginning the operator at zero is constructed, and then the optimal value is selected. Process of sequential learning of basic algorithms is used often in the construction of compositions.

Various criteria may be used depending on the specific of the problem.



Adaboost-algorithm

Video stream obtained with a help of a video camera, is a sequence of frames. For each frame integral image is calculated. Then, the window frame is scanned by small size window containing Haar-like features. For each j-th feature the corresponding classifier is defined by the formula.

$$h(x) = \begin{cases} 1, & p_j f_j(x) < p_j \theta_j \\ 0 & \end{cases}$$

where x – small size window; θ_j – threshold; p_j – direction of the inequality sign; f_j –Haar-like feature.

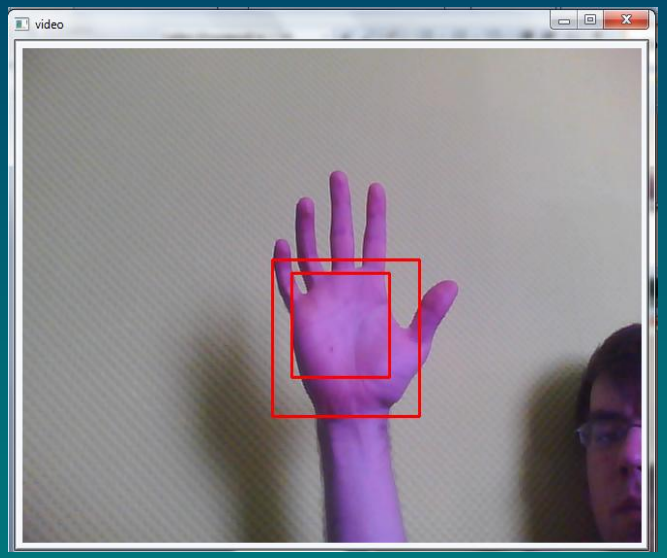
AdaBoost-algorithm can improve classification accuracy through a series of stages of weak classifiers. As a result, weighted combination of weak classifiers is calculated.

$$H(x) = \sum_{i=1}^{N} a_i h$$

(N –number of weak classifiers; α_i– a coefficient derived from the database; h – weak classifier).

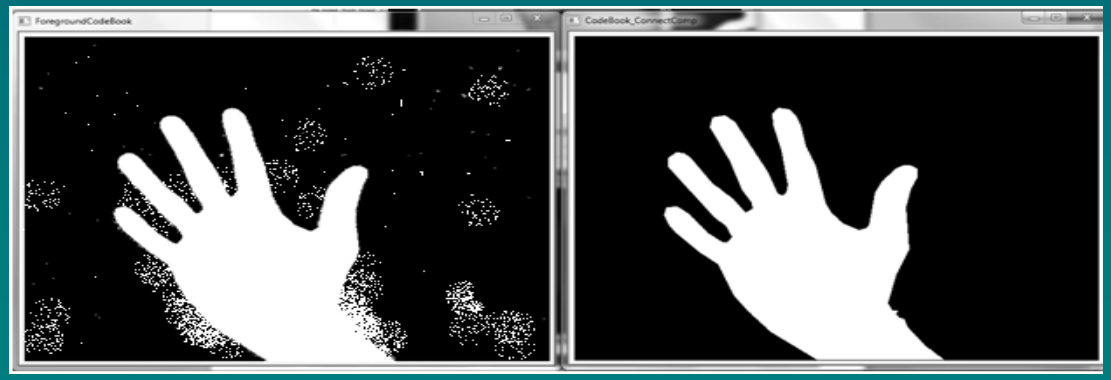
GESTURE CLASSIFICATION

Hand gesture recognition algorithm was created based on the detection algorithm. This algorithm builds a hand envelope, counts straighten fingers in real time and performs tracking that can be used for remote control of radio-physical and other technical devices.



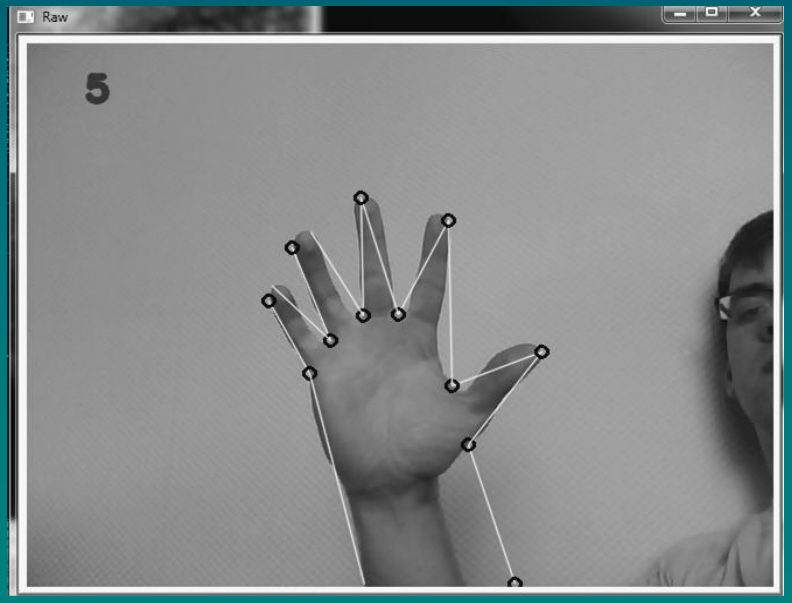
Background subtraction algorithm

Background pixels are not affected by the weighting, because background is a static object. Pixels that are constantly changing, are not the background image. After each iteration the background pixels are becoming more noticeable especially if they are compared with the pixels of the moving object (hands). Background subtraction involves the calculation of the reference image, the subtracting of each new frame of the image and the image segmenting which identifies nonstationary areas of objects.



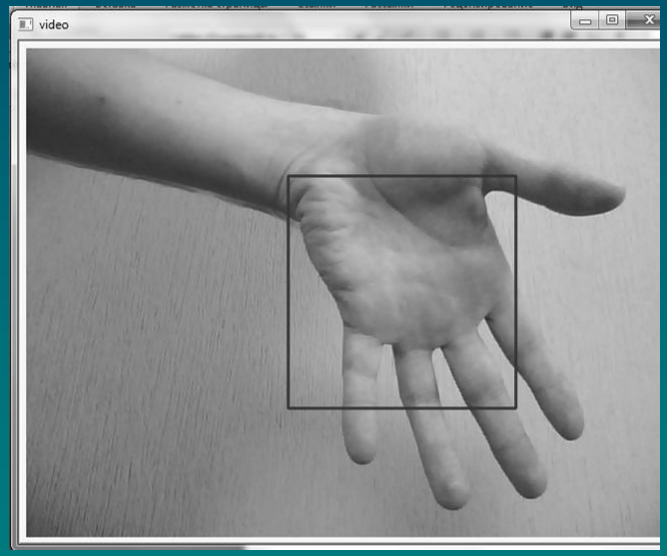
Convex hull and defects

Convex hull is in the fingers, because they are limbs. And hence this fact can be used to detect "NO fingers". But on the hand there are other special points - convex defects - the deepest points of diversion on the contour. To implement the hand tracking center of the palm have to be find. There are 2 convex defects for every convex point



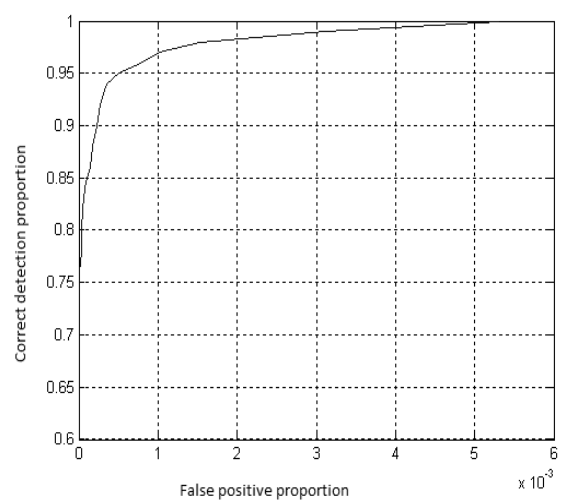
Hand detection algorithm

Considered hand detection algorithm was tested on a database of digital images. Database size was 1000 images. It was a set of gestures.

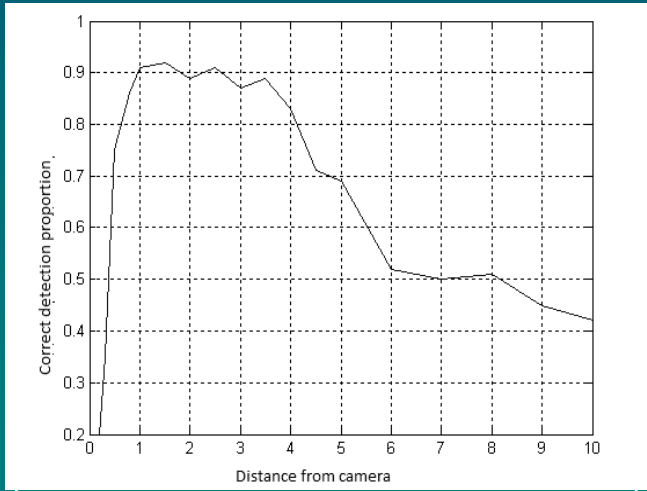


All measurements were performed on a personal computer with the following configuration: Intel(R) Pentium(R) D 930 CPU 3.00GHz, RAM 4GB 400MHz, Windows 7 64-bit.

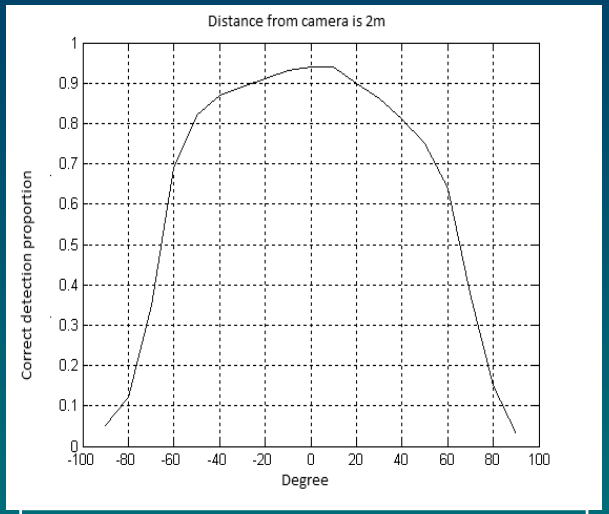
Measurements



ROC-curve for detection algorithm



The dependence of correct detection from the angular position



The dependence of correct detection from the distance

Conclusion

It is proposed a combined method of hand gesture recognition based on a statistical image processing algorithm. As a pre-processing algorithm there were applied Lucas–Kanade method, background subtraction algorithm, color detection and texture detection methods. Object recognition was performed with using of Haar classifiers. The possibility of using gestures for remote control of various devices with different hands position from the camera location was shown.

The analysis of the applicability of the algorithm is made under different external conditions. The accuracy of the algorithm is high enough and recognition is less exposed to errors of the first and second kind when the camera is oriented within the permissible angular aperture, which carried out the detection arm.

P.G.Demidov Yaroslavl State University

Recognition of Hand Gestures on the Video StreamBased on a Statistical Algorithm with Pre-treatment

Authors: Palochkin V.V.,
Maksimovskiy A.S.,
Priorov A.L.

FRUCT 15th, 2014