The effectiveness of automatic laparoscopic diagnostics of liver pathology using different methods of digital images classification

D. N. Bayazitov¹, N. V. Kresyun¹, A. B. Buzinovsky²,³, N. R. Bayazitov¹, A. V. Lyashenko¹, L. S. Godlevsky¹, T. V. Prybolovets¹, K. A. Bidnyuk¹

¹Odesa National Medical University, Ukraine, ²International center on researches and education in the field of informational technologies and systems at National Academy of Sciences and Ministry of Education of Ukraine, Kyiv, ³Odesa Regional Clinic Hospital, Ukraine

Key words: liver diseases, laparoscopy, diagnostic imaging, image processing, computer-assisted.

Aim: to evaluate the effectiveness of the CAD systems developed on the basis of two classifiers – HAAR features cascade and AdaBoost for the detection of cirrhotic and metastatic damages of the liver.

Materials and methods. The development of CAD was based on training of HAAR features cascade and AdaBoost classifiers with images/frames, which have been cropped out from video gained in the course of laparoscopic diagnostics. RGB frames which were gamma-corrected and converted into HSV have been used for training. Also descriptors were extracted from images with the modified method of Local Binary Pattern (LBT), which includes data on color characteristics (“modified color LBT” – MCLBT) and textural ones for AdaBoost classifier training. 1000 positive images along with 500 negative ones of both types of pathology were used for training. After cessation of training the tests were performed with the aim of the estimation of effectiveness of recognition. Test session images were different from those ones which have been used for training of the classifier. Test control sessions were performed with trained classifiers with 319 frames containing cirrhotic and 253 frames with metastatic deteriorations in liver tissue. 365 frames with the absence of mentioned pathology were used as a control group – practically healthy liver state.

Results. Classification of test video-images revealed that the highest recall for cirrhosis diagnostics was achieved after training of AdaBoost with MCLBT descriptors extracted from HSV images – 0.655, and in case for metastatic damages diagnostics – for MCLBT gained from RGB images – 0.925. Hence developed AdaBoost based CAD system achieved 69.0 % correct classification rate (accuracy) for cirrhotic and 92.7 % for metastatic images. The accuracy of Haar features classifier was highest in case of metastatic foci identification and achieved 0.701 (RGB) – 0.717 (HSV) values.

Conclusions. Haar features based cascade classifier turned to be less effective when compared with AdaBoost classifier trained with MCLBT descriptors. Metastatic foci are better diagnosed when compared with cirrhotic liver deterioration with the explored approaches to digital images classification.
Computer automatic diagnostic (CAD)/classification of videoimages is actual for minimal invasive abdominal surgery and endoscopy [1–5]. CAD systems are intensively developed for tracking laparoscopic instrumentation [6], identification of zones of pathology [4,7,8] and the diminution the risk of damage of healthy tissues [2]. Thus, A. La-hane et al. [2] used series of computer vision algorithms to track surgical instruments and coarse location of the cystic artery. It indicates the possibility of an injury to the cystic artery by automatically detecting the proximity of the surgical instruments with respect to the cystic artery [2].

Some peculiarities were stressed for video-laparoscopic images automatic classification [5]. Thus, authors pointed on the absence of tangible differences in elements of images as a result of rather high level of noise and lack of lightness. Besides, another obstacle, which complicates the CAD is confined to high variability of the shape of objects, which are searching for. Also the majority of pixels contain prevalently red color hues, which reduced their informative significance. The quick change of the angle of object along with changes of their illumination and specular reflections significantly impact on the result of automatic recognition [5]. Hence, the usage of gamma-correction in the course of preprocessing of primarily gathering information, which permits to identify relations between quantitative pixel characteristics and their actual brightness is justified [4].

Taking into consideration the necessity of fast working of algorithm for video-data analysis, it is reasonable to explore classifier based on Haar-features cascade [6]. Haar features system first time proposed by [9] for face recognition was described in details in a lot of works, including those ones devoted to tracking of laparoscopic instrumentation [8]. Exploration of Haar features presupposes the investigation of the sum of pixels intensity after broccoli initial image into rectangular tile format. The quick change of the angle of object along with changes of their illumination and specular reflections significantly impact on the result of automatic recognition [5]. Hence, the usage of gamma-correction in the course of preprocessing of primarily gathering information, which permits to identify relations between quantitative pixel characteristics and their actual brightness is justified [4].

Taking into consideration the necessity of fast working of algorithm for video-data analysis, it is reasonable to explore classifier based on Haar-features cascade [6]. Haar features system first time proposed by [9] for face recognition was described in details in a lot of works, including those ones devoted to tracking of laparoscopic instrumentation [8]. Exploration of Haar features presupposes the investigation of the sum of pixels intensity after broccoli initial image into rectangular tile format. The quick change of the angle of object along with changes of their illumination and specular reflections significantly impact on the result of automatic recognition [5]. Hence, the usage of gamma-correction in the course of preprocessing of primarily gathering information, which permits to identify relations between quantitative pixel characteristics and their actual brightness is justified [4].

Taking into consideration the necessity of fast working of algorithm for video-data analysis, it is reasonable to explore classifier based on Haar-features cascade [6]. Haar features system first time proposed by [9] for face recognition was described in details in a lot of works, including those ones devoted to tracking of laparoscopic instrumentation [8]. Exploration of Haar features presupposes the investigation of the sum of pixels intensity after broccoli initial image into rectangular tile format. The quick change of the angle of object along with changes of their illumination and specular reflections significantly impact on the result of automatic recognition [5]. Hence, the usage of gamma-correction in the course of preprocessing of primarily gathering information, which permits to identify relations between quantitative pixel characteristics and their actual brightness is justified [4].
the time for analysis and makes the process of recognition congruent with the flowing of video frames. For each window the sum of pixel intensities in adjacent windows is calculated and the difference is taken between those regions and the window in question. Obtained difference is indicative for the classification of the target window which is necessary for the comparison of gained data with those ones presented in the library specially composed for the training of the Haar cascade classifier. Hence, critical step for gaining results is confined to the correct creation of the library of Haar-features which provides with maximal versions of divergent appearance of pathological process in target zone.

The training of Haar features based cascade needs a lot of time which might be measured in months when tens of thousands of images are used [6]. That is clear disadvantage of the method and it is hard to avoid it as far as increasing the number of images is proportional to the effectiveness of diagnostics [6].

We have chosen the two versions of pathology with relatively low for the CAD – suspicion of focal metastatic damage of liver and suspicion of cirrhotic deteriorations of the liver. Such an approach permits to minimize the size of zone of interest as well as to avoid to some extent the drastic visual variability of pathological appearance. Besides, we hoped to get most informative and few/least numerous descriptors of texture, color and shape of pathological zones.

Texture features extracted from pathological endome trium were different from normal one, and were characterized by lower image intensity, while variance, entropy and contrast gave higher values [4]. Thus, for the hysteroscopic images the heightened mediana of grey scale is clearly identified and along with more homogenous and less contrast might be treated as informative differential index for normal state identification [4]. That is why we decided to use both color and texture features for training Haar-features based classifier. Besides, as an alternate to Haar features-based classifier we have explored AdaBoost classifier trained with minimal number of descriptors gained from Local Binary Pattern (LBP) method application. The classical method of LBT manipulates with grey scale of color and ignores other colors information. Instead of the modified LBT method, which includes data on color characteristics (modified color LBT – MCLBT) [11,12] was used in the present investigation for gaining color and texture descriptors [3,13].

The main aim of this work was to work out and to compare the effectiveness of CAD based on Haar-features cascade classifier with AdaBoost – based CAD and which were trained to distinguish between normal and pathological state of liver being damaged with metastases or cirrhosis.

**Materials and methods**

The next steps were performed in the course of collecting data and their analysis:

– Calibration of digital camera, which included white balance and conversion of color scale into digital code; that was performed in accordance to instructions of camera’s manufacturer;

– The frontal position of the object which was under inspection with deviation from right angle up to 15 ± 5° and distance to visualized zone from 3 to 5 cm was used [4]; those images which have been got in a such a fashion were used for both CAD system training and testing;

– There is no documentation for OpenCV that describes what size the positive samples should be scaled to. Some tutorials suggest using sizes around 20 x 20 or pixels for face recognition or 24 x 24 pixels for polyps endoscopic diagnostics [7]. In our investigation those zones which have been of interest from diagnostic point of view have been identified off-line with their size 60 x 60 pixels and used for training of classifier; in the course of laparoscopic intervention the speed of video frames was modified via using the low frequency filter and size of image was artificially modified from 30 x 30 up to 60 x 60 pixels, which was necessary for optimizing classificatory performance.

– Gamma-correction performing of gained image with the recalculation of gamma-coefficient;

– Conversion of RGB scale into HSV one; such a conversion was justified by orientation of Haar features on the estimation of the intensity of pixels;

The OpenCV library contains a built-in function for developing this library via Haar feature classifiers [10]. We have used the next OpenCV modules in our work: 1) OpenCV_core – for basic calculations, generation of pseudorandom numbers, XML import/export e. c.; 2) OpenCV_imgproc – images processing; 3) OpenCV_highgui – simple UI, upload/storage of images and video; 4) OpenCV_ml – methods and models of adaboost training; 5) OpenCV_video – movement analysis and tracking (optical stream, templates of movement, background abo lishing; 6) OpenCV_objdetect – detection of the objects on images (Haar, HOG e. c.); 7) OpenCV_calib3d — camera calibration.

– Training to Haar features, using both RGB and HSV images.

– Training AdaBoost classifier with MCLB templates [14]; in both cases key features, which were used, were confined to mean, entropy, contrast, homogeneity and excesses.

– Results of classification were stored at data base which permitted to perform additional analysis later on.

All the laparoscopic videos were got with laparo scopic camera with 5 mm aperture diameter “Carl Storz Tricam Camera (Carl Storz, Germany) during 2011–2016 years. That camera had the analogous input (PAL 475 horizontal lines) and incoming signal was digitalized with the pixel density of 720 x 576 and capture was made with video capture card “averMedia HD capture Studio 203” (Avermedia, France) and presented at CAD interface (Fig. 1).

Criteria (both technical and medical), which have been used for the inclusion were the next: documented digital camera calibration made in accordance to the manufacturer instruction; the average severity of patient’s state and approving of diagnoses by results of clinician and laboratory and instrumental investigations.

**Features extraction and classifiers training.** Being applied to RGB scale MCLBT calculates LBT for R and G channels of normalized RGB color space [11,12].
The more stable data were provided with RGB – MCLBT under different conditions of illumination intensity as far as after normalization the invariant state of “R” and “G” channels was achieved.

The texture characteristics calculation using HSV – MCLBT was performed via recalculations on Hue channel, which was invariant with regard to illumination and saturation variability. For LBT calculation the radius of 1.5 and 12 pixels was applied [12]. The pertinent pattern was created for each of scale vector, as a result and the characteristic vector was determined for templates of MCLBT, which included mean, entropy, contrast, homogeneity and excesses.

For the training of classifiers 32 laparoscopic video-images patients with metastatic liver damage and 35 with cirrhosis were used as “positive” ones (Fig. 2). Also for the classifier training 40 videos gained from normal liver surface were used as a control – “negative” images. Each video was classified in accordance to final diagnoses, which have been proved with clinical and instrumental methods of diagnostics or/and with histological data. Hence, in all the cases training was performed retrospectively.

Each video contained 2500–3000 frames, among which manually those ones for teaching and testing collections were verified, cropped out and storage.

For the classifiers training the next parameters were explored:
- False positive rate $f = 0.3$;
- Windows with the size of frame as 60 x 60 pixels;
- Number of positive images $n = 1000$ for each pathology;
- Number of negative images $n = 500$.

After cessation of training the tests were performed with the aim of the estimation of effectiveness of recognition.

Test session images were different from those ones which have been used for training of the classifier. Test control sessions were performed with 319 frames containing cirrhotic and 253 frames with metastatic deteriorations in liver tissue. 365 frames with the absence of mentioned pathology was used as a control group – practically healthy liver state.

Statistical procedures. To assess the performance of our classifiers, we use the measures precision, recall and F-score [7].

Precision measures the fraction of the detected-positive instances, which are true-positive:

$$\text{Precision} = \frac{TP}{TP + FP}.$$  

$TP$ is the number of true-positive instances, $FP$ is the number of false-positive instances.

Recall is the fraction of all true-positive instances, which are also detected positive. The formula for calculating the recall is:

$$\text{Recall} = \frac{TP}{P}.$$  

$P$ is the number of positive instances.

F-score (also F-measure or F1-score) is the harmonic mean of precision and recall. It is thereby a combination of these two measures in a single number.

$$F = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}.$$  

In several sections, the measure accuracy is used as well. Accuracy is the proportion of correctly classified items out of all the items classified.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}.$$  

$TN$ is the number of true-negative, $FN$ is the number of false-negative instances.

Results

Gained data revealed that the effectiveness of diagnostics of cirrhosis was lower than diagnostics of metastasis independently of type of frames which have been used for training (Table 1). The highest recall for cirrhosis diagnostics was registered after training with MCLBT obtained from HSV converted images – 0.655, while for metastasis diagnostics recall for RGB – MCLBT was higher than for HSV – MCLBT – 0.925 and 0.913 respectively.
Training with modified templates of both RGB and HSV images and minimal MCLBT derived descriptors substantially improved results of classification performed with AdaBoost classifier. Thus, MCLBT descriptors which were got both from HSV and RGB images permitted to reduce number of false negative diagnoses from 3.64 times to 4.57 times when pertained to the results of diagnostics of metastatic liver damage after training with only HSV or RGB – derived HAAR features correspondently. Less pronounced reduction was observed in case of diagnostics of cirrhosis – from 1.2 (HSV) to 1.65 (RGB) times. At the same time increasing the number of true positive diagnoses was more modest and raised up to from 1.3 to 1.5 times for both cirrhosis and metastases. A little bit better results gained with RGB images to some extent correspond with data after [15]. Authors showed the preference of RGB images for tissue classification, when used with widely applied feature descriptors and that combining the tissue texture with the reflectance spectrum improves the classification performance.

Hence, more pronounced improvement – reduction of false negative results points on meaningful role played by extracted features and used templates for classifier training. Despite main advantage of Haar features based classification, namely – small amount of data used for the machine learning process as well as avoiding overfitting of trained classifier [7], the exploration of MCLBT-based training of AdaBoost should be recognized as more effective for laparoscopic CAD of liver state. This result does not exclude principal possibility of high effectiveness of cascade classifier trained with MCLBT descriptors as well and AdaBoost trained with Haar descriptors [3,11].

Today we have no strict recommendations on the protocol of automatic classification of laparoscopic video-images. Meanwhile, being based upon delivered data it is reasonable to use MCLBT descriptors for AdaBoost classifier training for resolving problems of automatic diagnostics in laparoscopic surgery.

**Conclusions**

1. The CAD of laparoscopic images based on the AdaBoost classifier permitted effectively classify cirrhotic and metastatic deterioration in liver tissue with highest recall gained with MCLBT from HSV images used for training – up to 0.672, and for MCLBT RGB – up to 0.912 correspondently.

**Table 1. Detector performance of HAAR features-based and AdaBoost classifiers trained with different types of images**

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Frames type used for training</th>
<th>True positive</th>
<th>True negative</th>
<th>False positive</th>
<th>False negative</th>
<th>Precision</th>
<th>Recall</th>
<th>F1 Score</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cirrhosis</td>
<td>Haar-features cascade</td>
<td>133</td>
<td>225</td>
<td>140</td>
<td>186</td>
<td>0.487</td>
<td>0.417</td>
<td>0.450</td>
<td>0.523</td>
</tr>
<tr>
<td></td>
<td>HSV</td>
<td>162</td>
<td>202</td>
<td>163</td>
<td>157</td>
<td>0.498</td>
<td>0.508</td>
<td>0.503</td>
<td>0.532</td>
</tr>
<tr>
<td></td>
<td>MCLBT (RGB)</td>
<td>206</td>
<td>280</td>
<td>85</td>
<td>113</td>
<td>0.708</td>
<td>0.524</td>
<td>0.602</td>
<td>0.711</td>
</tr>
<tr>
<td></td>
<td>MCLBT (HSV)</td>
<td>209</td>
<td>263</td>
<td>102</td>
<td>110</td>
<td>0.672</td>
<td>0.655</td>
<td>0.663</td>
<td>0.690</td>
</tr>
<tr>
<td>Metastasis</td>
<td>Haar-features cascade</td>
<td>166</td>
<td>267</td>
<td>98</td>
<td>87</td>
<td>0.629</td>
<td>0.656</td>
<td>0.642</td>
<td>0.701</td>
</tr>
<tr>
<td></td>
<td>HSV</td>
<td>173</td>
<td>270</td>
<td>95</td>
<td>80</td>
<td>0.645</td>
<td>0.684</td>
<td>0.664</td>
<td>0.717</td>
</tr>
<tr>
<td></td>
<td>MCLBT (RGB)</td>
<td>234</td>
<td>339</td>
<td>26</td>
<td>19</td>
<td>0.900</td>
<td>0.925</td>
<td>0.912</td>
<td>0.927</td>
</tr>
<tr>
<td></td>
<td>MCLBT (HSV)</td>
<td>231</td>
<td>327</td>
<td>38</td>
<td>22</td>
<td>0.859</td>
<td>0.913</td>
<td>0.895</td>
<td>0.903</td>
</tr>
</tbody>
</table>

Nevertheless for both types of pathology MCLBT-based methods permitted to get better diagnostics results in comparison with the based on RGB and HSV images training of classifier.

The presence of rather low level of true positive results of diagnostics in case of training with RGB and HSV images independently of type of pathology should be stressed (Table 1). Further extraction of features and training with corresponded templates permitted to increase number of true positive results of diagnostics along with reduction of false negative results. Thus, the net increase of true positive diagnoses in case of MCLBT usage for training pertained to training with both RGB and HSV images was from 1.3 up to 1.5 times for both pathology, while decrease of false negative diagnoses was from 1.65 (RGB) to 1.42 (HSV) for cirrhosis and reduction of false negative diagnoses achieved from 4.57 times (RGB) to 3.64 times (HSV).

**Discussion**

Hence, gained data are in favor for the rather high effectiveness of automatic diagnostics of liver pathology with the created CAD.

Meanwhile, it was mentioned that only Haar-like feature based classifier is not enough to be a strong object classifier as far as it needs prolonged recalculation of features for robust detection of image [6]. Our results also have revealed that training with RGB and HSV images was less effective for proper diagnostics, especially in case of cirrhotic deteriorations detection. That might be a sign of rather weak potential of HAAR features based classifier for the correct recognition of suspected pathology. Thus, work performed by [7] which was devoted to automatic diagnostics colon polyps via endoscopics images analysis revealed that usage of Haar-features or Histogram of oriented Gradients based detectors were weak for reliable colon polyps automatic detection. Much more effective was an approach based on Global image feature such as Joint Composite Description, which permitted to gain precision of diagnostics of 93.3 % and weighted average recall of 98.5 %. Main reasons for not satisfactory results were the enormous variety of appearance of such lesions and their orientation. It should be noted that both factors are not so important and are rather well controllable in the course of laparoscopic intervention. That justifies our efforts in the direction of improving effectiveness of diagnostic which is based on Haar-features detectors.
2. MCLBT used descriptors for training AdaBoost classifier proved to increase the precision, recall, F1 score as well as accuracy of automatic diagnostics of cirrhotic and metastatic changes in liver.

Perspectives of future investigations. It is supposed to continue training of both classifiers with increased number of laparoscopic images and with extended forms of laparoscopy — defined pathology. Also the exploration of best version of developed CAD is supposed to perform in operation room.

References

Information about author:
Bayazitov D. N., Postgraduate Student, Surgery Department, Odesa National Medical University, Ukraine.

Godlevsky A. B., Postgraduate Student, International Center on Researches and Education in the Field of Informational Technologies and Systems, National Academy of Sciences and Ministry of Education of Ukraine, Kyiv, Surgeon in Odesa Regional Clinic Hospital, Ukraine.

Kresnew N. V., MD, PhD, DSc, Professor of the Ophthalmology Department, Odesa National Medical University, Ukraine.


