Local Rank Patterns –
Novel Features for Rapid Object Detection

Michal Hradis, Adam Herout, Pavel Zemcik

Graph@FIT, Brno University of Technology, Bozetechova 2, Brno, CZ
{ihradis, herout, zemcik}@fit.vutbr.cz

Abstract. This paper presents Local Rank Patterns (LRP) - novel features for rapid object detection in images which are based on existing features Local Rank Differences (LRD). The performance of the novel features is thoroughly tested on frontal face detection task and it is compared to the performance of the LRD and the traditionally used Haar-like features. The results show that the LRP surpass the LRD and the Haar-like features in the precision of detection and also in the average number of features needed for classification. Considering recent successful and efficient implementations of LRD on CPU, GPU and FPGA, the results suggest that LRP are good choice for object detection and that they could replace the Haar-like features in some applications in the future.

1   Introduction

Detection of objects in images has many real-world applications ranging from camera orientation to computer-human interaction, surveillance and robot vision. One of the approaches to real-time object detection is scanning images for specific 2D patterns using fast detection classifiers. Such classifiers are in many cases modifications of the classifier originally proposed by Viola and Jones [1], who designed the first real-time frontal face detector using a cascade of boosted classifiers and Haar like features. The cascade structure of the classifier together with rapid computation and relatively high discriminative power of the Haar like features resulted in very low average decision time which is essential in the detection task.

Aside of the frontal faces, the detection classifiers have been used to detect various objects such as occluded faces [2], pedestrians [3], general interest points [4], automobiles, stop signs, tea cups [5] etc. Recently, Sochman and Matas [6] have proposed a unified approach how to emulate behavior of any existing application-specific detector by sequential classifier which is optimal in terms of computational complexity for desired detection precision. In their approach, the main effort is put into finding a suitable set of features which are then automatically combined into a WaldBoost [9] ensemble.

The performance of the detection classifiers largely depends on the type of features they use. The ideal features should be computationally inexpensive, should be, to some degree, invariant to geometry and illumination changes and

should provide high discriminative power - all at the same time. High discriminative power is needed to achieve high precision of detection and it also implies more compact and faster classifiers as lower number of features is needed to be computed for the classifier to make a decision. In general, the ideal type of features can differ for different types of objects [6]. However, simple image filters have been proven to generalize well across various types of objects [4] [5].

We propose a novel feature set which is suitable for real-time detection of 2D patterns in images. These features, the Local Rank Patterns (LRP), are invariant to monotonous lighting changes, reflect the amplitude of local changes and also provide information about local patterns. The design of the LRP is based on the recently proposed Local Rank Differences (LRD) [10]. As shown below, the discriminative power of the LRD is unnecessarily constrained and the properties of the resulting classifiers can be significantly improved by removing these constraints. The LRP present one of the possibilities how to do this. The LRP share most of the computational steps necessary to evaluate the LRD; that suggests that the recent real-time implementations of engines evaluating classifiers with LRD on CPU, GPU [8] and FPGA [7] can be easily modified to also accommodate classifiers with the LRP.

The paper is structured as follows. In Section 2, general definition of the LRP is given and some of the properties of LRP are discussed. Section 3 presents the particular form of LRP which was used in experiments on frontal face detection task which are described in Section 4. The paper is concluded and hints for future work are given in Section 5.

2 Local Rank Patterns

Let us consider a scalar image $f : \mathbb{Z}^2 \to \mathbb{R}$. On such image, a *sampling function* can be defined ($\mathbf{x}, \mathbf{u} \in \mathbb{Z}^2, g : \mathbb{Z}^2 \to \mathbb{R}$)

$$S_{\mathbf{x}}^g(\mathbf{u}) = (f * g)(\mathbf{x} + \mathbf{u}). \tag{1}$$

This sampling function is parameterized by the convolution kernel g, which is applied before the actual sampling, and by the vector $\mathbf{x}$ which is the origin of the sampling. Next, let us introduce a vector of relative coordinates ($n \in \mathbb{N}$)

$$\mathbf{U} = [\mathbf{u}_1 \ \mathbf{u}_2 \ \ldots \ \mathbf{u}_n], \mathbf{u}_i \in \mathbb{Z}^2. \tag{2}$$

This vector of two-dimensional coordinates can define an arbitrarily shaped neighborhood and it will be used together with the sampling function to obtain a vector of values describing a neighborhood of this shape on position $\mathbf{x}$ in the image

$$M = [S_{\mathbf{x}}^g(\mathbf{u}_1) \ S_{\mathbf{x}}^g(\mathbf{u}_2) \ \ldots \ S_{\mathbf{x}}^g(\mathbf{u}_n)]. \tag{3}$$

This set of values will be referred to as the *mask* in the following text. The term mask is reasonable as the vector was created by "masking" global information from the image and leaving only specific local information. For each element

k in the mask, its rank can be defined as

$$R_k = \sum_{i=1}^{n} \begin{cases} 1, \text{ if } M_k < M_i \\ 0, \text{ otherwise} \end{cases}, \tag{4}$$

id est, the rank is the order of the given member of the mask in the sorted progression of all the mask members. Note that this value is independent on the local energy in the image, which is an important property useful for the behavior of the Local Rank Patterns image feature, which is defined as ($a, b \in \{1, \ldots, n\}$)

$$LRP(a, b) = R_a \cdot n + R_b. \tag{5}$$

Note that n is the number of samples taken in the neighborhood and therefore the result of LRP is unique for each combination of values of the two ranks R_a and R_b. This fact suggests an alternative definition of the LRP when we allow the results of LRP to be pairs of values instead of single value

$$LRP(a, b) = [R_a \ R_b]. \tag{6}$$

This implies that the LRP have higher descriptive power than the LRD which compute the difference of the two ranks [8].

The LRP have some interesting properties which make them promising for image pattern recognition. Mainly, the LRP are invariant to monotonous grayscale changes such as changes of illumination intensity. This invariance results from using ranks instead of absolute values to compute the value of the feature. In fact, using the ranks has the same effect as locally equalizing histogram of the convolved image $f * g$.

Further, the LRP are strictly local - their results are not influenced by image values outside the neighborhood defined by $\mathbf{U}$. This is a clear advantage over wavelet features (e.g. Haar-like features) which, in the way they are commonly used, need global information to normalize their results. This locality makes the LRP highly independent for example on changes of background and on changes of intensity of directional light.

The meaning of the values produced by the LRP can be understood in two ways. First and most naturally, the results give information about the image at the locations of the two ranks $\mathbf{x} + \mathbf{u}_a$ and $\mathbf{x} + \mathbf{u}_b$ and information about their mutual relation. On the other hand, the results also carry information about the rest of the neighborhood especially if the neighborhood is small. In such cases the results of LRP carry good information about local pattern in the image.

In the previous text, the LRP have been defined for two-dimensional images. However, the notation allows very simple generalization for higher-dimensional images by changing dimensions of $\mathbf{x}$, $\mathbf{u}$ and of the relative coordinates in $\mathbf{U}$ to $\mathbb{Z}^3$ for 3D or $\mathbb{Z}^k$ for general dimensionality. Further, it is possible to use more than two ranks to compute the results of the LRP. For example:

$$LRP(a, b, c) = R_a \cdot n^2 + R_b \cdot n + R_c. \tag{7}$$

3 Set of LRP Suitable for Object Detection

The definition of the LRP which has been given in the previous text is very general. It allows arbitrary sizes and shapes of the neighborhoods and arbitrary convolution kernels. In this section, we will define a set of LRP which is suitable for creating classifiers for detecting objects in images and which is used in the reported experiments.

First, let us define a suitable set of neighborhoods. The number of samples taken in the neighborhood should be high enough for the features to have good discriminative power; however, the number of samples must not be unreasonably high, because computational complexity of the LRP grows linearly with the number of samples. A good first choice is local rectangular subsampling which takes nine samples arranged in a regular 3×3 grid. Such neighborhoods can be defined by a base neighborhood $\mathbf{U}^{base}$ (see Figure 1) which is than scaled to generate all the required sizes.
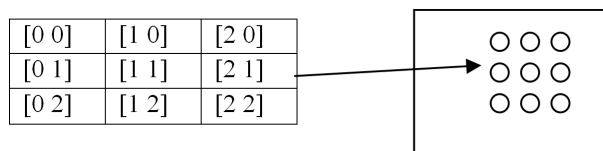
[0 0]	[1 0]	[2 0]
[0 1]	[1 1]	[2 1]
[0 2]	[1 2]	[2 2]

$$\mathbf{U}^{base} = \Big[[0\ 0]\,[1\ 0]\,[2\ 0]\,[0\ 1]\,[1\ 1]\,[2\ 1]\,[0\ 2]\,[1\ 2]\,[2\ 2] \Big]$$

Fig. 1. The base 3×3 neighborhood which is used to generate all other neighborhoods used in this work. The figure shows the relative coordinates of samples together with the locations in the image and also the vector $\mathbf{U}^{base}$ which defines the neighborhood.

In the further text, $\mathbf{U}^{mn}$ will refer to a neighborhood which is created from $\mathbf{U}^{base}$ by scaling the x-coordinates by m and scaling the y-coordinates by n:

$$\mathbf{U}_i^{mn} = \mathbf{U}_i^{base} \cdot \begin{bmatrix} m & 0 \\ 0 & n \end{bmatrix}. \tag{8}$$

The type and the purpose of the convolution kernel g in the LRP can differ. It could be a derivation filter or a wavelet, but the most basic purpose of g is to avoid aliasing when scaling the neighborhood. Aliasing could be avoided by using any low-pass filter such as the Gaussian filter. For efficiency reasons, we use rectangular averaging convolution kernels (see Figure 2). In the following text, g^{kl} will stand for a rectangular averaging filter of dimensions k by l:

$$g^{kl}(\mathbf{x}) = \begin{cases} 1, & \text{if } 0 \le \mathbf{x}_1 < k \text{ and } 0 \le \mathbf{x}_2 < l \\ 0, & \text{otherwise} \end{cases} \tag{9}$$

When scaling the neighborhood, it is reasonable to keep the size of the averaging filter the same as the distance between the samples. In such case m in $\mathbf{U}^{mn}$

equals to k in g^{kl} and n equals to l. In [8], the authors use a set LRD with only four neighborhoods $\mathbf{U}^{11}$, $\mathbf{U}^{12}$, $\mathbf{U}^{21}$ and $\mathbf{U}^{22}$ for samples of dimension 24×24 pixels. They report that the precision of detection of a classifier using this set of features rivals the performance of a classifier with Haar-like features. However, this small subset of all possible LRD is used only due to the constraints of the presented implementation on GPU and it is probable that classifiers using larger set of LRD would achieve better results in most cases. In this work, the set of LRP (and also LRD) is not restricted in this way and all neighborhoods which completely fit into the samples are used.

$$g^{22} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \qquad g^{42} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}, \qquad g^{23} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{bmatrix},$$

Fig. 2. Examples of some of the convolution kernels used in the experiments.

4 Experiments

To evaluate the discriminative power of the LRP and the suitability of the LRP for object detection, a series of experiments was concluded. In these experiments, the LRP were compared to LRD and the traditional Haar-like features. Frontal face detection was selected as the model task, as it is the task most frequently used to report performance of novel detection algorithms and also standardized test sets are available.

4.1 Training Algorithm

The WaldBoost [9] algorithm was used to train all classifiers in the experiments. This algorithm is similar to the algorithm used by Viola and Jones [1] in the way it uses features to create weak hypotheses and in the way it reduces the false positive rate. However, instead of cascade structure of classifiers, it creates a single sequential classifier with *early termination thresholds*. The early termination thresholds allow early classification decision about the "easy" samples. For detection classifiers, the algorithm sets the thresholds to create the fastest possible classifier which still satisfies the desired false negative rate. This desired false negative rate is a parameter of the algorithm and will be referred to as α in the following text.

In the experiments, weak classifiers were selected on 1500 unique positive and 1500 unique negative samples which were selected in each iteration of the WaldBoost algorithm by *importance sampling* from a *training set* which consisted of 100k positive and 100k negative samples. The selected weak hypotheses were then trained on the whole training set. The early termination thresholds were estimated on a *validation set* of the same size as the training set.

4.2 Datasets

The training set and the validation set were randomly bootstrapped from the same data, while guaranteeing that the two sets remain disjunct. The set of faces was generated by applying small random affine transformations to 5000 hand-annotated faces. The non-face samples were randomly sampled from a pool of 250 million sub-windows from more than 3000 non-face images downloaded from the Internet. All samples were normalized to 24×24 pixels.

The created classifiers were tested by scanning images from three separate test sets. The first test set is the MIT-CMU frontal face database[1] which is a standard benchmark for frontal face detection. The other two test sets contain photos of groups of people. The first of them, the GROUP1 test set was gathered by a person who is experienced in face detection. The second test set GROUP2 was gathered by a person who has no knowledge about this field. For more information about the test sets, see Table 1. The images were scanned with position displacement of 2/24 of the scanning window size and with scale factor 1.2. Non-maxima suppression algorithm proposed in [6] was used to post-process the raw detections.

Table 1. The test sets used in the experiments. The table shows the number of images, the number of faces and the number of scanned position in each of the sets.

DataSet	# Images	# Faces	# Positions
MIT+CMU	134	511	19M
GROUP1	89	1618	142M
GROUP2	109	2055	96M

4.3 The Classifiers

The tests included 5 different classifiers whose results are reported in the following section. They differ in the type of features they use and in the type of weak learners. See Table 2 for the complete list of the classifiers. The targeted false positive rate which is specified by the α parameter of the WaldBoost algorithm is 5 % for all of the classifiers. The final length of the classifiers is 1000 weak hypotheses.

The classifier *CHaar* uses the same set of Haar-like features as in [1]. The responses of the Haar-like features are normalized by the standard deviation of pixel intensities in the sample. The weak hypotheses have 10 equidistant bins such that the 8 middle bins contain 90 % of the training samples. The classifier *CLRP* uses the set of features defined in Section 3 and the weak hypotheses have 80 bins which correspond to the individual results of the LRP. The classifier *CLRD* uses a set of LRD features which is defined by the same convolution

---

[1] http://vasc.ri.cmu.edu/idb/html/face/frontal_images/

kernels and the same neighborhoods as are defined in Section 3. The weak hypotheses of CLRD have 17 bins which correspond to the individual results of the LRD features.

As the number of the bins of the three classifiers differs and it could affect the results of the tests, two more classifiers are introduced for better comparison of the individual types of features. The classifier *CLRP24* uses the same features as *CLRP*, except that the ranks R_k (see Eq. 4) are divided by 2. The effect of this change is that the weak hypotheses now have only 24 individual bins. On the other hand, the number of bins for classifier *CHaar80* was increased to 80 for better comparison with *CLRP*.

Table 2. The classifiers which were used in the experiments.

Classifier	Features	# of Bins
CHaar	Haar	10
CLRD	LRD	17
CLRP	LRP	80
CLRP24	LRP	24
CHaar80	Haar	80

4.4 Results

In each of the experiments, an average result of twelve classifiers is reported. The twelve classifiers were created with the same setting of the training algorithm, but with different training set and validation set.
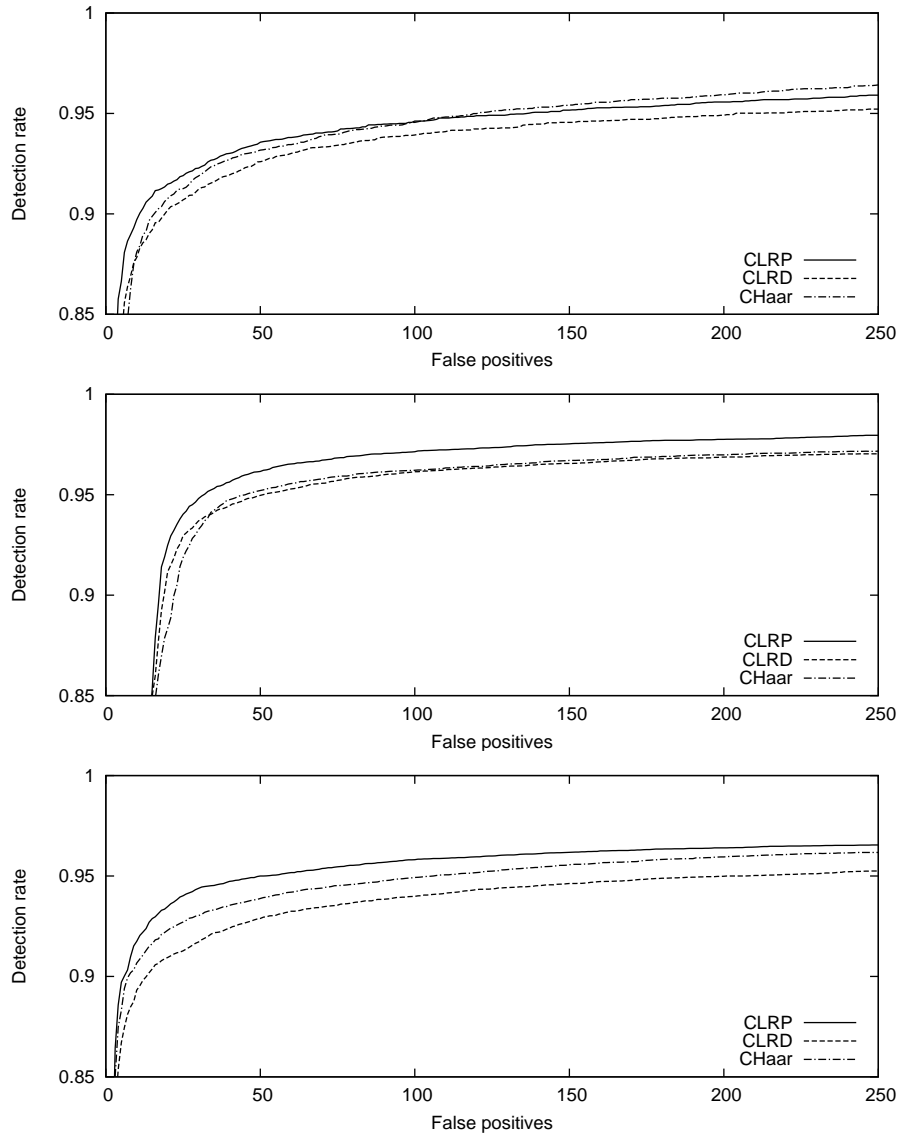
The *Receiver Operating Characteristics* (ROC) which characterize precision of detection of the classifiers can be seen in Figure 3 and in Figure 4. It can be seen that the *CLRP* classifier surpasses all the other classifiers. *CLRP* is definitely an improvement over the *CLRD* classifier and it also reaches better results than the *CHaar* classifier especially in the left part of the ROC.

Figure 4 suggests that the better results of the *CLRP* classifier are not caused by the higher number of bins of the weak hypotheses. For the *CHaar80* classifier, the higher number of bins did not improve performance of the Haar-like features. Also the results of *CLRP24* do not show any significant performance drop of the LRP when the number of bins is reduced.

The average speeds of the classifiers can be seen in Table 3. The fastest is the *CLRP* classifier.

5 Conclusions and Future Work

Novel features the Local Rank Patterns were introduced and the performance of these features was tested on the frontal face detection task. The results show that classifiers with the LRP achieve better precision of detection than classifiers

Fig. 3. The Receiver Operating Characteristics of the classifiers *CLRP*, *CLRD* and *CHaar* on the three test sets. top - MIT+CMU; middle - GROUP1; bottom - GROUP2
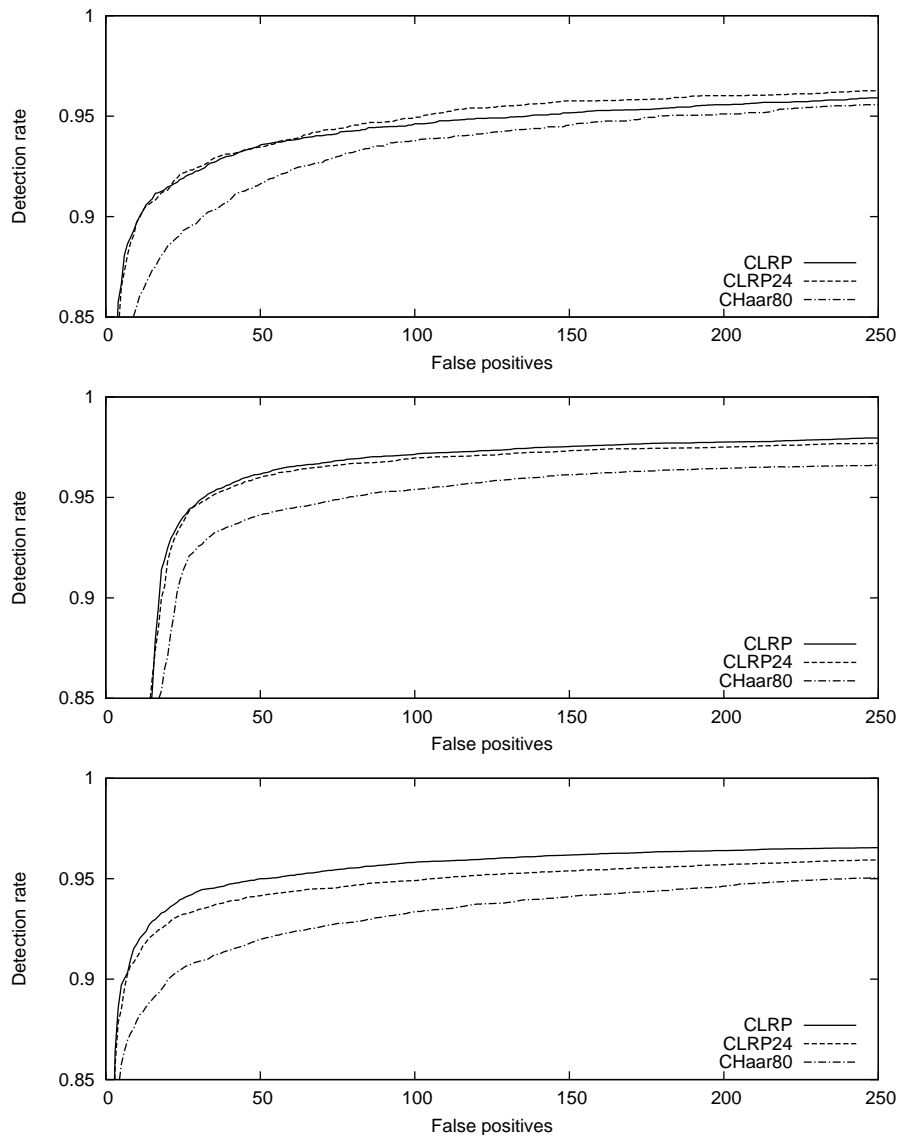
Fig. 4. The Receiver Operating Characteristics of the classifiers *CLRP*, *CLRP24* and *CHaar80* on the three test sets. top - MIT+CMU; middle - GROUP1; bottom - GROUP2

Table 3. The average speeds of the tested classifiers. The reported numbers are average number of weak hypotheses needed for decision.

Classifier	# of weak hypotheses per window
CLRP	4.8
CHaar	5.4
CLRD	6.0
CLRP24	5.0
CHaar80	6.5

with the Haar-like features and classifiers with the recently proposed LRD. The speed of the classifiers is also improved with the LRP. These results are promising and clear path to fast and precise detection classifiers on platform such as GPU and FPGA. However, question still remains how well do the results generalize to other types of objects and other learning algorithms.

Acknowledgements

References

1. Viola, P., Jones, M.: *Rapid object detection using a boosted cascade of simple features.* In: CVPR, 2001
2. Chen, J. et al.: *Modification of the AdaBoost-based Detector for Partially Occluded Faces.* In: ICPR '06: Proceedings of the 18th International Conference on Pattern Recognition, 2006.
3. Xinyi Cui et al.: *3D Haar-Like Features for Pedestrian Detection.* In: Proceedings of IEEE International Conference on Multimedia and Expo, 2007.
4. Matas, J., ochman, J.: *Wald's Sequential Analysis for Time-constrained Vision Problems.* Lecture Notes Electrical Engineering, Springer US, 2008.
5. Schneiderman, H.: *Feature-Centric Evaluation for Efficient Cascaded Object Detection.* IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, June, 2004.
6. Sochman, J., Matas, J.: *Learning A Fast Emulator of a Binary Decision Process.* In ACCV 2007.
7. Zemcik, P., Zadnik, M.: *AdaBoost Engine.* Proceedings of FPL 2007.
8. Polok, L., Herout, A., Zemcik, P., Hradis, M., Juranek, R., Josth, R.: *"Local Rank Differences" Image Feature Implemented on GPU,* accepted for publication in: Proceedings of Advanced Concepts for Intelligent Vision Systems (ACIVS 2008), 2008
9. Sochman, J., Matas, J.: *WaldBoost — Learning for Time Constrained Sequential Detection.* In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Volume 2

10. Zemcik, P., Hradis, M., Herout, A.: *Local Rank Differences - Novel Features for Image Processing,* In: Proceedings of SCCG 2007, Budmerice, SK, 2007, p. 1-12