Automatic Photo Selection for Media and Entertainment Applications

Ekaterina Potapova, Marta Egorova, Ilia Safonov
Moscow Engineering Physics Institute (National Nuclear Research University), Moscow, Russia
ep.slip at gmail.com, marta.egorova at gmail.com, ilia.safonov at gmail.com

Abstract

We propose an algorithm for automatic photo selection for media and entertainment applications like photobook and slide-show. The technique comprises three main steps: photo quality estimation and elimination of poor-quality photos, adaptive quantization of survived photos in time-camera plane, and selection of the most appealing photos from each quantized group. For detection of low-quality photos complex classifier comprising of two AdaBoost classifiers committees is created. Photos with exposure defects, such as over- and underexposed, backlit, blurred photos as well as images affected by strong JPEG artifacts are detected confidently. For quantization of photos the method similar to median-cut color quantization is proposed. The appealing photos are selected basing on novel scheme via comparison of visual salience among several images as well as face detection. Our method of identification of the most salient photo among others is based on Itti-Koch-Niebur algorithm of saliency map building. Obtained results of selection as well as time performance issues are discussed. The majority of observers were pleased with the results of the algorithm.

Keywords: *low-quality photo detection, automatic photo selection, visual salience.*

1. INTRODUCTION

At present time almost all persons have digital photo camera and capture a thousand photos. Frequently some individuals use several cameras simultaneously, for example, Digital Still Camera (DSC), camera in mobile phone, camcorder. Browsing and viewing very large photo collections are exhausting work. How to keep order in such huge heap of photos? How to select good photos for slide-show, printed photobook and similar applications? Manual selection from huge set of photos takes a long time and it is tiresome. Researchers from HP Labs [1] remark the following: "people do not create as many photobooks as they would like to, one of the reasons being the image selection process is too painful in the current digital photography landscape, in which hundreds or even thousands of photos are taken in one single event". Accordingly automatic photo selection for media and entertainment applications is a topical task.

Traditional album with printed photos, printed photobook, Web-album, slide-show for PC or digital photo frame, DVD slide-show and so forth relate to media and entertainment applications, which telling about events such as travelling or party with friends and relatives. It is required for such applications to select M photos from N, where $N>M$, sometimes $N>>M$ in order to target a specific final photo count, while preserving a good coverage of the event as well as selection of high-quality photos only. In given paper we solve the problem of selection about 100-200 photos among collection from 500-1000 photos or selection of 20-30 photos from several hundreds. In the scope of the task additional sub-tasks occur for example selection of several representative photos for cover of photobook or DVD, selection of K photos for each page of photobook, which are related to the same event, etc. Fig. 1 demonstrates collection from 30 photos. We use the collection for demonstration of automatic selection of 10 photos by proposed algorithm. The selected photos are outlined by red dot line.

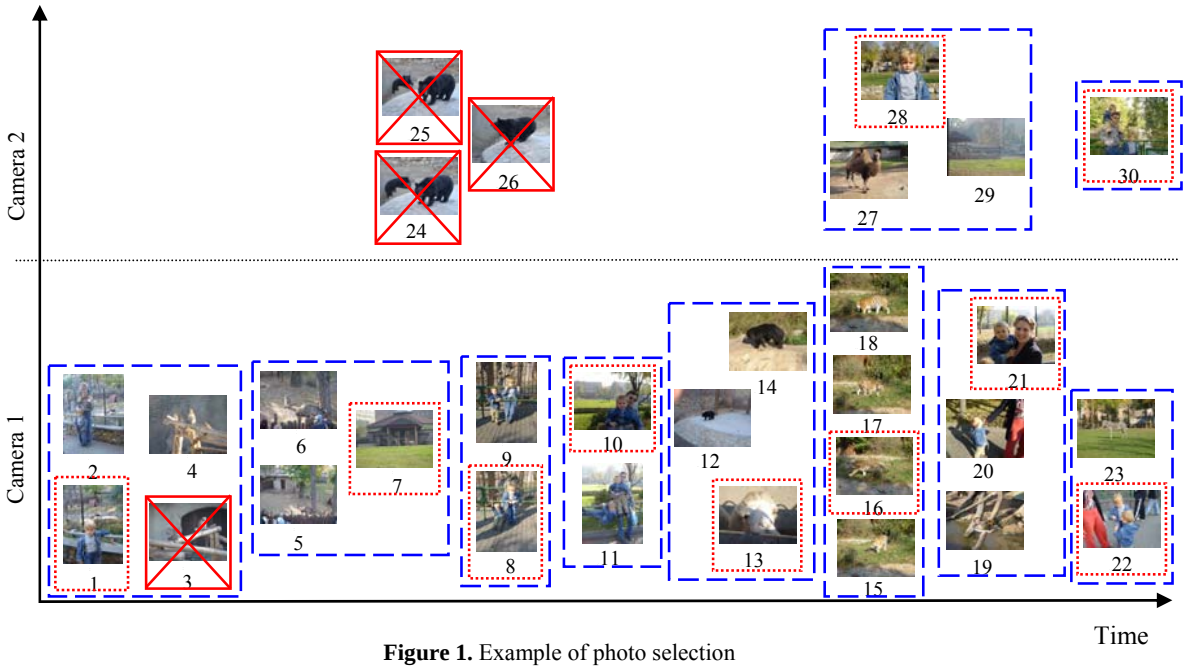


Figure 1. Example of photo selection

High photo processing performance is an important challenge for such kind of applications. The thousand photos can be processed and total selection procedure should take acceptable time, for example several minutes on modern PC. That means we should reach processing time for one photo less or at least about 1 second.

2. RELATED WORKS

Just appeared paper [1] discusses the same problem: automatic selection of images for photobook. The paper describes hierarchical time clustering, which is traversed at a specific hierarchy level in order to select images by alternating among all time clusters, and selecting the most relevant images in that cluster. The relevance ordering is based on a combination of features such as detected faces and smile, image appeal measures, where the measures such as sharpness, contrast, colorfulness, homogeneity are calculated for segmented regions. Exclusion of duplicate photos is based on the time analysis. In general the algorithm looks reasonable but it has a lot of non-obvious heuristic parameters and it looks too complex for fast implementation.

At present time for photo browsers and media applications clustering, which is based on the time of photo, where time is extracted from EXIF, is a common approach [2, 3, 4]. In addition these techniques try to exploit content-based information. In [5] blurred, underexposed and overexposed photos are excluded from analysis automatically. However other types of low-quality images are not considered. Time-based clustering is used to select photos for a slideshow.

Paper [6] is devoted to collage creation including automatic photos selection for the collage. The described algorithm is realized in impressive software application MS Research AutoCollage. The representative images are selected in three different ways: textually "interesting", mutually distinct and presence of faces in the image. "Interestingness" of the image is assessed applying entropy of histograms *ab* in color space *Lab*, mutual difference in distance between their histograms *ab*. Images with greatest entropy are interesting. However, in our opinion it is not right to consider informativeness from the viewpoint of the information theory. For example, it is known that image entropy increases with growing of image noises. Thus, images with high level of noises are selected.

3. AUTOMATIC PHOTO SELECTION

3.1 General workflow

In our algorithm we try to inherit positive trends from prior-art and overcome disadvantages of existing methods. The technique comprises three main steps: detection of low-quality photos and excluding of these photos from further processing, adaptive quantization of survived photos in time-camera plane, and selection the most appealing photos from each quantized rectangle (see fig. 2).

For detection of low-quality photos we use machine learning, namely complex classifier which comprises of consecution of one simple threshold classifier and two AdaBoost classifiers committees. Photos with exposure defects, blurred and images affected by strong JPEG artifacts are detected fast and confidently. For adaptive quantization we propose the method similar to median-cut color quantization algorithm [7]. For each quantized rectangle required number of photos is selected. Our hypothesis is the following: the most salient photo in sense of human vision model is more appealing for people. Until now, universal model of human vision did not exist, but pre-attentive vision model based on feature integration theory is well-known [8, 9]. We propose novel scheme for photos ranking based on visual salience. As a rule, presence of humans on a photo increases the appeal of the image and so we use face detection for strengthening of salience for photos with faces. For time optimization we work with downsampled versions of original photos during all steps of the algorithm.

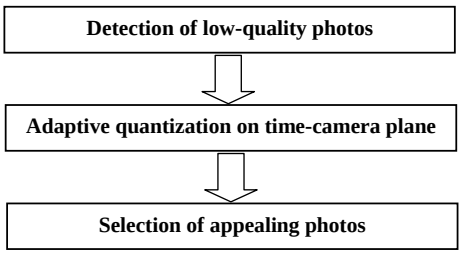


Figure 2. General workflow of proposed algorithm.

3.2 Detection of low-quality photos

Our investigations have revealed that up to 25 % of user's photos have serious defects such as blurriness, noise, compression artifacts, color misbalance as well as various types of exposure defects. Part of these images can be corrected but another part is irreparably defaced. Obviously such low-quality photos should be excluded from further processing. We consider fast algorithms only because of processing time should be extremely low. General scheme of our detection of low-quality photos algorithm is shown on fig. 3.

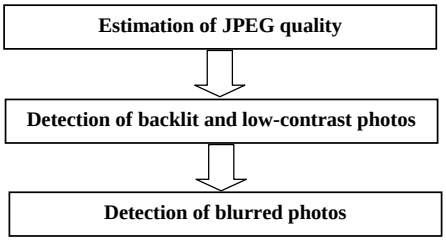


Figure 3. Scheme of detection of low-quality photos.

Sometimes photos are affected by color cast, for example, when indoor scene illuminated by an incandescent lamp is photographed. The simplest and fast color cast detection method is based on "gray world" assumption, i.e. averages in red, green and blue channels are equal. In [10] various information on a huge data set of 4.8 million photos is analyzed. Of those photos, only 74% have no dominant color that supports the general thesis about "gray world". In general it is hard to distinguish between intentional color cast and cast caused by color misbalance. Therefore exclusion of photos with any dominating color from collection is unwise.

Another important factor affecting image quality is noise. We could not find fast and reliable algorithm for noise level estimation that would provide adequate results for real-world photos. Approaches like described in [11] confuse textures with noise; error rate can reach 30% according to our experiments. It is inapplicable. More comprehensive approaches provide slightly better results but require significantly more time. Fortunately

modern DSCs have comprehensive noise reduction schemes and as a rule high noise presents on dark under-exposed photos, which can be detected easily.

Frequently compression artifacts look as noise. Absolute majority of user's photos are in JPEG format. In [12] a filter for deblocking and deringing was proposed. For adjustment of filter parameters the top left corner of the square of 3x3 quantization table of brightness channel is analyzed:

$$K = \frac{1}{9}\sum_{i,j=1}^{3} q_{i,j}.$$

Our research has shown that this metric correlates with visual assessment of JPEG images better than the compression ratio, which is strongly dependent upon the content. Figure 4 demonstrates plot *K* vs. compression ratio for high-quality photos and images with irritating artifacts level. *K* allows to separate good-quality JPEG images from JPEGs with strong artifacts. Our experiments have shown the best threshold for detection JPEG images with irritating artifacts is $K = 6.5$. The approach does not require any image content analysis and is extremely fast.

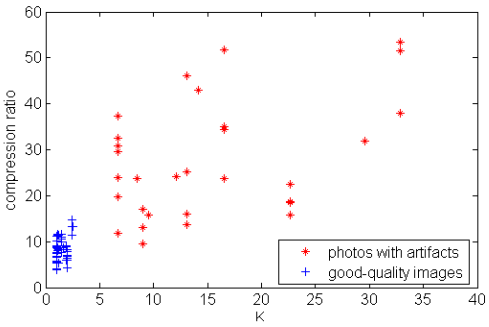


Figure 4. *K* vs. compression ratio.

There are a lot of image selection approaches where overexposed and underexposed photos are excluded from further processing. However modern DSCs have sophisticated algorithms; under-exposed and especially over-exposed photos happen rarely. More often photos damaged by backlighting can be found. Such photos have low dynamic range in dark areas. The comprehensive review of the problem was done in [13]. *Ibidem* the correction algorithm was described, but part of backlit images is irreparably defaced and correction produce unsightly outcomes.

Paper [13] describes features based on brightness histogram analysis and decision tree for adjusting correction parameters. We have repeated investigation of these features for our own test set and have found out that thresholding several of the features allows to detect backlit photos with probability 0.6-0.8. Combination of the simple classifiers and classifier for detection of low-contrast photos in AdaBoost committee as it is shown on fig. 5 allows to build classifier with high detection rate.

There are several AdaBoost algorithms which differ in approaches for optimization of weights w_i. Some realizations of these algorithms are capable to adjust thresholds of simple classifiers. We used GML AdaBoost Matlab Toolbox for feature selection and building of classifiers committee. GML AdaBoost Matlab Toolbox is a set of Matlab functions and classes, which implement Real AdaBoost, Gentle AdaBoost and Modest AdaBoost techniques. Real AdaBoost is the generalization of a basic AdaBoost algorithm first introduced by Fruend and Schapire [16]. Gentle AdaBoost [17] is a more robust and stable version of Real AdaBoost. In our case Gentle AdaBoost performs slightly better than Real and Modest AdaBoost [18].

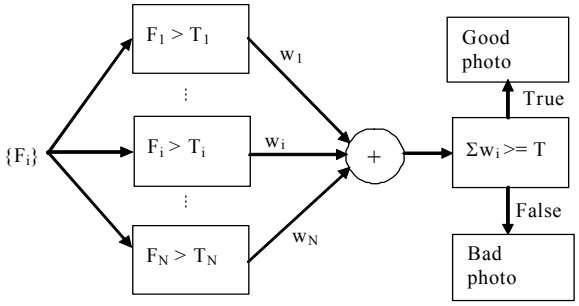


Figure 5. Scheme of AdaBoost classifiers committee.

The following features, which are calculated from brightness histogram *H* for image size *MxN* and color depth of brightness 8 bpp, were selected for classifiers committee:

S_1 / S_2 - ratio of tones in shadows to midtones, S_{11} / S_{12} - ratio of tones in first to second part of shadows, M_1 / M_2 - ratio of the histogram maximum in shadows to global histogram maximum, P_1 - location of the histogram maximum in shadows, *C* – global contrast, where

$$S_1 = \sum_{[0,85]} H(i)/(M \times N),\ S_2 = \sum_{(85,170]} H(i)/(M \times N),$$

$$S_{11} = \sum_{[0,42]} H(i)/(M \times N),\ S_{12} = \sum_{(42,85]} H(i)/(M \times N),$$

$$M_1 = \max_{[0,85]}(H(i))/\max_{[0,\ 255]}(H(i)),\ M_2 = \max_{(85,170]}(H(i))/\max_{[0,\ 255]}(H(i)),$$

$$P_1 = l \mid H(l) = \max_{[0,85]}(H(i)),$$

$$C = high - low,$$

$$low = \min(\min\{i \mid H[i] \ge H_0\}, \min\{i \mid \sum_{k=0}^{i} H[k] \ge C_0\}),$$

$$high = \max(\max\{i \mid H_R[i] \ge H_1\}, \max\{i \mid \sum_{k=i}^{1} H_R[k] \ge C_1\}),$$

where H_0, H_1 and C_0, C_1 are threshold values for histogram area and intensity correspondingly.

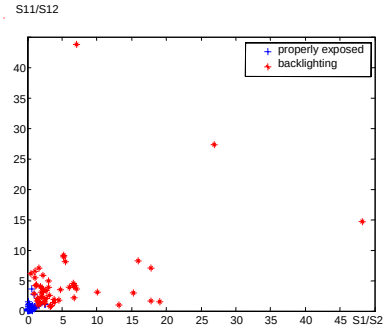


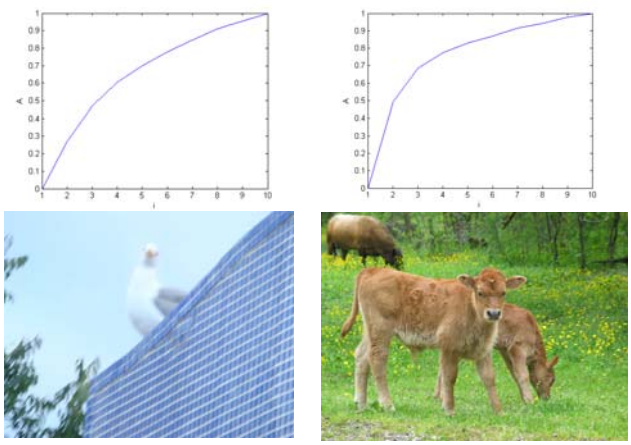
Figure 6. *S1/S2* vs. *S11/S12* for properly exposed and backlit.

The plot on fig. 6 demonstrates distribution of properly exposed and backlit photos on plane features *S1/S2* and *S11/S12*, backlit

photos can be detected with high probability based on these features. Our training set contains 188 photos with various exposure defects and 292 high-quality photos. Error rate on cross-validation test is about 0.055. Our testing set contains 1830 photos but only about 2% of the photos have low-contrast or affected by backlit. The number of False Positives (FP) is 10, number of False Negatives (FN) is 3. As a rule False Positives are night shots with wide dark background.

The proposed technique is very fast because only features calculated from histogram are used and classifiers committee is very simple in sense of computational complexity.

Blurriness is one of the most common image defects. It can be caused by mistake of focusing or camera shaking. Paper [14] proposes non-reference automatic sharpness level estimation, which is based on analysis of variations of edges histograms, where edge-images are produced by high-pass filters with various kernel sizes, array of integrals of logarithm of edges histograms characterizes photo sharpness. Let's consider that approach in more details. On the first step gray channel I of initial image is scaled to destination size according to viewing or printing conditions.



a) b)

Figure 7. Array $\{A_i\}$ for blurred (a) and sharp (b) photos.

AdaBoost classifiers committee is applied to detect out-of-focus photos. Further I is filtered by set from 10 high-pass filters with convolution kernels Z_i [-1 1], [-1 0 1], [-1 0 0 1] … [-1 0 0 0 0 0 0 0 0 1]:

$$E_i = I \otimes Z_i$$

For each abs(E_i) histogram He_i is calculated. The entropy characterizes the 'flatness' and 'peakedness' of histogram, but value of entropy strongly depends on total number and magnitude of edges that depends on photo content. It was proposed to normalize entropy by dividing by a number of edges for all edge magnitudes:

$$A_i = \sum_k \frac{-He_i(k)\log(He_i(k)+1)}{-He_i(k)} = \sum_k \log(He_i(k)+1).$$

Array $\{A_i\}$ varies for blurred and sharp photos. Figure 7 demonstrates $\{A_i\}$ for such photos as well as corresponding photos itself.

Thereupon several features $\{F_i\}$ to characterize $\{A_i\}$ were formulated and AdaBoost classifiers committee was constructed.

We propose to boost the committee by means of addition Crete's sharpness metric [15] as one more feature. Crete's non-reference sharpness estimation is based on the idea that a high variation between the original and the blurred image means that the original image was sharp whereas a slight variation between the original and the blurred image means that the original image had been already blurred.

In our committee the following features are applied:

$$F_1 = An_3 - An_2,\ F_2 = \sum_{i=2}^{11} An_i,\ F_3 = A_2,$$

where An_i elements are A_i normalized to [0, 1] range by means of dividing on max(A_i).

The fourth feature F_4 is similar to Crete's sharpness metric but is calculated for rows only:

$$F_4 = (SDI - SV_h)/SDI,$$

$$B_h = I \otimes LPF,\ DI = I \otimes HPF,\ DB_h = B_h \otimes HPF,$$

$$SDI = \sum_{\forall r,c} DI(r,c),$$

$$SV_h = \sum_{\forall r,c} (DI(r,c) - DB_h(r,c)) \times \frac{1}{1+e^{-100\times(DI(r,c)-DB_h(r,c))}},$$

where LPF is a low-pass filter with convolution kernel [1 1 1 1 1 1 1 1 1 1]/9, HPF is a high-pass filter with convolution kernel [1 -1], (r,c) are coordinates of image pixels.

Our training set contains 205 blurred photos and 311 high-quality photos. The error rate on cross-validation test is about 0.07. Our testing set of 1830 photos contains 171 blurred photos. The number of FP is 34, the number of FN is 10.

The processing time for classification of one photo on sharp or blurred classes is about 0.4 c. Whole classification cascade takes about 0.5 c for one photo, including downsampling to target printing or viewing size.

3.3 Time and camera-based quantization

In order to divide whole collection on M groups, where one photo is selected from each group, various quantization algorithms are intended. Time-based clustering or quantization provides satisfactory outcomes in event coverage sense in cases where all photos are photographed by just one camera only. Papers in prior-art do not discuss image selection from collection of photos captured by various cameras. Is it typical for users to collect photos from several cameras for one and the same event? We conduct user study to define whether amateurs collect images from several cameras for one storytelling application. Survey participants were asked three questions:

1) Do you have some photos for one event captured by several cameras at your paper or web photo album?

2) How many events from your album are captured by several cameras?

3) What are your arguments to collect photos from different cameras?

31 participants assisted in our survey. Albums of 83% of the users contain event photos captured by more than one camera and portion of such events is about 30%. Fig. 8 presents answers distribution for third question. So existence application of several cameras for one event is a widespread practice.

In case of usage of several cameras time quantization only is ineffective sometimes due to imperfection of cameras time synchronization and the same event can be presented on photos

with different time in EXIF. Usually mistiming is about several minutes, but can be equal to years. We propose quantization in 2D plane, where the first axis is time and the second axis is a camera name obtained from EXIF information. If EXIF is absent, then all images are merged into a separate virtual camera.

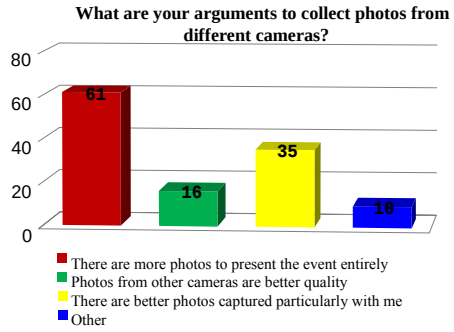


Figure 8. Motives to collect photos from different cameras.

Fig. 9 illustrates rule for time-camera plane creation. Cameras are sorted in ascending order according to the number of photos in collection. *L* is the time between the least and the biggest time for the camera with the largest number of photos. Yp_i coordinates on Ticks on *Camera* axis are calculated as follows:

$$Yp_i = \begin{cases} H \times (i+1)/2: & i \text{ is odd} \\ H \times (Nps+1-i/2): & i \text{ is even} \end{cases}$$

where *i* is index of camera, *Nps* is number of cameras,

$$H = 2 \times L / M .$$

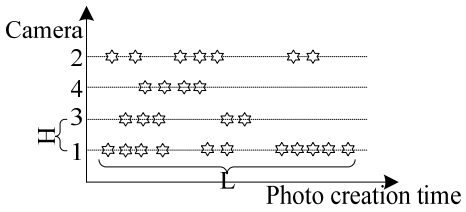


Figure 9. Time-Camera 2D plane

The general idea of the proposed quantization algorithm is the following: each group should contain approximately equal number of photos. The idea is similar to well-known Heckbert's color quantization technique [7]. In present time there is a trend to use modifications of the algorithm for plenty of tasks, for example for light probe sampling [23]. Taking inspiration from the median cut algorithm we can partition time-camera plane in the rectangular regions as follows:

1. Calculate bounding boxes for regions;
2. For region with the greatest number of photos divide on to sub-regions along the longest dimension such that the new subregions contain approximately equal number of photos;
3. If the number of current regions is less than number of the required groups, then return to step 1.

In addition we introduce limitation on minimal dimension of region: if the longest dimension is less than *Tr* then the region is not divided. The condition is intended for combining of duplicate photos in one group. The example of partitioning is shown on fig. 1. Further for each group one of the most appealing photo is selected.

3.4 Salient photo selection

We assume that the most appealing and relevant photo in a set is the most noticeable. This assumption leads us to conclusion that the most appealing photo is the most salient photo. General approach for construction of saliency map is described in Itti-Koch-Niebur [8, 9] papers. Usually saliency map is used for tasks related to pre-attention vision and scene analysis in order to determine the most valuable parts of the image or the scene, so-called regions of interest (ROI) [19]. Until now saliency map did not apply for comparison of images with each other; we propose the appropriate way for selection of the most salient photo.

The schema of saliency map building is shown on fig. 10. Every image in the set has red (r), green (g) and blue (b) channels. Intensity map is obtained as:

$$I = (r+g+b)/3 .$$

Four color channels *R, G, B, Y* are created from *r, g, b* in the following way:

$$R = r - \frac{g+b}{2}, \quad G = g - \frac{r+b}{2},$$

$$B = b - \frac{r+g}{2}, \quad Y = \frac{r+g}{2} - \frac{|r-g|}{2} - b.$$

For *I, R, G, Y* 8-level Gaussian pyramids are constructed using Gauss separable filter with convolution kernel [1 5 10 5 1]. From intensity map 8-level Gabor pyramids for different orientations $\theta \in \{0,45,90,135\}$ are created to obtain local orientation information. We compute 42 feature maps using center-surround difference:

$$I(c,s) = |\, I(c) - I(s) \,|$$
$$RG(c,s) = |\, (R(c) - G(c)) - (G(s) - R(s)) \,|$$
$$BY(c,s) = |\, (B(c) - Y(c)) - (Y(s) - B(s)) \,|$$
$$O(c,s,\theta) = |\, O(c,\theta) - O(s,\theta) \,|$$

where $c \in \{2,3,4\}$ and $s = c + \delta,\ \delta \in \{2,3\}$.

All feature maps are normalized using local maximum technique and combined into conspicuity maps using across-scale addition:

$$\bar{I} = \sum_{c=2}^{4} \sum_{s=c+3}^{c+4} N(I(c,s))$$

$$\bar{C} = \sum_{c=2}^{4} \sum_{s=c+3}^{c+4} [N(RG(c,s)) + N(BY(c,s))],$$

$$\bar{O} = \sum_{\theta \in \{0,45,90,135\}} N\Big(\sum_{c=2}^{4} \sum_{s=c+3}^{c+4} N(O(c,s,\theta))\Big),$$

where N() is a normalization operator which increases strong peaks and decrease noise.

Normalization operator consists of two parts. In the first part Gaussian filter is applied to the image in order to decrease noise. In the second part average local maximum is computed and the whole image is multiplied by the difference of the maximum value on the image and local maximum value. This operation helps to prevent strong but individual peaks and also helps not to take into account such things as very bright background.

So, we obtain the following conspicuity maps: $\bar{I}$ for intensity, $\bar{C}$ for color, $\bar{O}$ for orientation. Conspicuity maps are summed with specific weights into final image which is called saliency map:

$$S = \frac{weightI \cdot N(\bar{I}) + weightC \cdot N(\bar{C}) + weightO \cdot N(\bar{O})}{weightI + weightC + weightO} .$$

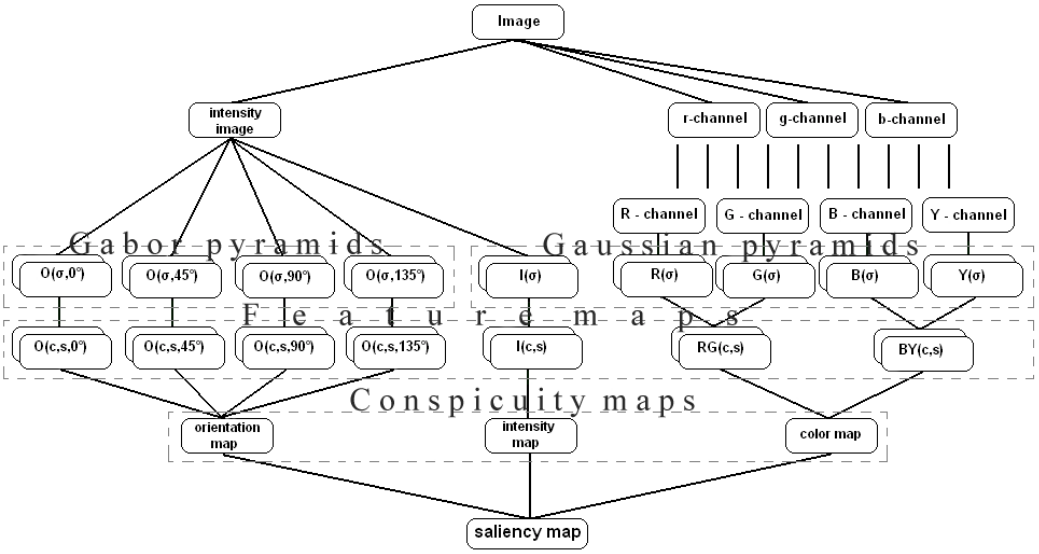


Figure 10. The schema saliency map building.

The main problem is to find right weights cause due to normalization different conspicuity maps have different contribution to final result. The majority of the previous works consider to sum conspicuity maps in equal proportions which in our opinion is not completely right. To solve this problem we considered to make an experiment. This experiment as input data has a number of images (normally from 30 to 50 images). For each picture in the set the most salient regions were marked by several experts. In order to determine the best weights we were finding maximum of the following function using simplex algorithm:

$$\sum_{p \in ROI} S(p) \to \max$$

where ROI are noted areas on the image, $p \in ROI$ and $S(p) \geq S_{\max} / 4$.

Mathematical expectations of weights were calculated after finding values for every image in the set.

Experiment has shown that weights locate in the following ranges:

$$weightI = 0.2..0.5,\ weightC = 0.4..0.6,\ weightO = 0.2..0.5.$$

Specific values depend on the person and his perception of the surrounding world, his preferences and features, everyone can choose what he or she likes more. Fig. 11 demonstrates the photo and its conspicuity maps as well as the final saliency map (weightI = 0.5, weightC = 0.25, weightO = 0.3).

The last step was to find a criterion which ranges photos in the set and gives clear answer what photo is the most salient among others. "Saliency Index" *SI* is counted as following:

$$SI = \sum S(x,y)/(w \cdot h)$$

where $S(x,y) \geq S_{\max} / 4$, w is image width, h is image height.

This criterion was applied to different photo sets and it was found that itproduced appropriate results. Thr example of how photos are ranged by *SI* can be seen on fig. 12. The photos "camel" and "boy" have close values of *SI* but camel's *SI* is a little bit greater.

Algorithm works less than 1 second for color images with size 500 x 500. Processing time can be decreased considerably due to parallel calculation on GPU as it is described in [20].

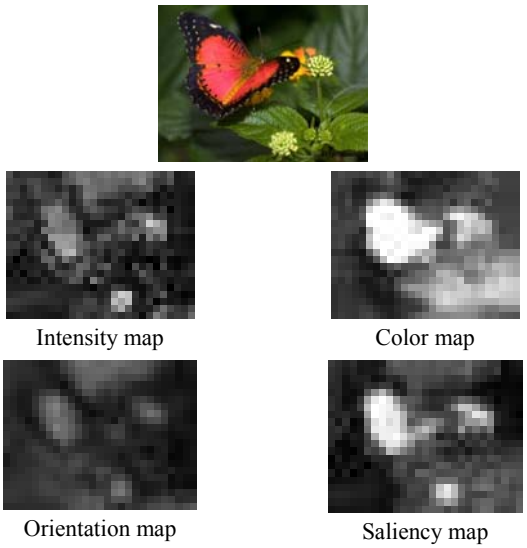


Figure 11. Photo and its conspicuity and saliency maps.

3.5 Face detection

Images of people prevail in a lot of amateur photo collections. This kind of photos attracts attention more than images without humans. A face detection algorithm can be used for search of human presence. Face processing is a rapidly expanding area and a lot of researches have been conducted in recent years. One of the acknowledged algorithms for face detection is the one developed by Viola and Jones [21]. The Intel OpenCV library provides an efficient implementation of the Viola-Jones face detector.

We analyzed implementation of Viola-Jones algorithm in OpenCV library for typical user's photos with sizes from 4 to 6 Mpix and found out the following main disadvantages: average number of FP is more than 3; the processing time is more than 5 seconds for modern PC. The face detection outcomes of initial

OpenCV version is shown on the photo “boy” on fig. 12. The face is detected correctly but two places were detected erroneously.

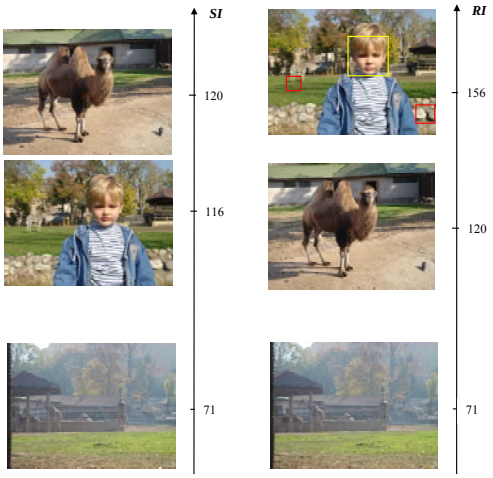


Figure 12. Ranking results by *SI* and summary index *RI*.

We have carried out some modifications to improve the algorithm efficiency. Specifically, conducted modifications add up to the following: the optimal image downsampling is applied at preprocessing step for speed increasing; optimization of search region using color information is performed. Detailed information about the experiment conditions, obtained data and comparative results were described in [22]. There are no False Positives on photo “boy” after our modifications.

Detection time is about 1 second; now it is greater than it is necessary of our task. In particular it is connected with extra conversion to internal OpenCV structures and other programming issues. We expect to reach 0.2-0.3 s time for processing of image with size 850x640 on PC.

3.6 Photos ranking

We propose to combine saliency index *SI* and face detection outcomes for calculation of summary appealing index *RI* as follows:

$$RI = SI + w \times \sqrt{NF} \ ,$$

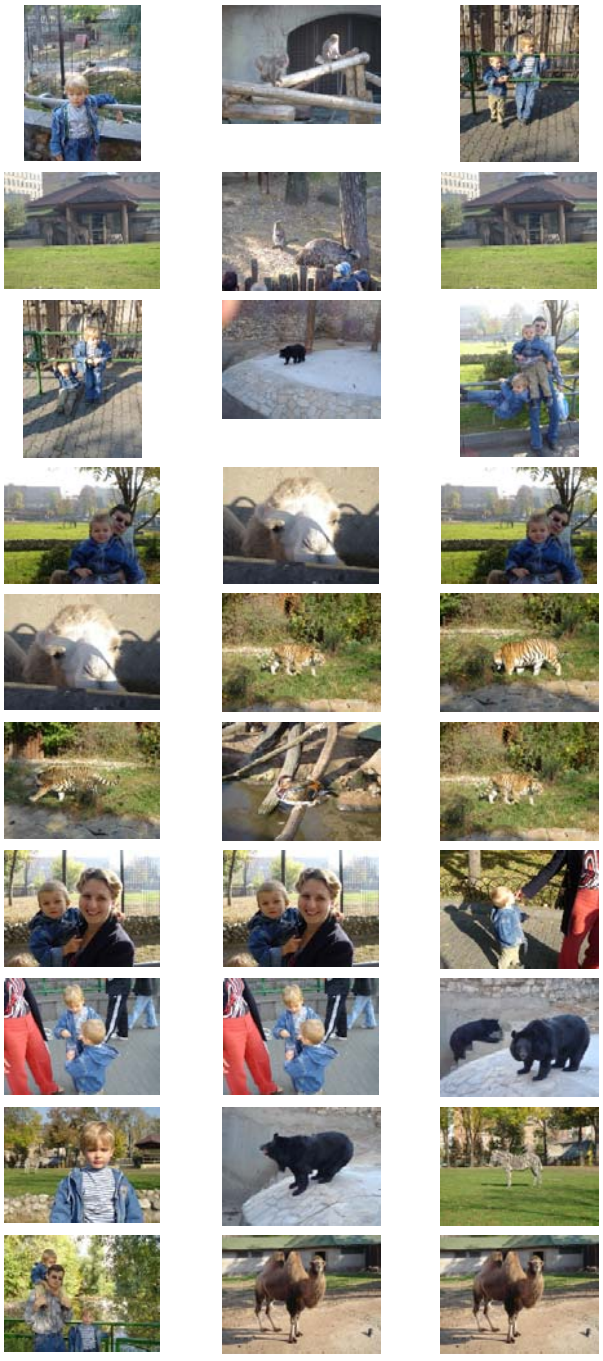
where *NF* is the number of detected faces, *w* is weight.

The heuristic formula and preferable *w*=25 value were obtained during plenty of experiments. The final ranking results by *RI* are shown on fig. 12. Accordingly photo “boy” is selected as the most appealing from the group of three photos.

4. RESULTS AND DISCUSSION

The set of 30 photos captured by two cameras is shown on fig. 1.

Let 10 photos should be selected. At the first stage low-quality photos are detected. For given set four poor images were detected. These photos are blurred actually and they are excluded from further processing. On fig. 1 excluded images are crossed out by solid red line. Next stage is median-cut-like quantization on time-camera plane. The 10 groups, which are the result of quantization, are outlined by blue dash line. The final stage is selection of the most appealing photo among images of each group. The selected 10 photos are outlined by red dot line. The owner of the collection evaluates achieved outcomes as accurate: 6 photos coincide with manual selection by the expert (# 1, 8, 10, 21, 28, 30 on fig. 1), 3 photos are considered as acceptable (# 13, 16, 22) and only one photo (# 7) the expert counts as uninteresting.



a) b) c)

Figure 13. Results of photo selection by proposed technique (a), MS AutoCollage (b) and random selection (c).

Unfortunately the majority of existing solutions for automatic photos selection is inaccessible for testing, but we had possibility to compare our selection results with outcomes of MS Research AutoCollage (ver. 1.1.2009.0130) and with simple random selection. The function of AutoCollage software application is automatic creation of photo collage and the first stage is selection

of photos from a collection. Two photos selected by AutoCollage (# 3, 26) are blurred and expert counts their as inacceptable. Only two photos (# 19, 21) coincide with manual selection; other 6 photos (#6, 12, 13, 17, 22, 27) are considered as acceptable.

So ground truth is manual selection by the owner of photo collection and his/her expertise. In our opinion the number of inacceptable photos in selected set is principal criterion. For estimation of efficiency of proposed technique we have processed 5 sets of photos; each set contains 30 photos. Table 1 reflects obtained results.

TABLE 1 RESULTS OF AUTOMATIC SELECTION FOR 5 SETS.

		Set 1	*Set 2*	*Set 3*	*Set 4*	*Set 5*	**Sum**
Proposed	Agree with expert	6	5	6	5	7	**29**
	Acceptable	3	4	4	4	2	17
	Inacceptable	1	1	0	1	1	**4**
AutoCollage	Agree with expert	2	2	2	6	5	**17**
	Acceptable	6	7	7	0	4	24
	Inacceptable	2	1	1	4	1	**9**
Random	Agree with expert	2	2	3	4	4	**15**
	Acceptable	5	5	4	2	5	21
	Inacceptable	3	3	3	4	1	**14**

Both tested solutions demonstrate high efficiency and good coverage of event. Random selection demonstrates serious errors. Automatic selection is capable to improve creation of photo album for media and entertainment applications such as photobook and slide-show.

Sometimes AutoCollage selects low-quality images whereas proposed algorithm has no such drawback. In AutoCollage number of FP on face detection stage is high enough. The number of FP in our face detection module is less in 2 times with preserving the same detection rate. As regards processing time, AutoCollage spends about 1.1 s per image. Our method is a little bit slower; it spends about 1.4 s per image. Bottleneck is face detection module. We expect to speed up face detection in the future.

5. REFERENCES

1. P.Obrador, N.Moroney, "Automatic Image Selection by means of a Hierarchical Scalable Collection Representation", Proc. of SPIE-IS&T Electronic Imaging, 2009.
2. J.C. Platt, M. Czerwinski, B.A. Field, "PhotoTOC: Automatic Clustering for Browsing Personal Photographs", Proc. IEEE Pacific Rim Conf. Multimedia, pp. 6-10, 2003.
3. A.Graham, H.Garcia-Molina, A.Paepcke, T.Winograd, "Time as essence for photo browsing through personal digital libraries", Proc. of the 2nd ACM/IEEE-CS conference on Digital libraries, 2002.
4. D.F.Huynh, S.M.Drucker, P.Baudisch, C.Wong, "Time Quilt: Scaling up Zoomable Photo Browsers for Large, Unstructured Photo Collections", CHI'05 extended abstracts on Human factors in computing systems, 2005.
5. J.Chen, W.Chu J.Kuo, C.Weng, J.Wu, "Tiling Slideshow: An Audiovisual Presentation Method for Consumer Photos", In Proc. of ACM Multimedia Conference, pp. 36-45, 2007.
6. C.Rother, L.Bordeaux, Y.Hamadi, A.Blake, "Autocollage", ACM Transaction on Graphics, SIGGRAPH-2006, vol 25, pp. 847-852, 2006.
7. P.Heckbert, "Color Image Quantization for Frame Buffer Display", SIGGRAPH-1982, ACM Press. pp. 297-207, 1982.
8. L.Itti, C.Koch, E.Niebur, "A model of saliency-based visual attention for rapid scene analysis", IEEE Transactions on Pattern analysis and machine intelligence, Vol. 20, No. 11, pp. 1254-1259, 1998.
9. L.Itti, C.Koch, "A saliency-based search mechanism for overt and covert shifts of visual attention", Vision Research 40, pp 1489-1506, 2000.
10. D.Wueller, R.Fageth,"Statistic Analysis of Millions of Digital Photos", Proc. of SPIE/IS&T Electronic Imaging, 2008
11. P.Meer, J.M.Jolion, A.Rosenfeld, "A Fast parallel algorithm for blind estimation of noise variance", IEEE Tran. on Pattern Analysis and Machine Intelligence, Vol. 12, No. 2, 1990.
12. A.Foi, V.Katkovnik, K.Egiazarian, "Pointwise Shape-Adaptive DCT for High-Quality Denoising and Deblocking of Grayscale and Color Images". IEEE Transaction on Image Processing, v.16, No 5, 2007.
13. I.Safonov, "Automatic correction of amateur photos damaged by backlighting", Graphicon-2006.
14. I.V.Safonov, M.N.Rychagov, K.M.Kang, S.H.Kim, "Adaptive sharpening of photos", Proc. of SPIE-IS&T Electronic Imaging, 2008.
15. F.Crete, T.Dolmiere, P.Ladret, M.Nicolas, "The Blur Effect: Perception and Estimation with a New No-Reference Perceptual Blur Metric", Proc. of SPIE-IS&T Electronic Imaging, 2007.
16. Y.Freund, R. Schapire. Experiments with a new boosting algorithm, Int. conf. on Machine Learning, pp. 148-156, 1996.
17. J.Friedman, T.Hastie, R.Tibshirani, "Additive logistic regression: A statistical view of boosting", The Annals of Statistics, 38(2), pp. 337-374, 2000.
18. A.Vezhnevets, V.Vezhnevets, Modest AdaBoost – teaching AdaBoost to generalize better, Graphicon, pp.322-325, 2005.
19. C.M.Privitera, L.W.Stark, "Algorithms for defining visual regions-of-interest: comparison with eye fixations", IEEE Transactions on Pattern Analysis and machine intelligence, vol. 22, no. 9, pp. 970-982, 2000.
20. P.Longhurst, K.Debattista, A.Chalmers, "A GPU-based saliency map high-fidelity selective rendering", Proc. of ACM AFRIGRAPH, 2006.
21. P.Viola, M.Jones, "Robust Real-time Object Detection", Compaq Technical report, 2001
22. M.A.Egorova, A.B.Murynin, I.V.Safonov, "An improvement of face detection algorithm for colour photos", Proc. of Pattern Recognition and Image Analysis, 2008.
23. P.Debevec, "A Median Cut Algorithm for Light Probe Sampling", ACM Transaction on Graphics: SIGGRAPH-2005 posters, 2005.

About the authors

Ekaterina Potapova is a 5 year student at National Nuclear Research University (Moscow Engineering Physics Institute, MEPhI).

Marta Egorova received her MS degree in cybernetics from MEPhI in 2008. At present time she is a post-graduate student of MEPhI.

Ilia Safonov received his MS degree in automatic and electronic engineering from MEPhI in 1994 and his PhD degree in computer science from MEPhI in 1997.