Animated thumbnail for still image

Ilia V. Safonov, Victor V. Bucha
Samsung Research Center,
Moscow, Russia
ilia . safonov at samsung . com, v . bucha at samsung . com

Abstract

The conventional thumbnails do not provide a user-friendly way to view and browse still images on portable devices with a small display. The issue is that a still thumbnail is too small to recognize image details such as human faces, document title and photo quality.

To cope with this problem we propose an automatic algorithm for creation of animated thumbnail from still photos and scanned documents. The approach comprises image classification on photo and scanned document, detection of attention zones, creation of animation that simulates camera tracking-in, tracking-out, panning between detected zones and whole scene.

Short looped movie demonstrates whole image, main objects of the scene one by one as well as quality of the photo. User study demonstrates effectiveness of proposed animated thumbnail in comparison with conventional still thumbnail.

Keywords: *animated thumbnail, creation video from still image, vision model, attention zones, document segmentation*

1. INTRODUCTION

The thumbnails are used widely to browse image collection. It is essential part of the user interface for the various devices, PC and Web-applications. In general, thumbnail represents down-sampled copy of original image. However, often it is difficult to recognize original image from observed thumbnail. The small details and overall image quality are undistinguished especially when a size of original image is big enough or has a complex layout (Fig. 1).

Figure 1: Conventional thumbnails on various devices.

For example, the correlation between the image of a document and its thumbnail is not obvious. To browse photo collection with the small thumbnails is not so easy too. Often user is forced to make zoom-in in order to recognize the people on it.

In the paper we propose the approach for demonstration of still image thumbnail as an animation comprising the frames with the most important zones of the image. The resulting animation looks attractive and provides a user-friendly way to view and browse images especially on small displays widely used in the mobile devices. The animated thumbnail may look like a slide-show application for full-size photos. However a thumbnail animation for document images has no well-known examples.

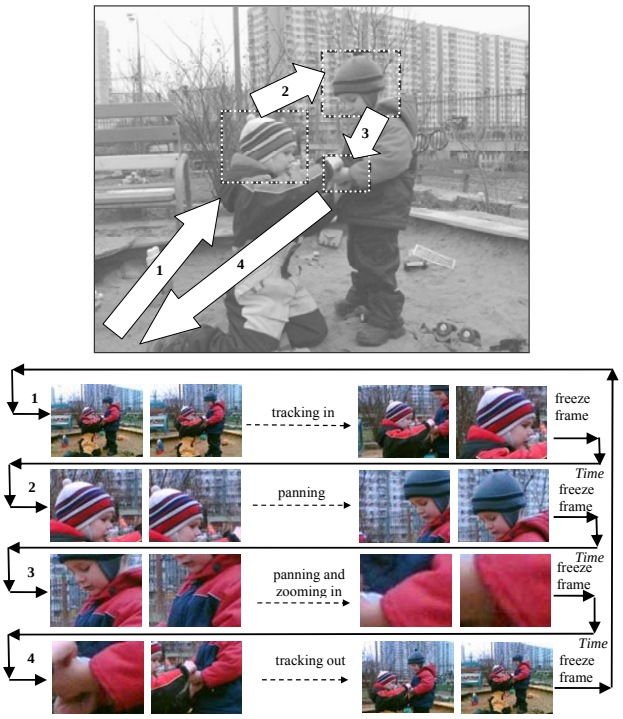


Figure 2: Frames of animated thumbnail for photo.

2. RELATED WORK

There are several approaches to create “intellectual” thumbnail as another still image. Way for creation of single still thumbnail by means of auto cropping for photo retrieval is described in [1]. The approach is based on the face detection and saliency map building. The technique has several drawbacks: the overall initial view is destroyed; it is not clear how thumbnail is created for photos with several faces as well as for photos with several salient regions; aspect ratio of images is altered.

Another automatic thumbnails creation process is discussed in [2]. The proposed approach does not modify the image composition and rather better reflects the image quality of the originals. The

main purpose of this approach is to reflect noise level and blurriness of photo on a generated thumbnail.

So called SmartNails for browsing of document images is described in [3]. SmartNail consists of a selection of cropped and scaled document segments that are recomposed in order to fit the available display space while maintaining recognizability of document image, readability of text and keeping the layout close to the original document layout. Evidently, the recognizability of SmartNail is high. Nevertheless the overall initial view is destroyed as well as sometimes layout alteration is estimated by observer ambiguously.

3. ANIMATED THUMBNAIL

3.1 General idea

In order to recognize image content as well as to estimate its quality confidently it is necessary to see zoomed-in fragments of the image. Sometimes several regions of the image should be inspected. We propose to create smooth animated transitions between *attention zones* of image. Obtained video frames are cropped from initial still image and scaled to thumbnail or display size. Further the frames are stored in appropriated animation file format (animated GIF, Adobe flash) or played on-the-fly.

Duration of the movie should not be too long. Optimal duration is less than 10 seconds. Therefore, the number of attention zones is limited to 3-5. Animation can be looped. We named such small size movie an *animated thumbnail*.

The algorithm for animated thumbnail generation has the following three key stages:

- detection of attention zones;
- selection of zone for quality estimation;
- generation of video frames which are transitions between the zones and whole image.

The attention zones are important for recognition of the image and they differ depending on image content. At least two types of still images can be selected: photos and document images. Title, headers, embedded pictures and other emphasized text elements are sufficient for identification of a document image. Human faces are adequate for identification of photos for the most part. For photos which do not contain faces the preattentive human vision model can be used for detection of visual attention zones.

There are plenty of methods for detection of faces and other visual attention zones as well as methods for document layout analysis. Below we describe one possible way for important zones detection without comparison with any alternatives. We demonstrate advantages of animated thumbnail as a concept.

It should be clear, that to select approaches for important zones detection one should take into account application scenario and hardware platform for implementation. Fortunately panning over image during animation allows to recognize image content even if important zones were detected incorrectly. So we prefer more simple techniques for attention zone selection rather than more comprehensive ones.

For visual estimation of blurriness, noise and compression artifacts observer should investigate fragment of image without any scaling or with modest magnification. We propose several simple rules for selection of appropriate fragment: the fragment should contain at least one contrasting edge and at least one flat region, histogram of the fragment should be wide enough but without clipping on limits of dynamic range. Proposed algorithm uses these rules for automatic zone selection in the central part of an image or in the attention zones.

Fig. 2 demonstrates example of animated thumbnail for still photo. Two faces are detected on the first stage. Hands of kids are selected as a zone for quality estimation. Movie consists of four transitions between whole image and these three zones. The first sequence of the frames looks like a camera tracking-in to a face; after that frame is frozen on a moment for better observer's focus on zoomed face. The second sequence of the frames looks like a camera panning between faces. The third sequence of the frames looks like a camera panning and zooming-in between face and hands; after that the frame with the hands is frozen on a moment for visual quality estimation. Final sequence of frames looks like a camera tracking-out to whole scene and freeze frame takes place again.

3.2 Image classification

Every image type has its own approach to detect areas of attention. That is why the first processing step is image classification into two types: photo or document. The classification is performed by analyzing of average *energy En* of normalized co-occurrence matrices of *rgb* channels of downsampled image:

$$\underset{\forall x,y}{C_r(i,j) = C_r(i,j) + 1 \; if \; r(x,y) = i \; and \; r(x+dx, y+dy) = j,}$$

$$\underset{\forall x,y}{C_g(i,j) = C_g(i,j) + 1 \; if \; g(x,y) = i \; and \; g(x+dx, y+dy) = j,}$$

$$\underset{\forall x,y}{C_b(i,j) = C_b(i,j) + 1 \; if \; b(x,y) = i \; and \; b(x+dx, y+dy) = j,}$$

$$N_r(i,j) = \frac{C_r(i,j)}{\sum_i \sum_j C_r(i,j)}; \quad N_g(i,j) = \frac{C_g(i,j)}{\sum_i \sum_j C_g(i,j)};$$

$$N_b(i,j) = \frac{C_b(i,j)}{\sum_i \sum_j C_b(i,j)};$$

$$En = \frac{1}{3} \sum_i \sum_j (N_r{}^2(i,j) + N_g{}^2(i,j) + N_b{}^2(i,j));$$

where *x, y* – coordinates of pixels of image, *dx, dy* – displacements.

Table 1 demonstrates energy *En* for various images. Average energy of normalized co-occurrence matrices *En* differs for photos and document images on several orders. For photos *En* is less than 0.01, whereas for typical document images *En* is greater than 0.1.

Given approach works well for both contone and halftoned photos as well as for color and grayscale photos. Application of three co-occurrence matrices for all *rgb* channels allows to detect document images with a color background unlike the application of one co-occurrence matrix of gray channel.

Naturally, documents which contain huge built-in photo or several photographic illustrations are classified as a photo. The reason of such misclassification is that the relative area of the text is small. On the other hand black and white graphic arts and business graphics are classified as an image document. Despite the misclassification detected attention zones looks pretty good in general.

TABLE 1 NORMALIZED CO-OCCURRENCE MATRICES OF RGB CHANNELS FOR PHOTOS AND SCANNED DOCUMENT IMAGES

Photos	*En*	*Document Images*	*En*
	*0.0009*		*0.3*
	*0.0001*		*0.1*
	*0.0005*		*0.2*

3.3 Attention zones detection for photo

Information about humans on photo is important to recognize the scene. Thus, it is reasonable to apply the face detection algorithm to detect attention zones in the photo. There are many methods for face detection. For example, well-known OpenCV software library contains implementation of face detection for front and profile faces. The implementation is based on Viola-Jones face detector [4]. In general the methods which are based on state-of-the-art Viola-Jones face detector provide good results. However they can't detect rotated faces confidently and find out a lot of false positives. The number of false positives can be decreased with additional skin tone segmentation and processing of downsampled image [6]. In recent years, methods for multi-view face detection were proposed, for example, such an approach is described in [5]. Fig. 3 illustrates detected faces as attention zones.

Very often face may characterize the photo very well but a lot of photos do not contain face or faces may be non-informative as they relate to random person. Thus an additional mechanism has to be used to detect zones of attention. We propose an algorithm to detect the zones of attention based on human visual model of attention. Until now, universal model of human vision does not exist, but pre-attentive vision model based on feature integration theory is well-known [7], [8].

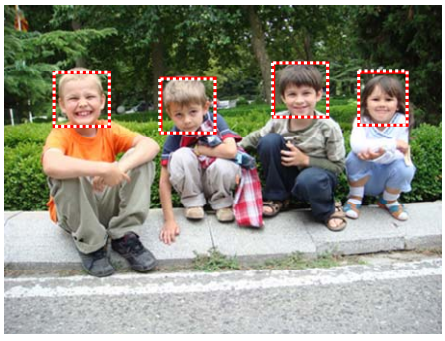


Figure 3: Example of detected zones for photo with people.

In the strict sense a model of human preattentive vision does not quite fit in this case, since the observer is on attentive stage while viewing thumbnail. However existing approaches for the detection of regions of interest are based on saliency map and they often provide reasonable outcomes, whereas the use of attentive vision model requires too much *prior* information about the scene and it is not generally applicable.

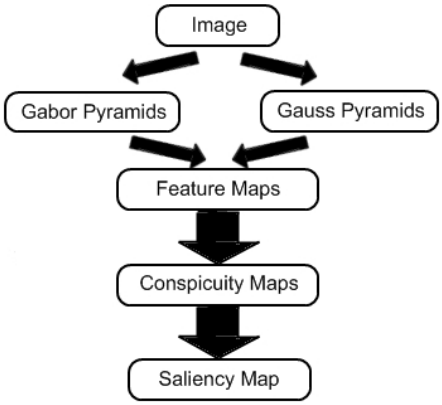


Figure 4: The schema of saliency map building.

We build the saliency map and detect attention zones as described below. The schema of saliency map generation is shown on Fig. 4. Let us define the intensity map *I* as:

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

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

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

For *I, R, G, Y* 8-level Gaussian pyramids are constructed using Gauss separable filter.

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.

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 the individual peaks and also helps not to take into account such things as the very bright background.

Conspicuity maps $\bar{I}$ for intensity, $\bar{C}$ for color, $\bar{O}$ for orientation are summed with specific weights into final image which is called saliency map *S*:

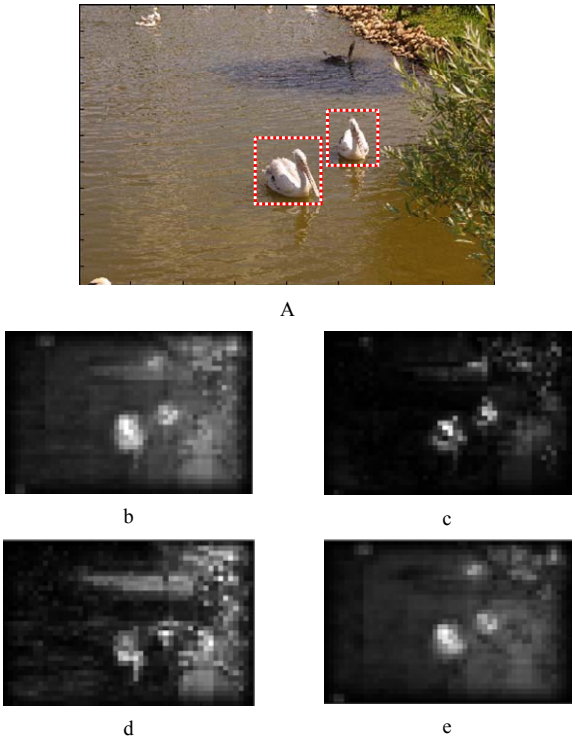
$$S = \frac{wI \cdot N(\bar{I}) + wC \cdot N(\bar{C}) + wO \cdot N(\bar{O})}{wI + wC + wO}.$$

The main problem is to find weights *wI, wC, wO* since due to normalization, different conspicuity maps have different contribution to final result. To solve this problem we applied training by expert procedure. For each photo in the training set the attention zones were marked by several experts. In order to determine the best weights we found maximum of the following function using simplex algorithm:

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

where ROI are marked zones on the image.

Mathematical expectations of weights were calculated after finding values for all images in the training set. The most salient regions correspond to attention zone. Fig. 5 demonstrates two detected zones as well as corresponding conspicuity and saliency maps.



A b c d e

Figure 5: Detected attention zones on initial photo (a) and corresponding saliency map (b), intensity map (c), color map (d), orientation map (e).

For time optimization we work with down-sampled versions of original photos during all steps of the algorithm. Generation of saliency map works less than 0.2 second on PC with quad-core CPU 2.4 GHz for color image with size 500x500 pixels. Processing time can be decreased considerably with the fast approach for saliency map generation via quaternion transform [9] and parallel calculation on GPU [10].

3.4 Attention zones detection for document images

Most of the thumbnail images of documents look the same. It is difficult to distinguish from one another . To recognize the document it is important to see title, headers and embedded pictures. There are a lot of document layout analysis methods which allow to perform the segmentation and detection of different important regions of the document [11], [12]. However existing techniques are computationally expensive.

We propose simple and fast algorithm to detect a block of text from the very large size that relates to title and headers. Our algorithm for detection of such zones of attention includes the following steps. First of all, initial *rgb* image is converted to grayscale *I*. Next step is to downsample the original document image to a size that provides recognizability of text with the size 16-18 pt or more. For example, scanned document image with a resolution 300 dpi should be downsampled 5 times. The resulting image of A4 document has size 700x500 pixels. Handling of grayscale downsampled copy of initial image allows to decrease processing time significantly.

Downsized text regions look like a texture. The areas contain bulk of edges. So, to reveal text regions edge detection technique can be applied. For this purpose we use Laplacian-of-Gaussian (LoG) filtering and zero-crossing. LoG filtering is a convolution of downsampled image *I* with kernel *kg*:

$$k(x,y) = \frac{(x^2 + y^2 - 2*\sigma^2) * k_g(x,y)}{2\pi\sigma^6 \sum_{x=-N/2}^{N/2} \sum_{y=-N/2}^{N/2} k_g(x,y)},$$

$$k_g(x,y) = e^{-(x^2+y^2)/2\sigma^2},$$

where *N* is size of convolution kernel, σ - standard deviation, (x,y) – coordinates of Cartesian system with origin at the center of the kernel.

Zero-crossing approach with fixed threshold *T* is preferable for edge segmentation. The edges are labeled on the binary image *BW* using following rules:

BW(x, y) = 1 if (|Ie(x, y)- Ie(x, y+1)| >= T
and Ie(x, y)<0 and Ie(x, y+1)>0)
or (|Ie(x, y)- Ie(x, y-1)| >= T
and Ie(x, y)<0 and Ie(x, y-1)>0)
or (|Ie(x, y)- Ie(x-1, y)| >= T
and Ie(x, y)<0 and Ie(x-1, y-1)>0)
or (|Ie(x, y)- Ie(x+1, y)| >= T
and Ie(x, y)<0 and Ie(x+1, y-1)>0);
otherwise BW(x, y) = 0,

where *Ie* is outcome of LoG filtering, (*x*,*y*) are coordinates of pixel.

For segmentation of text regions we look for the pixels which have a lot of edge pixels in vicinity according to the equation:

$$L(x,y)=\begin{cases}1, & \sum_{i=x-dx/2}^{x+dx/2}\sum_{j=y-dy/2}^{y+dy/2}BW(i,j)>Tt\\ 0 & \end{cases},$$
$$\forall x,y$$

where *BW* – image with detected edges, *L* – image of segmented text regions, *dx, dy* – sizes of block, *Tt* is a threshold.

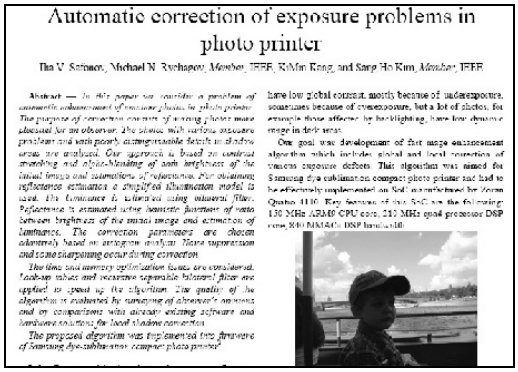
In addition, regions corresponding to vector graphics such as plots and diagrams are segmented too. Next step is Further for *L* labeling of connected regions in *L* and calculation of its bounding boxes. Regions with a small height or width are eliminated.

The calculation of the average size of character for each region of the text and selection multiple zones with a maximal average size of the character are performed on the next step. Let us consider how to calculate the average size of the character of the text region that corresponds some connected region in the image *L*. The text region can be designated as:

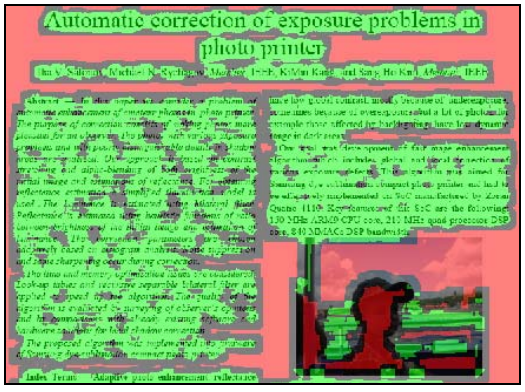
$$Z(x,y)=I(x,y)\times L(x,y),$$
$$\forall x,y\in\Omega$$

where Ω is some connected region in image *L*.

The image *Z* is binarized by threshold. We use well-known Otsu algorithm to calculate threshold *Tz* for histogram calculation the pixels which belongs to Ω are considered only. We apply optimized version of the algorithm as described in [13]. Thus connected regions on the binary image *Zb* are labeled. The size of bounding box is calculated for all connected regions in *Zb*. The average size of character is calculated as the average of the sizes for the Ω zone from *L*. If the number of connected regions in *Zb* is too small then the text region is eliminated.



a)



b)

Figure 6: Detection of text regions: (a) downsized scanned document, (b) detected text regions (in green).

Fig. 6 illustrates our approach for detection of text regions. Detected text regions *L* are marked in green color in fig. 6b. Image *Z* consists from all the connected regions. The average size of the character is calculated for the dark connected areas inside the green region. This is the right way to detect title of paper for image in fig. 6a.

At the final stage of our approach photographic illustrations are identified because they are important for the document recognition as well. The image *I* is divided into non-overlapped blocks with a size *NxM* for the detection of embedded photos. For each block *Ei* energy of the normalized co-occurrence matrix is calculated:

$$C_I(i,j)=C_I(i,j)+1 \text{ if } I(x,y)=i \text{ and } I(x+dx,y+dy)=j,$$
$$\forall x,y$$

$$N_I(i,j)=\frac{C_I(i,j)}{\sum_i\sum_j C_I(i,j)},\quad Ei=\sum_i\sum_j N_I{}^2(i,j),$$

where *x, y* are coordinates of pixels of a block, *dx, dy* are displacements.

If $Ei < 0.01$ then all pixels of the block are marked as related to photo. Further all adjacent marked pixels are combined to connected regions. Regions with the small area are eliminated. Regions with too large area are eliminated too because they also belong to the complex background of a document as a rule. The bounding box of region with the largest area defines zone of the embedded photo. Fig. 7 shows results of detection of the blocks related to photographic illustration.

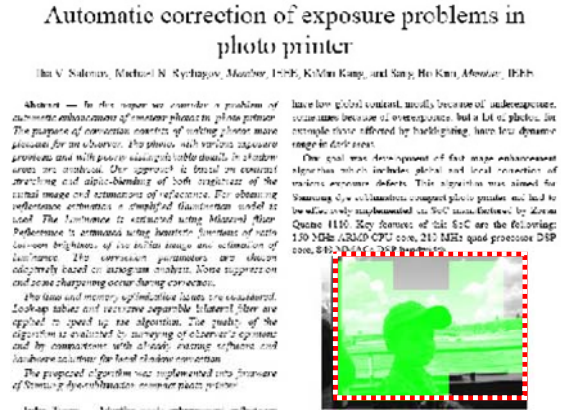


Figure 7: Detection of photographic illustration inside document image.

3.5 Animation creation

First of all the sequence order of zones is selected for animation creation. Always first frame represents a whole downsampled image that is conventional thumbnail. The subsequent zones are selected to provide the shortest path across the image. The animation can be looped. In this case final frame is a whole image too. The animation simulates the following camera effects: tracking in, tracking out and panning between attention zones, slow panning across large attention zone and pauses on the zones. Tracking in, tracking out and panning effects between two zones are created by constructing a sequence from N frames. Each frame of the sequence is prepared with the following steps:

- calculate coordinates of bounding box for cropping zone using line equation in parametric form:

  $x(t) = x1 + t\times(x2 - x1),$

$$y(t) = y1 + t\times(y2 - y1),$$

where (x1, y1) are coordinates of start zone, (x2, y2) are coordinates of end zone, t is increased from 0 to 1 with step dt=1/(N-1);

- crop image using coordinates of calculated bounding box with preserving of aspect ratio;
- resize cropped image to size of destination thumbnail.

Fig. 8 demonstrates example of the animated thumbnail for the image of scanned document. The title and embedded photo are detected on the first stage. The fragment of image with a title is appropriate for quality estimation. Animation consists of 4 transitions between entire image and these two zones as well as viewing a relatively large zone of title. The first sequence of frames look like a camera tracking-in to the left side of title zone.

The second sequence of frames look like a slow camera panning across zone of title; after that the frame is frozen on a moment for visual quality estimation. The third sequence of frames look like camera panning from the right side of title zone to embedded photo; after that frame is frozen on a moment. The final sequence of frames look like a camera tracking-out to entire page; after that frame with entire page is frozen on a moment. The sequence of the frames allows to identify the image content confidently.

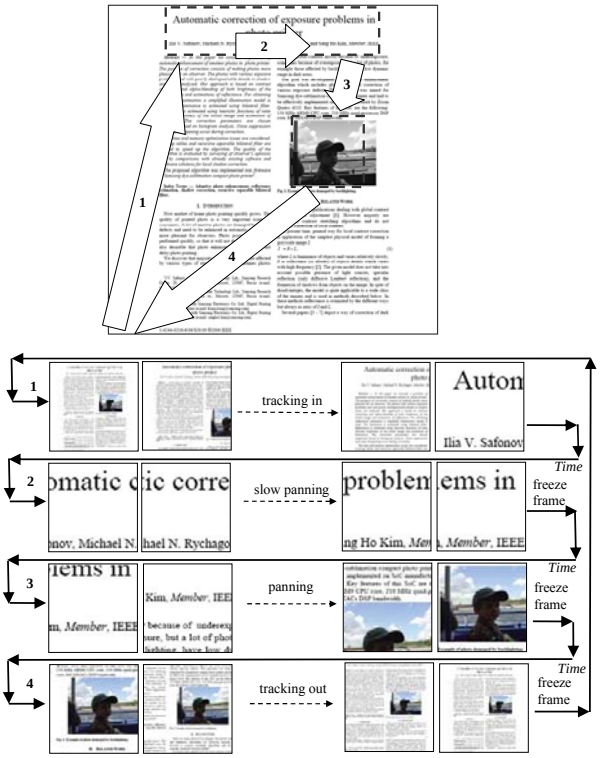


Figure 8: Frames of animated thumbnail for document image.

4. RESULTS AND DISCUSSION

We conducted user study to estimate effectiveness of animated thumbnail in comparison with conventional thumbnail. The study was focused on recognition of image content and its quality. The survey was held among ten persons. Surely ten people is not enough for deep and confident investigation. However it is enough to demonstrate advantages of animated thumbnail.

Survey participants were asked to complete three tasks on one laptop PC with WinXP independently of one another but in the similar conditions. Conventional thumbnails were viewed in the standard Windows Explore application. Animated thumbnails were created as animated GIF files. Each participant has one minute to solve the tasks.

First task was selection of two photos with the particular person known to the respondent. Total number of viewed thumbnail was eight. The photos used for thumbnail generation were never seen before by respondents. Fig. 9 shows conventional thumbnails. The majority of faces are too small for confident recognition. Nevertheless percentage of right answers was not so bad; 62% respondents selected photos with the person. It is probably explained by very high recognition and cognitive abilities of the people. Such characteristics as hair color, head form, build of body, height, typical pose and expression allow to identify known person even if size of photo is extremely small. However the recognition results for animation thumbnails are much better; 95% of respondents selected right photos. Fig. 10 shows animated thumbnails frames which contain enlarged faces. In most cases the enlarged face allows to identify person. Even if a face is not detected as attention zone, walk through zoomed in image fragments allows to see the face in detail often. Perhaps 5% of errors are explained by carelessness because faces were frozen on a moment only and time for task completion was limited.

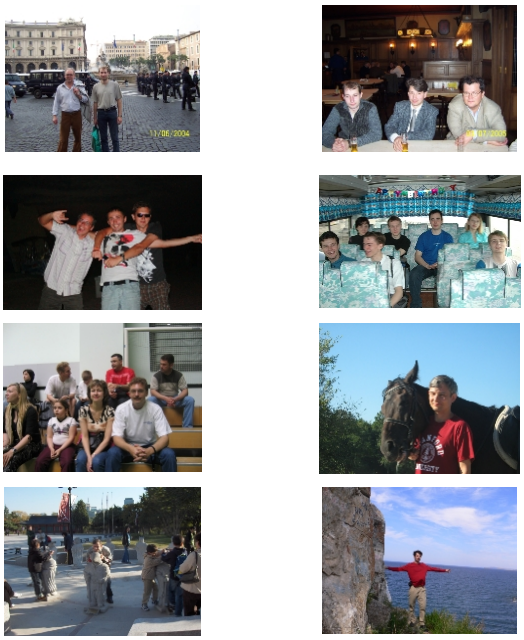


Figure 9: Conventional thumbnails in the survey for selection of photos with the certain person.

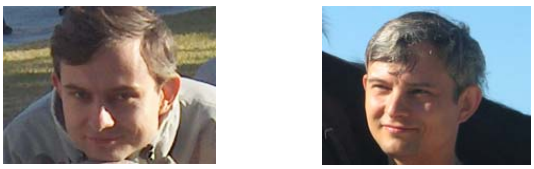


Figure 10: Frames of animated thumbnails in the survey for selection of photos with the certain person.

Second task was selection of two blurred photos from eight. Fig. 11 shows conventional thumbnails of the survey. It is almost impossible to detect blurred photos by means of thumbnail viewing. Only 35% gave the right answers. It is a little bit better than random guessing. Two responders had better results than another because they had a big experience in photography and understood shooting conditions which can cause a blurred photo. Animated thumbnails demonstrate zoomed in fragments of photo. It is allow to select low-quality photos. In our survey 89% of respondents detected right photos by viewing of animated thumbnail frames.

Fig. 12 shows enlarged fragments of sharp photos in top row and blurred photos in bottom row; difference is obvious and blurriness is detectable. 11% of errors are explained by subjective interpretation of blurriness concept probably. Indeed sharpness and blurriness are not formalized strictly and depends from viewing conditions.

Third task was selection two scanned images which represent the document related to “Descreening images” topic. Total number of documents was nine. The conventional thumbnails of the scanned images are shown on fig. 13. Percentage of right answers for conventional thumbnail is 22. It corresponds to random guessing. In general it is impossible to solve the task properly using conventional thumbnails of small size. Contrary animated thumbnails provide high enough level of right answers. 85% of respondents selected pages related to “Descreening” topic due to zooming and panning through title of the papers as it is shown on the fig. 14.

Summary percentage of right answers for the 3 tasks is shown on fig. 15. Animated thumbnail provides capability to recognize image content and to estimate quality confidently and outperforms conventional thumbnail considerably.

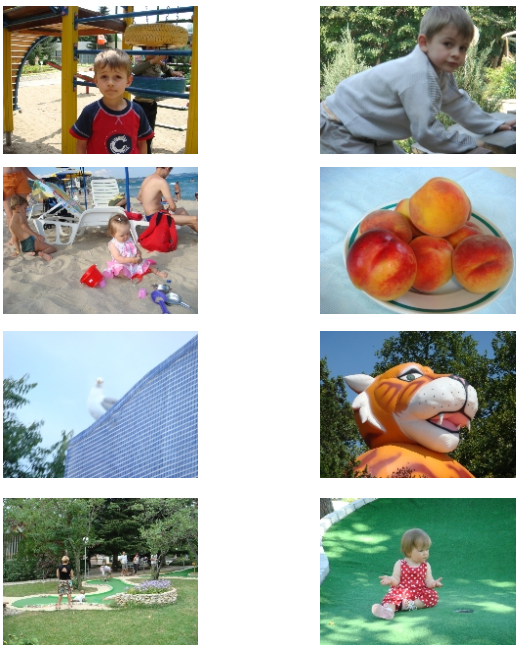


Figure 11: Conventional thumbnails in the survey for selection of blurred photos.

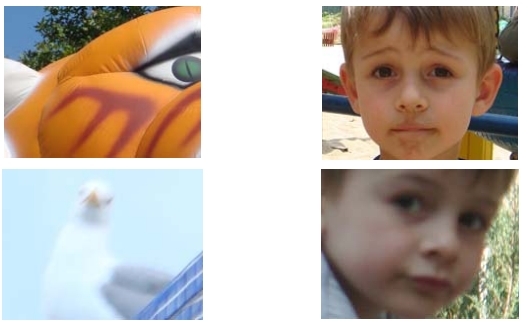


Figure 12: Frames of animated thumbnails in the survey for selection of blurred photos: top row for sharp photos, bottom row for blurred photos.

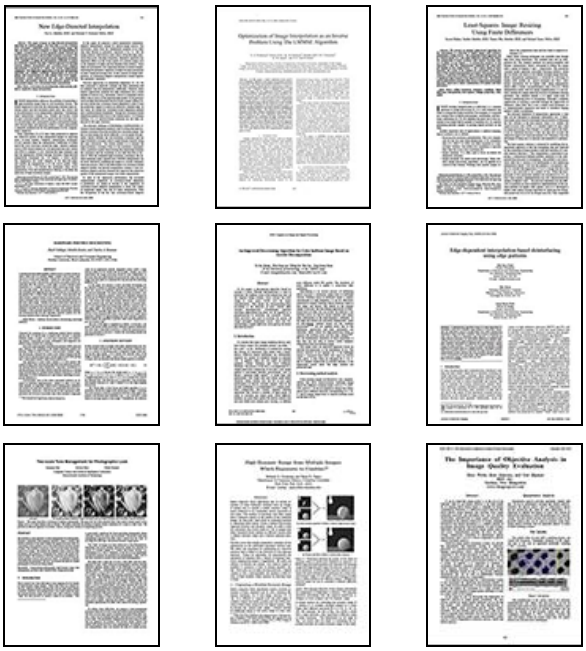


Figure 13: Conventional thumbnails in the survey for selection of documents related to particular topic.

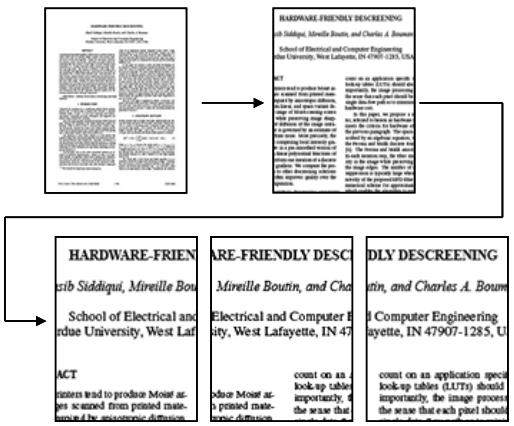


Figure 14: Frames of animated thumbnail with panning through title of the paper.

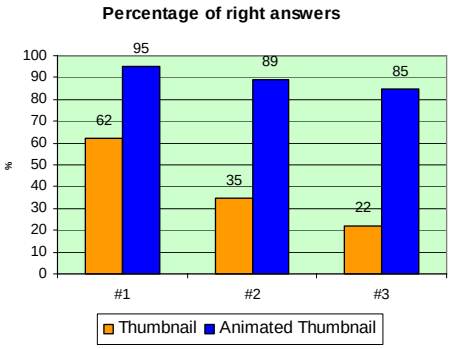


Figure 15: Summary results of the user study.

5. CONCLUSION

Animation thumbnail is applicable in devices with small display for user-friendly viewing of still image thumbnails. First of all it is valuable for mobile phones, Digital Still Cameras, Digital Photo Albums, Digital Photo Frames, Photo Printers and MFP. In addition it is impressive way for navigation through image collections in software and web-applications.

The general idea of animated thumbnail can be extended on thumbnails for separate frames of video and documents such as PDF, MS Word etc.

ACKNOWLEDGMENT

Authors would like to thank Ekaterina Tolstaya for discussions about this paper. Thanks also go to Graphicon's anonymous reviewers for their comments.

REFERENCES

[1] B.Suh, H.Ling, B.B.Bederson, D.W.Jacobs *"Automatic Thumbnail Cropping and its Effectiveness", Proceedings of ACM UIST, 2003.*

[2] R.Samadani, T.Mauer, D.Berfanger, J.Clark, B.Bausk *"Representative Image Thumbnails: Automatic and Manual", Electronic Imaging, 2008.*

[3] K.Berkner, E.L.Schwartz, C.Marle *"SmartNails - Display and Image Dependent Thumbnails", Electronic Imaging, 2004.*

[4] P.Viola, M.Jones, *"Rapid object detection using a boosted cascade of simple features", In Proc. of Conference Computer Vision and Pattern Recognition, 2001.*

[5] B.Ma, W.Zhang, S.Shan, X.Chen, W.Gao, *"Robust head pose estimation using LGBP", Proc. of Conference on Pattern Recognition, 2006.*

[6] M.A.Egorova, A.B.Murynin, I.V.Safonov, *"An improvement of face detection algorithm for colour photos", Proc. of Conference on Pattern Recognition and Image Analysis, 2008.*

[7] 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.*

[8] L.Itti, C.Koch, *"A saliency-based search mechanism for overt and covert shifts of visual attention", Vision Research 40, pp 1489-1506, 2000.*

[9] W.F. Lee, T.H. Huang, Y.H Huang, M.L. Chu and H.H. Chen, *"Efficient Construction of Saliency Map", Proc. SPIE Vol. 7240, 2009.*

[10] P.Longhurst, K.Debattista, A.Chalmers, *"A GPU-based saliency map high-fidelity selective rendering", Proc. of ACM AFRIGRAPH, 2006.*

[11] S. Klink, A. Dengel, T. Kieninger, *"Document structure analysis based on layout and textual features", Proc. 4th IAPR International Workshop on Document Analysis, pp. 99–111, 2000.*

[12] M. Aiello, C. Monz, L. Todoran, M. Worring, *"Document understanding for a broad class of documents", International Journal on Document Analysis and Recognition, vol. 5, no. 1, pp. 1–16, 2002.*

[13] K.C.Lin, *"On improvement of the computation speed of Otsu's image thresholding", Journal of Electronic Imaging, 14(2), 2005.*

About the authors

Ilia V. Safonov received his MS degree in automatic and electronic engineering from Moscow Engineering Physics Institute/University (MEPhI), Russia in 1994 and his PhD degree in computer science from MEPhI in 1997. Since 1998 he is an assistant professor of faculty of Cybernetics of MEPhI (now National Research Nuclear University) while conducting researches in image segmentation, features extraction and pattern recognition problems. Since 2004, Dr. I. Safonov joined Image Enhancement Group, Samsung Research Center, Moscow, Russia where he is engaged in photo, video and document image enhancement projects.

Victor V. Bucha received his MS degree in computer science from Belorussian National Technical University in 2003 and his PhD degree in computer science from United Institute of Informatics Problems, Belarus in 2006. Since 2007, Dr. V. Bucha joined Samsung Research Center, Moscow, Russia where he is engaged in image processing and computer vision projects.