Efficient Super-Resolution Up-Conversion Algorithm for HDTV

Vadim Vashkelis, Natalia Trukhina, Ivan Chirkov
{vashkelis, ntrukhina, chirkov.ivan}@gmail.com

Abstract

Video up-conversion takes significant place in various application areas. One of important application areas is standard-definition television (SDTV) video processing to get high-definition television content (HDTV) for broadcast. However, high-quality up-conversion is a challenging task. Most practical implementations use spatial domain processing such as video frame interpolation for video up-scale. Meanwhile, due to sampling limitation the high-frequency component of output HD video cannot be efficiently reconstructed by applying only the spatial domain processing and high-quality up-conversion usually requires temporal domain processing as well. The authors propose practical implementation of such up-conversion technique providing significantly better visual results in comparison to traditional methods of SDTV to HDTV up-conversion.

Keywords: *Super-Resolution, video up-conversion, HDTV, video enhancement.*

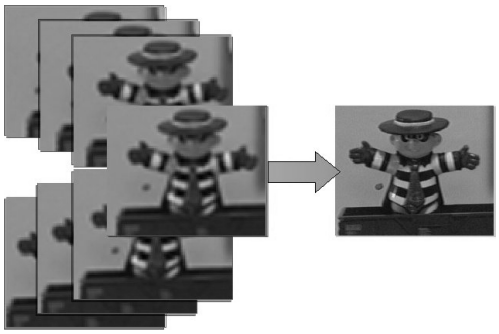

Figure 1: Reconstruction of single super-resolution image based on analysis of several images.

1. INTRODUCTION

There are many up-conversion algorithms widely used. Most of them use spatial signal processing to construct new data points within a set of existing pixels. These methods are based on general interpolation approach, namely on construction of new data points within a set of existing data points with fixed sampling rate. As an example of such interpolation algorithms the nearest neighbour, bilinear, bicubic, spline, sinc, lanczos and some others can be mentioned. These methods use various mathematical interpretation of the spatial signal to construct necessary points.

Another widely used set of methods is frequency domain processing, usually fast Fourier transform or wavelet analysis based algorithms. These methods are based on the shifting property of the Fourier transform, the aliasing relationship between the continuous Fourier transform and the discrete Fourier transform. Accurate processing of the Fourier transform results can give us the frequency domain coefficients of the original scene, which may then be recovered by inverse Fourier transform. [1] However, frequency domain processing has several important disadvantages. These methods require the existence of a transformation which is the Fourier domain equivalent of the spatial domain motion model what is not always feasible. Also, it is difficult to include spatially varying degradation models in the frequency domain reconstruction formulation. [2]

For the last years the variety of works were dedicated to relatively novice approach to video up-conversion addressing to resolve the bandwidth limitation of other methods. This set of algorithms is called super-resolution (SR). This work represents the computational efficient high-quality implementation of super-resolution technology for video up-conversion aimed on television and broadcast applications.

2. TRADITIONAL APPROACH

SR is technique that enhances the resolution of an imaging system. This technique uses additional image information for high resolution. It may be information from single or multiple input images. Single image SR method extracts high-resolution image details from a single low-resolution image, which cannot be achieved by simple sharpening [3] and traditional interpolation. It uses other parts of the low resolution images to guess how the high resolution image should look like.

Multiple-frame SR is a method based on idea of using information from several images to create one up-scaled image. In particular, the source video sequence contains similar, but not identical information. The additional information available in these frames makes possible the reconstruction of visually superior frames at higher resolution than that of the original data. This method tries to extract details from one frames to reconstruct other frames. The SR algorithms are possible only if aliases exist, and the images have sub-pixel shifts. [3] This approach differs a lot from some sophisticated image up-scaled methods which try to synthesize artificial details.

Generally, there are three critical factors affecting super-resolution restoration. Firstly, reliable sub-pixel motion information is essential. Poor motion estimates are more detrimental to restoration than a lack of motion information. Secondly, observation models must accurately describe the imaging system and its degradations. Thirdly, restoration methods must provide the maximum potential for inclusion of a priori information. [4]

3. GENERAL SUPER-RESOLUTION APPROACH

Multiple-frame SR for video sequences uses information from the sub-pixel shifts between several frames of the same scene within a video. This pixel shift is caused by a relative motion between the scene and camera. The video with improved resolution can be created by merging the data from a set of low-resolution frames taking the relative pixel shifts into the account. SR works when several low resolution images $LR(x, y)$ contain slightly different views of the same objects. In this case total information about the object is much higher than information in one frame. Using existing information from current frame $LR_i(x, y)$ and getting additional sub-information from several previous ..,LR_{i-n},..,LR_{i-2},LR_{i-1} and several next LR_{i+1},LR_{i+2},..,LR_{i+n},.. frames we can reconstruct high resolution image $HR_i(x, y)$. Simplified method of reconstruction may be defined as function F from several frames(1).

$$HR_i(x,y) = F\,(\ LR_{i-n}\,(x_{i-n}, y_{i-n})\,, \\ \ldots, \\ LR_{i+n}\,(x_{i+n}, y_{i+n})\ ) \tag{1}$$

The methods discussed in other papers usually describe the reconstruction of HR in ideal conditions. Under these conditions all compensated objects precisely assist in concerned frames LR_j and every sub-pixel shifts are found and position in other frames (x_j, y_j) are exactly known. Actually it is important to know if we afford to use sub-pixels from neighbour frames or we have to amount current low-resolution frame information only. It involves irregular structure of reconstruction on non-uniformly spaced sampling grid and smart image recognition. The registration of low-resolution image sequence results in a composite image of samples on a non-uniformly spaced sampling grid. These sample points are interpolated and positioned over the high-resolution sampling grid. However, despite the simplicity of such model it does not take into consideration the fact that samples of the low resolution images cannot be results of ideal sampling and relative pixel shifts cannot be known a priori. This results in the fact that the reconstructed image does not contain the full range of frequency content that ideally could be reconstructed. Practical implementation of SR shows that high resolution and quality are unachievable without strong model of data points recognition and detection. Also high quality interpolation base is necessary. The expanded formalization of SR implementation can be described as

$$HR_i(x_{hr}, y_{hr}) = F\,(\ LR_{i-n}\left(x_{lr_{i-n}}, y_{lr_{i-n}}\right), \\ \ldots, \\ LR_{i+n}\left(x_{lr_{i+n}}, y_{lr_{i+n}}\right), \\ INT\,(LR_i, x_{hr}, y_{hr})\ ) \tag{2}$$

Due to all these limitations the most important key factors for efficient super-resolution processing are quality of generalized up-scaling, accurate motion estimation and robustness of the super-resolution construction procedure that creates uniform set of data points from non-uniform points mesh. Below, in this work, we will consider all of these three factors separately.

3.1 Non-uniformly sampled grid interpolation

The first step of SR is high quality up-scaling. It is fundamental point for both the motion estimation and super-resolution construction procedure. The best decision is to use multivariate harmonic interpolation with non-uniform mesh nodes. In signal processing, a sinc filter is an idealized filter that removes all frequency components above a given bandwidth, leaves the low frequencies alone, and has linear phase. The filter's impulse response is a sinc function in the time domain, and its frequency response is a rectangular function. [3] In digital signal processing and information theory, the normalized sinc function is commonly defined by

$$sinc(x) = \frac{\sin \pi x}{\pi x} \tag{3}$$

The Lanczos filter is a windowed form of the sinc filter. Its impulse response is the normalized $sinc$ function $sinc(x)$ windowed by the Lanczos window. The Lanczos window is itself the central lobe of a scaled $sinc$, namely $sinc(x/a)$ for a from $-a$ to a (the central lobe scaled to run from $-a$ to a). The resulting function is then used as a convolution kernel to resample the input field. [5] Its formula is given by:

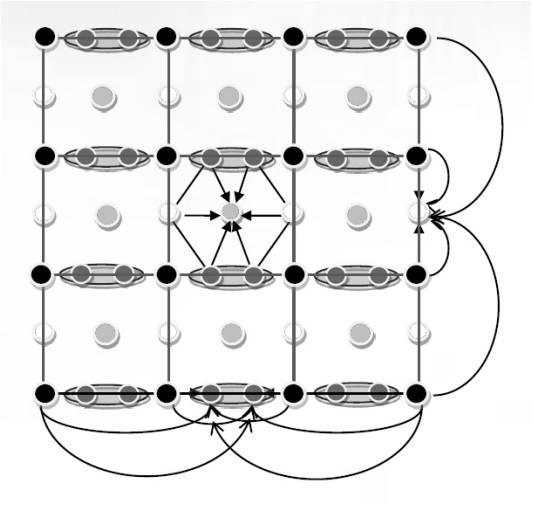

Figure 2: Smart Lanczos interpolation with non-uniform mesh nodes.

$$L(x) = \begin{cases} sinc(x) \cdot sinc(\frac{x}{a}), & -a < x < a, x \neq 0 \\ 1, & x = 0 \\ 0, & otherwise \end{cases} \tag{4}$$

The Lanczos filter has been compared with other filters, particularly other windowing of the sinc filter. Lanczos is the best compromise in terms of reduction of aliasing, sharpness, and minimal ringing. Nevertheless, the regular structure and linear nodes do not give the best results.

Proposed approach uses different weight coefficients for different mesh nodes to improve visual quality of interpolated images. Analysis of the morphological structure of image and individual choice of weight coefficients allow to select correct nodes for high quality interpolation and make precise motion estimation. One way of possible adapted smart Lanczos interpolation with non-uniform mesh nodes is described in Figure 2. It is an example of 4x up-scaling where:

- ☐ black nodes are pixels of based input image,
- ☐ white nodes are pixels of regular vertical Lanczos interpolation,
- ☐ dark gray nodes are pixels of possible horizontal interpolation,
- ☐ light gray nodes are results of total non-uniform interpolation.

Schema of non-uniform mesh nodes can be different and adapts for input image structure and combines different Lanczos windows for different types of images.

3.2 Heterogeneous motion estimation

The large regions overlap that usually exists between successive frames of the same sequence and the multiple sampling of this regions in several frames, yield the conclusion that it is possible to combine this information to achieve higher spatial resolution images. Motion estimation techniques are used to find this overlapping areas from frame to frame. [6] The resulting motion vectors

must be at least sub-pixel precision to provide useful information for SR. For the best quality results quarter-pixel precision is used.

Most papers dedicated to super-resolution construction claim that image registration is known a priori. Meanwhile, image registration or in other words accurate inter frame motion estimation is a crucial component of super-resolution processing. Insufficient accuracy of image registration inevitably leads to significant quality degradation and makes super-resolution approach nearly useless.

Most popular yet powerful enough practical motion estimation approach utilizes the sum of absolute differences (SAD) as a criterion for image templates matching (5).

$$SAD = \sum_{j=0}^{j=n-1} \sum_{i=0}^{i=m-1} |I_1(i,j) - I_2(i,j)| \tag{5}$$

This technique is used in many video coding applications [7] and characterized by high computational simplicity. However, there are number of know lacks of SAD approach what makes it less applicable for super-resolution image registration then for video compression [8]. A Most noticeable problem of SAD-based matching is inconsistency in the case of sufficient noise additions and average brightness (DC component) change.

Many papers dedicated to the problems of image registration and template matching point on morphological hit-or-miss criteria for image matching [9, 10, 11, 12]. The proposed approach combines the power of both methods for creating computational efficient and effective image registration approach. The proposed method of motion estimation combines computational simplicity of SAD based methods and efficiency of morphological analysis MSC(Morphological search criteria)(6) It as defined as morphological SAD - $MSCSAD$(7) where $kSAD$ and $kMSC$ are weighting factors.

$$MSC = \sum_{j} (MAX_i \ |I_1(i,j) - I_2(i,j)| - MIN_i \ |I_1(i,j) - I_2(i,j)|) \tag{6}$$

$$MSCSAD = (SAD \ MSC) \times \begin{pmatrix} kSAD \\ kMSC \end{pmatrix} \tag{7}$$

Position that turns out to be the most similar to the current image pixel in the reference frame, is given by the candidate with the minimum $MSCSAD$ value:

$$MSCSAD(x_{lr}, y_{lr}) = MIN_{\forall x, \forall y} (MSCSAD(x,y)) \tag{8}$$

For precise search and future accurate reconstruction it is important to select the best candidate from operating positions and to understand if this candidate affords to give additional resolution indeed or such position does not exist at all. It is proposed to use complicated pyramid structure of motion estimation with several steps for parcelling out input images and separating background and foreground with objects combining.

As a result of motion estimation and object detection we have half/quarter pixel motion vectors, values of $MSCSAD$ and map of objects and theirs motion. This parameters give us information for SR reconstruction and possibility to construct strong criteria for its employment.

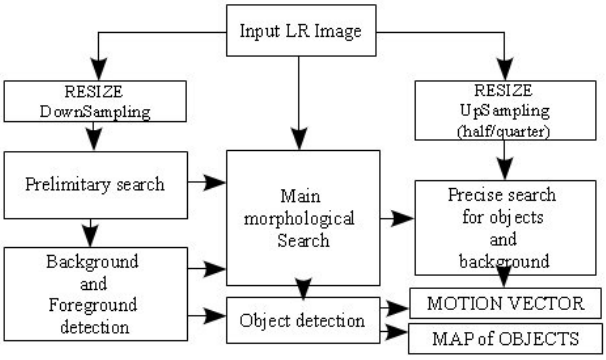

Figure 3: Precise motion estimation.

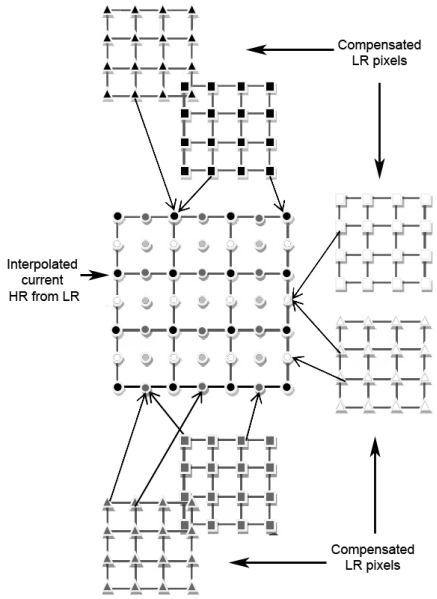

Figure 4: Precise motion estimation.

3.3 Super-Resolution frame construction

Registering a set of low-resolution images using motion compensation results in a single, dense composite image of non-uniformly spaced samples. The super-resolution image can be constructed from this composite using techniques for reconstruction from non-uniformly spaced samples. Restoration techniques are sometimes applied to compensate for degradations [11]. Description of iterative reconstruction techniques can also be noticed [13]. Such interpolation methods are unfortunately overly simplistic. Since the observed data result from severely down-sampled, spatially averaged areas, the reconstruction step (which typically assumes impulse sampling) is incapable of reconstructing significantly more frequency content that is present in a single LR frame. Degradation models are limited, and no a priori constraints are used. There is also question of the optimality of separate merging and restoration steps.

Hence, the construction of super-resolution frame can be formulated as an approximation of non-uniform mesh by the uniformly positioned set of samples. Proposed approach uses different weight coefficients for different mesh nodes to improve visual quality of

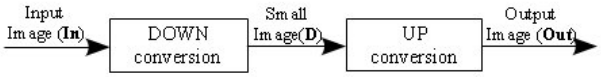

Figure 5: Schema of down-up-conversion for objective quality measurement.

interpolated images. Analysis of morphological structure of image and individual choice of weight coefficients allows to select correct nodes for high quality SR reconstruction.

4. QUALITY MEASUREMENTS

We will relay on two quality measurement criteria subjective and objective. As for objective quality metric a peak signal to noise ratio ($PSNR$) is usually used in practice. It is an engineering term for the ratio between the maximum possible power of a signal and the power of corrupting noise that affects the fidelity of its representation. This criterion is usually defined via mean squared error (MSE)

$$MSE = \frac{1}{W \cdot H} \sum_{j=0}^{j=H-1} \sum_{i=0}^{i=W-1} [I_1(i,j) - I_2(i,j)]^2 \quad (9)$$

The PSNR can be defined as

$$PSNR = 20 \cdot \log_{10}(\frac{MAX_I}{\sqrt{MSE}}) \quad (10)$$

Meanwhile, in case of up-conversion we do not have the reference video with the appropriate frame size because such video can be only obtained by some other up-conversion method from the input and any alternative up-conversion approach will add its own conversion error. To avoid this problem we can use the results of some typical down-conversion as input data (see Figure 5).

In this case the output results of up-conversion routine can be objectively measured against input signal such as $PSNR(In, Out)$.

Subjective evaluation is another important measurement approach because objective metrics such as $PSNR$ cannot fully substitute manual visual perception. During the subjective testing such visual characteristics as video stability, aliasing effect and overall impression were manually evaluated.

For quality testing, we use in this paper two typical image sequences. The first one, *Shields*, is a sequence with moving background containing many small details and texts and local motion on foreground. The second one, *Mobcal*, is a sequence containing global motion on background and fast motion on foreground. Both HR test image sequences are first down-scaled to LR by a factor of 2 in both vertical and horizontal directions. After LR sequence is reconstructed by different interpolation methods and concerned SR algorithm and compared with subjective and objective quality measurement criteria (see Figure 6).

Figures 6 and 7 show input LR frame and results of its bilinear, bicubic and Lanczos (window size a is 3) interpolations and SR transformation. Figure 7 also contains diagram of peak values of $PSNR$ in comparison.

Quantitative $PSNR$ comparisons between reconstructions for whole test sequences and distinctive selected regions in the images are in Table 1.According to quality measurements and visual comparison SR allows efficiently reconstruct low-resolution video to high quality and resolution video. As compared with different inter-

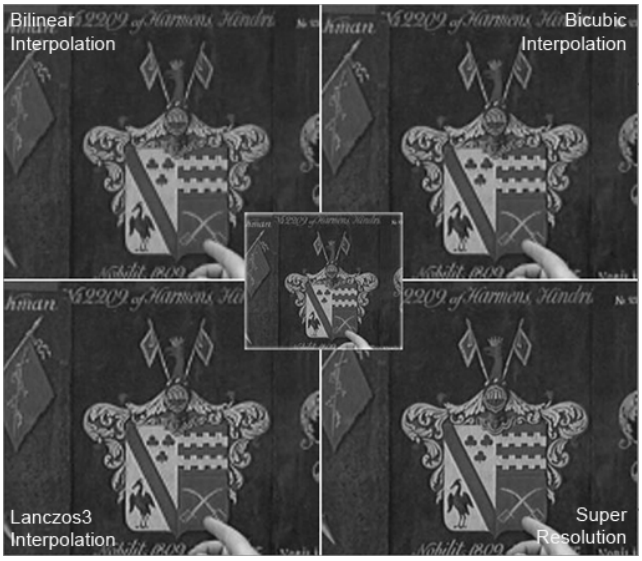

Figure 6: Visual quality comparison.

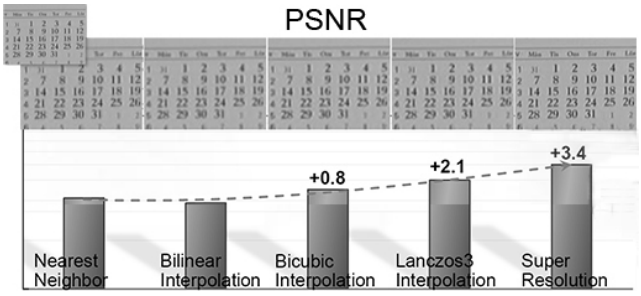

Figure 7: Objective PSNR measurement.

polation methods SR significantly increases resolution within small detailed and slow motion videos. Fast moving videos are reconstructed with the same quality and objective quality measurements as Lanczos interpolation.

5. CONCLUSION

The work demonstrates one possible approach for efficient implementation of high-quality up-conversion solution. The authors demonstrate how to resolve the problem of high computational complexity of every super-resolution solution without degradation of output visual quality. Proposed algorithms allow to build computational efficient solutions on various DSP or GPU platforms. At the same time the efficiency of computation does not affect visual quality of the proposed solution.

The results of objective and subjective quality comparison against well-known spatial domain based alternatives displays that method given exceeds the results of traditional algorithms in both subjective and objective fields.

A reasonable trade-off between quality of up-conversion and relatively low computational complexity of proposed method allows to design the real-time high-quality video up-conversion devices on various DSP or GPU platforms that will address the problem of efficient SD-to-HD video conversion.

	Bicubic Interpolation		Lanczos3 Interpolation		Super Resolution	
	Average PSNR	Max PSNR	Average PSNR	Max PSNR	Average PSNR	Max PSNR
Shields (moving background with small details)	30,69	31,03	31,22	31,97	**32,13**	**32,94**
Shields (moving background without small details)	33,48	33,71	**33,99**	**34,22**	**34,24**	**34,57**
Shields (local motion on foreground)	31,82	32,25	32,98	33,26	**33,21**	**33,78**
Shields	31,89	33,12	32,14	33,69	**33,12**	**35,00**
Mobcal (global motion on background)	32,24	32,96	32,68	33,09	**33,56**	**33,97**
Mobcal (moved text on foreground)	29,22	29,87	30,12	30,65	**31,31**	**31,86**
Mobcal (fast motion on foreground)	31,21	31,94	**31,58**	**32,15**	**31,64**	**32,12**
Mobcal	30,52	31,64	31,19	32,78	**32,01**	**33,14**

Table 1: PSNR comparison on different parts of images.

	Reconstruction of details with slow or without motion	Reconstruction of details with fast motion	Anti-aliasing effect	Clearness
Shields	Better	Same	Better	Better
Mobcal	Better	Same	Better	Better

Table 2: Subjective visual quality SR in comparison with Lanczos3 interpolation.

6. REFERENCES

[1] S. Borman and R. Stevenson, "Super-resolution from image sequences - a review," *Circuits and Systems, 1998. Proceedings. 1998 Midwest Symposium on*, pp. 374-378, 1998.

[2] K. Kim, M. Franz, and B. Scholkopf, "Kernel Hebbian Algorithm for Single-Frame Super-Resolution," *Statistical Learning in Computer Vision (SLCV 2004)*, pp. 135–149, 2004.

[3] S. Borman and R. Stevenson, "Spatial Resolution Enhancement of Low-Resolution Image Sequences - A Comprehensive Review with Directions for Future Research," Department of Electrical Engineering, University of Notre Dame, 1998

[4] Sina Farsiu, M. Dirk Robinson, Student Member, Michael Elad, and Peyman Milanfar, Senior Member, "Fast and Robust Multiframe Super Resolution," *IEEE Transactions on Image Processing*, vol. 13, no. 10, pp. 1327–1344, 2004.

[5] Claude E. Duchon, "Lanczos Filtering in One and Two Dimensions," *Journal of Applied Meteorology*, no. 18(8), pp. 1016–1022, 1979.

[6] Andrew S. Glassner, Ken Turkowski and Steve Gabriel, "Filters for Common Resampling Tasks," *Graphics Gems*, pp. 147–165, 1990.

[7] D. Barreto, L. D. Alvarez and J. Abad, "Motion Estimation Techniques in Super-Resolution Image Reconstruction. A Performance Evaluation," *Virtual observatory. Plate content digitalization, archive mining and image sequence processing*, Sofia, Bulgary, vol. 1, pp. 254–268, 2006.

[8] I. Richardson, "H.264 and MPEG-4 Video Compression: Video Coding for Next-generation Multimedia," Chichester: John Wiley & Sons Ltd., 2003.

[9] Lisa Gottesfeld Brown, "A survey of image registration techniques," *ACM Computing Surveys (CSUR)*, vol. 24, iss. 4, pp. 325–376, 1992.

[10] E. Aptoula, S. Lefevre and C. Ronse, "A hit-or-miss transform for multivariate images," *Pattern Recognition Letters*, vol. 30, iss. 8, pp. 760–764, 2009.

[11] M. Khosravi and R. Schafer, "Template matching based on a grayscale hit-or-miss transform," *IEEE Transactions on Image Processing*, vol. 5, no. 5, pp. 1060–1066, 1996.

[12] B. Perret, S. Lefevre and Ch. Collet, "A robust hit-or-miss transform for template matching applied to very noisy astronomical images," *Pattern Recognition*, vol. 42, no. 11, pp. 2470–2480, 2009.

[13] A. M. Tekalp, M. K. Ozkan, and M. I. Sezan, "High-resolution Image Reconstruction from Lower-resolution Image Sequences and Space-varying Image Restoration," *Proceedings of the IEEE international conference on acoustics, speech, and signal process*, vol. 3, pp. 169–172, 1992.

[14] T. Komatsu, T. Igarashi, K. Aizawa, and T. Saito, "Very high resolution imaging scheme with multiple different aperture cameras," *Signal Processing: Image Communication*, vol. 5, no. 5-6, pp. 511–526, 1993.

[15] Barbara Zitova, Jan Flusser, "Image registration methods: a survey," *Image and Vision Computing*, vol. 21, no. 11, pp. 977–1000, 2003.

ABOUT THE AUTHOR

Vadim Vashkelis is a Ph.D. at St. Petersburg State Polytechnical University. His contact email is `vashkelis@gmail.com`.

Natalia Trukhina is a graduate at St. Petersburg State University of Aerospace Instrumentation. Her contact email is `ntrukhina@gmail.com`.

Ivan Chirkov is a graduate at St. Petersburg State Polytechnical University. His contact email is `chirkov.ivan@gmail.com`.