DOI: 10.2352/ISSN.2470-1173.2016.15.IPAS-187

Video segmentation in presence of static and dynamic textures

Frantc V. A.(a), Makov S. V.(a), Voronin V. V.(a), Marchuk V. I.(a), Stradanchenko S. G.(a), Egiazarian K. O.(b); (a) Don State Technical university, Dept. of Radio-Electronics Systems, Gagarina 1, Rostov on Don, Russian Federation, 344010; (b) Dept. of Signal Processing, Tampere University of Technology, Korkeakoulunkatu 1, Tampere, Finland FI-33720

Abstract

This paper describes an approach to video sequence over–segmentation. The objective is to split the video up to set of disjoint spatio–temporal regions with homogeneous texture properties. In the work we consider three possible types of regions: static texture, dynamic texture and non-textured region. Video over-segmentation is useful for wide range of applications, including perceptual video coding, video-based object recognition and high-level video segmentation. We treat the problem as a labeling problem on a Markov Random Field. Observed data are represented by output of fully-connected layer of convolutional neural network trained on static and dynamic textures. The hidden states of our model represent appropriate region labels. To provide robust over-segmentation we employ energy function composed of terms associated with neighboring voxels similarity and smoothness of obtained supervoxels. We show that our approach is able to segment static and dynamic textures in simultaneous fashion. We have tested our approach on several video sequences rich of static and dynamic textures and it has shown promising results.

Keywords*—over-segmentation, supervoxel, video segmentation, CNN-based video features, multilabel graph-cut*

Introduction

Image and video segmentation has a long history and wide range of applications in the field of computer vision [1]. In this paper we present a novel approach to video over–segmentation targeted to video coding and object detection applications. The objective of our approach is to split the video sequence to several disjoint regions according to their local appearance. We consider as different static texture (ST), dynamic texture (DT) and non-textured regions (NT). The goal is to determine spatio-temporal regions with different static and dynamic textures within the source video sequence. We refer to such regions as supervoxels in contrast to particular element of video sequence which we call voxel. To be included in the same supervoxel neighboring voxels should have similar texture in local neighbourhood. The main goal of our approach is to mark each pixel of video stream with label unique for each textured region. Various high-level computer vision and video processing techniques could benefit from preprocessing in such a way.

There is no universal definition of the term texture. Usually texture is defined as partly regular and partly stochastic pattern. In existing literature on the subject particular definitions depend of particular task. It is widely agreed to be very subjective matter. As well as still image textures the dynamic textures could be important during automatic video analysis [2]. Division of the video sequence to regions with similar texture properties is useful tool for many applications.

It's widely acknowledged that the task of segmentation in general cannot be tackled without particular application in the mind. The problem of unsupervised segmentation is ill-defined because semantically meaningful objects do not usually correspond to homogeneous spatio-temporal regions described by homogeneous color, texture or motion. It makes the task of image and video segmentation rather difficult and give a rise to over-segmentation based techniques. Supervoxels obtained by over–segmentation could be grouped to form semantically meaningful segmentation by more high level approaches.

Despite to significant amount of papers on the subject there are no approach suitable in different situations at the time. In this paper we present framework for video sequence over-segmentation based on deep convolutional neural network (CNN) features and markov random field formalism to model relations among neighboring voxels. The main contribution of this work is joint spatio–temporal CNN-based features, specially suited for simultaneous static and dynamic textures description. The aim is not to determine spatio-temporal boundaries of individual objects but rather to provide elements of more high level that voxels for subsequent processing.

Previous work

During recent years the subject of over-segmentation or even hierarchical subdivision of video data was actively investigated by researches in the field of computer vision and video processing. Early approaches teded to produce supervoxel with partially occluded boundaries. The good overview of existing methods is presented in work [1]. Authors empirically compare five state-of-the-art superpixel algorithms for their ability to adhere to image boundaries, speed, memory efficiency, and their impact on segmentation performance. However, there are many plausible supervoxel methods and little understanding as to when and where each is most appropriate. Moreover the solid part of existing approaches are aimed to produce supervoxels without taking in account texture properties, only more low level ones.

Turns out, the main problem of many existing approaches is their high computational requirements. To address the problem in the work [3] introduces superpixel algorithm, simple linear iterative clustering (SLIC), based on k-means clustering approach to efficiently generate superpixels. Approach is rather simple and well suited for still image over-segmentation but have lack of generality.

In the work [5] Liang et al. proposed the video supervoxel generation algorithm using partially absorbing random walks to get more accurate supervoxels in these regions. Authors aim to use joint use of appearance and motion cues, which effectively exploits the temporal consistency in video sequence. However, approach tends to produce a lot of small regions in the case of

DOI: 10.2352/ISSN.2470-1173.2016.15.IPAS-187

highly textured videos.

The work [4] addresses the limitation caused by large memory requirement even for quite short video sequences when use convinient techniques by proposing an approximation framework for streaming hierarchical video segmentation motivated by data stream algorithms: each video frame is processed only once and does not change the segmentation of previous frames. The Chang et al. [5] have developed a generative probabilistic model for temporally consistent superpixels in video sequences. In contrast to supervoxel methods, object parts in different frames are tracked by the same temporal superpixel. The drawback of such an approach susceptible to undersegmentation at coarse levels and over-segmentation at fine levels, which make it a challenge to adopt the hierarchies for later use. The work by Xu et al.[6] presents the method called the uniform entropy slice, seeks a selection of supervoxels that balances the relative level of information in the selected supervoxels based on some post hoc feature criterion such as object-ness.

The interesting strategy to generate supervoxels is use of probability graph-theory based strategies. It's widely applied in bunch of different computer vision problems. Another popular way to address the problem is coarse-to-fine energy minimization strategy for semantic video segmentation [7]. The strategy is based on a hierarchical abstraction of the supervoxel graph that allows us to minimize an energy defined at the finest level of the hierarchy by minimizing a series of simpler energies defined over coarser graphs. In [9] is introduced and addressed the problem of video object cosegmentation, which concerns the task of segmenting the common object in a pair of video sequences. We present a new algorithm that works on supervoxels in videos to solve this task. The algorithm compute the intra-video relative motion derived from dense optical flow and the inter-video co-features based on Gaussian mixture models. The work [12] address spatio-temporal detection of actions and events in video is a challenging problem. Besides the difficulties related to recognition, a major challenge for detection in video is the size of the search space defined by spatio-temporal tubes formed by sequences of bounding boxes along the frames.

In the task of activity recognition in videos [13] to segment the video into supervoxels, we explore two recent video segmentation algorithms. The proposed representation enables localization of common regions across videos in both space and time. Importantly, since the video segments are meaningful regions, we can interpret the proposed features and obtain a better understanding of why two videos are similar. Evaluation on classification and retrieval tasks on two datasets further shows that Motion Words achieves state-of-the-art performance. Based on the recent developments of visual recognition in static images, many concepts have been successfully extended to video sequences. Similar to object recognition in images, bag-of-features based methods have recently shown excellent results for action recognition. Despite recent developments, the representation of local regions in videos is still an open field of research.

Despite the progress made in recent years there is still place for investigation on the subject of video over-segmentation. However, most of the traditional supervoxel algorithms do not perform well in the regions with complex textures or weak boundaries. These methods may generate the supervoxels with overlapping boundaries.

DCNN–based supervoxels

In this section we describe our approach to video supervoxels generation. The paper [3] – introduce the superpixel partitioning problem in an energy minimization framework, and optimize with graph cuts. Our energy function explicitly encourages regular superpixels. We explore variations of the basic energy, which allow a trade-off between a less regular tessellation but more accurate boundaries or better efficiency. Our advantage over previous work is computational efficiency, principled optimization, and applicability to supervoxel segmentation. Authors achieve high boundary recall on 2D images and spatial coherence on video. We also show that compact superpixels improve accuracy on a simple application of salient object segmentation.

The encoding process can be divided into two stages: analysis and synthesis. On the stage of the analysis of the video sequence is segmented into non-overlapping regions. Each of the areas is classified. The options are not part of the redeveloped, static texture, dynamic texture. The second step is to perform calculations descriptors for slices of the video sequence. Segmentation is performed using the method multilabel graph cut optimization. As a descriptor, a combination of local area descriptor HOF (histogram of optical flow) [10] and the exit convolution network [9]. The architecture of the network used by the convolution presented on figure 1.

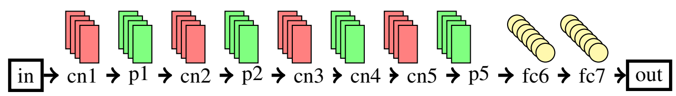


Figure 1: CNN for feature extraction

The paper [10] describe a new system for searching video databases using free-hand sketched queries. But, this method does not employ texture information. We parse space-time volumes from video to form graph representation, which we match to sketches under a Markov Random Field (MRF) optimization. The MRF energy function is used to rank videos for relevance and contains unary, pairwise and higher-order potentials that reflect the colour, shape, motion and type of sketched objects.

The algorithm

A grayscale video sequence is typically modelled as a function $u : \Omega \leftarrow \mathbb{R}$ where $\Omega \subseteq \mathbb{R}^2$ is usually a rectangle and $u(x)$ is the intensity of the grey level at the point x.

The most important aspect of the problem is features employed to describe important aspects of video sequence. It's rather difficult to make a decision's based on raw data only. These methods open the possibility to use strong but computationally expensive features since only a relatively small number of detection hypotheses need to be assessed. In this paper we make two contributions towards exploiting detection proposals for spatio–temporal detection problems. First, we extend a recent 2D object proposal method, to produce spatio–temporal proposals by a randomized supervoxel merging process. We introduce spatial, temporal, and spatio-temporal pairwise supervoxel features that are used to guide the merging process. Second, we propose a new efficient supervoxel method. We experimentally evaluate our detection proposals, in combination with our new supervoxel method as well as existing ones. This evaluation shows that our supervoxels lead to more accurate proposals when compared to using existing state-of-the-art supervoxel methods. The overall workflow of our

DOI: 10.2352/ISSN.2470-1173.2016.15.IPAS-187

approach is presented on figure 2.

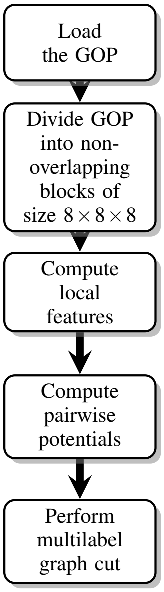


Figure 2: Proposed approach

Our approach are process one group of pictures at the time. Schematic representation of the video sequence is presented on the figure 3.

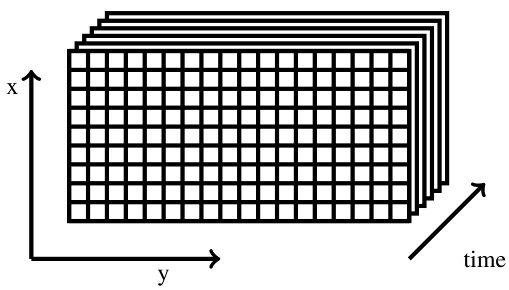


Figure 3: The group of pictures (GOP)

We parse space-time volumes from video to form graph representation, which we match to sketches under a Markov Random Field (MRF) optimization. The MRF energy function is used to rank videos for relevance and contains unary, pairwise and higher-order potentials that reflect the colour, shape, motion and type of sketched objects.

Assumptions and data representation

Another result of the algorithm is to classify types of blocks. The descriptors used as a feature vector at the output of the second layer is fully connected network convolution (fc7) in figure 1.

The network is composed of:

- Input unit of a video sequence
- Convolutional layer
- Filter bank, trained on the data

The novelty of our approach is to use hierarchical representations as video descriptors and global model based on markov random field to simultenously segment and label blocks.

Training of convolution network is made on the basis of texture DTD. The basis of the approach is based on studying features.

The evaluation of the effectiveness of spending as follows. From the binary stream removes all data related to the ejected blocks. So as regression analysis involves finding mathematical expectation of a random variable depending on the available variables, it is composed of additive random variable.

Energy function (Objective function)

To formulate the objective function we use Markov Random Field formalism. The equation for pairwise potentials is:

$$E_{int}(x,y) = s_x(x,y)\cdot(I(x,y)-I(x+1,y)) + s_y(x,y)\cdot\rho_{int}(I(x,y)-I(x,y+1)) \quad (1)$$

where ρ is a monotonically increasing function of the fine $\rho(d) = |d|$. The figure 4 illustrates an enlarged portion of the Markov random field. Where $X_{x,y}$ are hidden variables. Each latent variable is connected to four neighbors ribs with weights w (indicated by squares). The yellow color shows the observed data - descriptor block. The external energy $E_{ext}(x,y)$ for each pixel depends on the difference between the state of latent variable and the observed value.

$$E_{ext}(x,y) = w(x,y)\cdot\rho_{ext}(X(x,y)-I(x,y)) \quad (2)$$

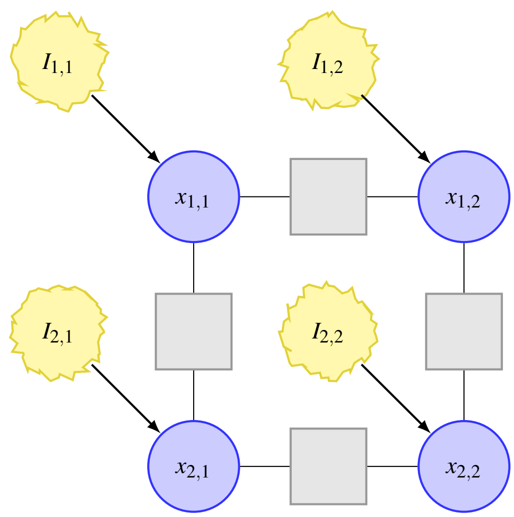


Figure 4: Video model

Consider a multi-label segmentation task, where we want to partition an image into object and background, i.e, we have to decide for each pixel p whether it can be classified as "object" or "background". Informally speaking, a "good" segmentation is obtained when all pixels of the same class have similar intensities (or colors) and if the boundaries between pixels labeled as "object" and "background" are located such that adjacent pixels of different labels have dissimilar intensities/colors. The mapping of these criteria to the two terms of the energy of is straightforward as follows:

$$E_{MRF} = \sum_{p} E_d(p) + \sum_{p,q\in N} E_{int}(p,q) \quad (3)$$

DOI: 10.2352/ISSN.2470-1173.2016.15.IPAS-187

Within multi-label graph-cut technique in order to incorporate the data-driven terms E_d, two additional so-called terminal nodes (one "source" node s, and the sink t are introduced. Every non-terminal node can be connected to s as well as t by additional edges e_{ps} and e_{pt}. The main workflow of our approach is composed of several steps. We work with groups of the pictures (GOP) of length 8 at the time. Figure 5 presents block-scheme of our approach. Here E_d measure data fidelity can be derived from an evaluation how well the observed data at pixel p fits into intensity/color statistics obtained from pixels with known labelled, which could, e.g. be brightness histograms of Gaussian mixture models of color distributions. image features representation. The interaction potentials E_{int}, which measure the cost of assigning different labels to adjacent pixels, can be based on the intensity or color difference between neighbours p and q. This kind of energy function can be modeled by an MRF, where each pixel p is represented by node. In accordance to the pixel grid, these nodes are arranged in a rectangular two-dimensional grid. Adjacent nodes p and q are connected by an edge e_{pq}. For example, a node p can be connected to all nodes within its 4-neighborhood. Another common choice is connecting all nodes. The weight w_{pq} of each edge e_{pq} represents the interaction potentials $E_{int}(p,q)$ between adjacent pixels. w_{pq} of each edge e_{pq} represents the interaction potentials $E_{int}(p,q)$ between adjacent pixels. w_{pq} should be hight when the intensities/colors of p and q are similar.

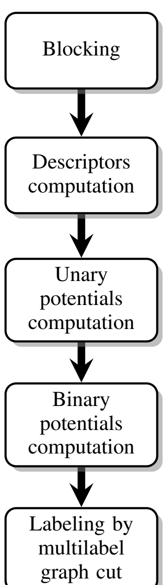


Figure 5: Main workflow of proposed approach

The weights w_{ps} reflect the fidelity to the statistics of the background pixels. Hence, w_{ps} should be high (and w_{pt} low) if the data observed at p fits well into the statistics of all known object pixels. Respectively, w_{ps} should be low (and w_{pt} high) when the appearance at p is in accordance with background appearance. As a summary, the whole setting is described by a graph $G = (N,E)$. If we consider the case of two labels, then segmentation now be represented by a so-called graph cut. A cut C of a graph is a subset of edges. For each cut C, we can define a so-called cut function E_C, which can be set to the sum of the weights of the edges it severs:

$$E_C = \sum_{e \in C} w_e \tag{4}$$

Consequently, the optimal segmentation can be found through the minimum cut, e.g., the subset of edges which minimized:

$$C^* = \arg\min \sum_{e \in C} w_e \tag{5}$$

MRF is an undirected graph where each node represents a pixel in an image $\mathbf{I}$, and each edge represents relation between pixels. Each node is associated with a binary latent variable, $y_u^i \in 0,1$, indicating whether a pixel i has label u. We have $\forall u \in L = 1,2,\ldots,l$, representing a set of l labels. The energy function of MRF is written as

$$E(\mathbf{y}) = \sum_{\forall i \in \mathscr{V}} \Phi(y_i^u) + \sum_{\forall i,j \in \mathscr{E}} \Psi(y_i^u, y_j^v), \tag{6}$$

where $\mathbf{y}$, $\mathscr{V}$, and $\mathscr{E}$ denote a set of latent variables, nodes, and edges, respectively. $\Phi(y_i^u)$ is the unary term, measuring the cost of assigning label u to the i-th pixel. For instance, if pixel i belongs to the first category other than the second one, we should have $\Phi(y_i^1 < \Phi(y_i^2)$. Moreover, $\Phi(y_i^u, y_j^v)$ is the pairwise term that measure the penalty of assigning labels u, v to pixels i, j respectively. The labeling problem is to assign a label from the label set L to each of the sites S. The goal of GOP segmentation is to assign a level f_i from the set $L = edge, nonedge$ to site $i \in S$, where elements in S in the image pixel. The

$$f = \{f_1, \ldots, f_m\} \tag{7}$$

is called labeling of the sites in S in terms of the labels in L. When each site is assigned a unique label, $f_i = f(i)$ can be regarded as a function with domain S and image L. Because the support of the function is the whole domain S, it is a mapping from set of sites to set of labels:

$$f : S \to L \tag{8}$$

All the sites have the same label set L. In the MAP-MRF labeling, $P(f|d)$ is the posterior distribution of an MRF. An important step in Bayes labeling of MRF's is to derive this distribution. Here we use a simple example to illustrate the formulation of a MAP-MRF labeling problem. The problem is to segment video to several disjoint regions $\Omega = \{\Omega_1, \Omega_2\}$. Assuming that the image surfaces are flat, the the joint prior distribution of f is:

$$P(f) = \frac{1}{Z} e^{-U(f)} \tag{9}$$

Here $U(f) = \sum_i$ is the prior energy for the type of surface. The MAP estimate is equivalently found by minimization the posterior energy function $f^* = \arg\min_f U(f|d)$. There is only one parameter is this simple example, σ_i. When it is determined, $U(f|d)$ is fully specified and the MAP-MRF solution is completely defined. The prior model depends on the type of the scene we expect. In image analysis, it is often one of the Gibbs models. The likelihood model depends on physical considerations such as the sensor process (transformations, noise, etc.). It is often assumed to be Gaussian. The parameters in both models need to be specified for the definitions of the models to be complete. The sites in S are related to one another via a neighborhood system. A neighborhood system of S is defined as N_i is the set of sites neighboring i. For our goal we use simple pair-wise cliques.

Let $F = \{F_1, \ldots, F_m\}$ be a family of random variables defined on the set S in which each random variable F_i takes a value f_i in L. The family F is called a random field. We use the notation $F_i = f_i$. A set of random variables F is said to be a *Gibbs random field* on *w.r.t*. N if and only if its configurations obey a Gibbs distribution. A Gibbs distribution takes the form

$$P(f) = Z^{-1} \times e^{-\frac{1}{T}U(f)} \tag{10}$$

where $Z = \sum_{f \in \mathbb{F}}$. Due to Markov-Gibbs Equivalence stated by Hammersley-Clifford markov random field could be represented using equivalent GRF. . When learning from video, $\mathbf{x}$ and $\mathbf{y}$ are image patches (expressed as ectors) at identical spatial locations in sequential frames, and $\mathbf{z}$ is a latent representation of the transformation between $\mathbf{x}$ and $\mathbf{y}$. The energy of any joint configuration $\{\mathbf{y}, \mathbf{z}; \mathbf{x}\}$ is converted to a conditional probability by normalizing it with the partition function, $Z(x) = \sum_{y,z} \exp(-E(y, z; x))$ which is intractable to compute exactly since it involves a sum over all possible configurations of the output and latent variables. However, we do not need to compute this quantity to perform either inference or learning. Given an input-output pair of image patches, $\{x, y\}$, it follows from quations that

$$p(z_k = 1|x, y) = \sigma(\sum_{ij} W_{ijk} x_i y_j + b_k) \tag{11}$$

where $\sigma z = \frac{1}{1+\exp(-z)}$ is the logistic function. Maximinizing the marginal conditional likelihood, $p(y|x)$, over parameters $\theta = \{W, b, c\}$ is difficult for all but the smallest models due to the intractability of computing Z. Learning, however, still works well if we approximately follow the gradient of another function called the contrastive divergence (CD).

Experimental results

The main goal of our approach is to segment video sequence (or group of pictures in particular) to disjoint set of voxels – 3D blobs of arbitrary shape. Video should be segmented in such a way that each voxel will be ”uniformely textured” in some way. We consider two kinds of texture – static texture and dynamic texture. Segmentation is carried out within macro blocks of size 8×8 or 16×16. Experimental results for 3 subsequent frames of video sequence ”controlled burn” are presented on figure 6. Supervoxels are separated by yellow edges. Regions with static textures are marked by green color. Dynamic textures are market by blue color. As you can see, proposed method give results consistent with human perception.

Conclusion

The paper presents a video segmentation algorithm. We treated this problem as a labeling problem on a Markov Random Field using energy function composed of terms associated with neighboring voxels similarity and smoothness of obtained supervoxels. Observed data are represented by output of fully-connected layer of convolutional neural network trained on textures. Proposed approach shown promising results on several video sequences with static and dynamic textures.

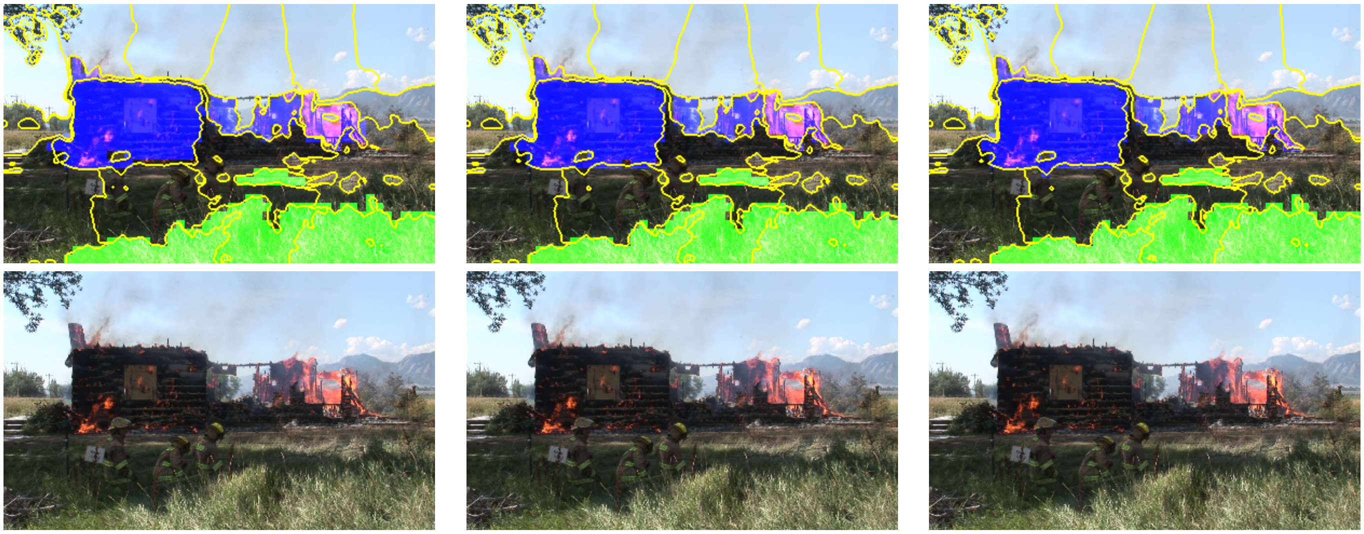
ACKNOWLEDGEMENT

The reported study was supported by the Russian Foundation for Basic research (RFBR), research projects #15-01-09092/15, #15-37-21124/15 and by the Ministry of Education and Science of the Russian Federation, research project #1.10.14.

References

[1] Achanta, Radhakrishna, et al. ”SLIC superpixels compared to state-of-the-art superpixel methods.” Pattern Analysis and Machine Intelligence, IEEE Transactions on 34.11 (2012): 2274-2282.
[2] Xu, Chenliang, and Jason J. Corso. ”Evaluation of super-voxel methods for early video processing.” Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on. IEEE, 2012.
[3] Veksler, Olga, Yuri Boykov, and Paria Mehrani. ”Superpixels and supervoxels in an energy optimization framework.” Computer VisionECCV 2010. Springer Berlin Heidelberg, 2010. 211-224.
[4] Xu, Chenliang, Caiming Xiong, and Jason J. Corso. ”Streaming hierarchical video segmentation.” Computer VisionECCV 2012. Springer Berlin Heidelberg, 2012. 626-639.
[5] Chang, Joana, Donglai Wei, and John W. Fisher. ”A video representation using temporal superpixels.” Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on. IEEE, 2013.
[6] Xu, Chenliang, Spencer Whitt, and Jason J. Corso. ”Flattening supervoxel hierarchies by the uniform entropy slice.” Computer Vision (ICCV), 2013 IEEE International Conference on. IEEE, 2013.
[7] Jain, Abhishek, Saptarshi Chatterjee, and Ren Vidal. ”Coarse-to-fine semantic video segmentation using supervoxel trees.” Computer Vision (ICCV), 2013 IEEE International Conference on. IEEE, 2013.
[8] Galasso, Fabio, et al. ”A unified video segmentation benchmark: Annotation, metrics and analysis.” Computer Vision (ICCV), 2013 IEEE International Conference on. IEEE, 2013.
[9] Chen, Ding-Jie, Hwann-Tzong Chen, and Long-Wen Chang. ”Video object cosegmentation.” Proceedings of the 20th ACM international conference on Multimedia. ACM, 2012.
[10] Hu, Rui, et al. ”Markov random fields for sketch based video retrieval.” Proceedings of the 3rd ACM conference on International conference on multimedia retrieval. ACM, 2013.
[11] Trichet, Remi, and Ramakant Nevatia. ”Video segmentation with spatio-temporal tubes.” Advanced Video and Signal Based Surveillance (AVSS), 2013 10th IEEE International Conference on. IEEE, 2013.
[12] Oneata, Dan, et al. ”Spatio-Temporal Object Detection Proposals.” Computer VisionECCV 2014. Springer International Publishing, 2014. 737-752.
[13] Taralova, Ekaterina H., Fernando De la Torre, and Martial Hebert. ”Motion words for videos.” Computer VisionECCV 2014. Springer International Publishing, 2014. 725-740.
[14] Neubert, Peer, and Peter Protzel. ”Evaluating superpixels in video: Metrics beyond figure-ground segmentation.” Proc. British Machine Vision Conference. 2013.
[15] De Souza, Kleber Jacques, et al. ”Hierarchical video segmentation using an observation scale.” Graphics, Patterns and Images (SIB-

©2016 Society for Imaging Science and Technology
DOI: 10.2352/ISSN.2470-1173.2016.15.IPAS-187



Frame 1 Frame 2 Frame 3

Figure 6: Segmentation results for the video sequence "Controlled burn"

GRAPI), 2013 26th SIBGRAPI-Conference on. IEEE, 2013.

Author Biography

Vladimir Frantc *was born in Dushambe, Tadjikistan in 1989. He recieved his BS in electrical engeniering from South-Russian university of economics and service in 2011. He recieved the M.Sc. degree in technics from Don State Technical university(Russian Federation) in 2013. At the time he is a PHD student at Don State Technical University. His research interests lay in the areas of applied mathematics, digital signal processing and computer vision.*

Voronin Viacheslav *was born in Rostov (Russian Federation) in 1985. He received his BS in radio engineering from the South-Russian State University of Economics and Service (2006), his MS in radio engineering from the South-Russian State University of Economics and Service (2008) and his PhD in technics from Southern Federal University (2009). Voronin V. is member of Program Committee of conference SPIE. His research interests include image processing, inpainting and computer vision..*

Sergey Makov *was born in 1978. In 2001 he has graduated Don State University of Service as radio engineer. Since then he is working in the engineering company as designer of telecommunication hardware. He has got PhD in 2011. Since then he has worked as a professor assistant in the department of electronic systems of Don State Technical University.*

Vladimir Marchuk *was born in 1951. Marchuk Vladimir received the Ph.D. degree in technics from Southern Federal University (Russian Federation) in 2006. Since 2006, he has been a Professor. His research interests are in the areas of applied statistical mathematics, signal and image processing.*

Stradanchenko Sergej *was born in 1972 in the Rostov region. He graduated Novocherkassk State Technical University in 1994. From 1994 to 2001, he worked in the NSTU as an assistant, senior lecturer and associate professor. In 2003, he defended his doctorate thesis.*

Karen Egiazarian *(SM'96) was born in Yerevan, Armenia, in 1959. He received the M.Sc. degree in mathematics from Yerevan State University in 1981, the Ph.D. degree in physics and mathematics from Moscow State University, Moscow, Russia, in 1986, and the D.Tech. degree from the Tampere University of Technology (TUT), Tampere, Finland, in 1994. He has been Senior Researcher with the Department of Digital Signal Processing, Institute of Information Problems and Automation, National Academy of Sciences of Armenia. Since 1996, he has been an Assistant Professor with the Institute of Signal Processing, TUT, where he is currently a Professor, leading the Spectral and Algebraic Methods in DSP group. His research interests are in the areas of applied mathematics, signal processing, and digital logic.*