Real-time Stable Texture Regions Extraction for Motion-Based Object Segmentation

Dubravko Ćulibrk
http://www.dubravkoculibrk.org
Borislav Antić
http://www.ursusgroup.com
Vladimir Crnojević
http://www.ursusgroup.com

Faculty of Technical Sciences
University od Novi Sad
Serbia

Although used extensively for object recognition, texture has for the most part been ignored as a feature used for background modelling and object segmentation. The complexity of working with texture descriptors for segmentation in videos is two-fold: the descriptive features cannot be calculated in real time and features extracted based on arbitrarily chosen regions or blocks in the frame are not stable enough to allow for building models sufficiently accurate, yet simple enough to be used for real-time segmentation. The paper proposes an approach that can be used to detect regions of texture, stable enough to be modelled using probabilistic models commonly used for foreground segmentation. Based on the evaluated stable texture regions, a discriminative texture descriptor is proposed that can be evaluated in real time.

When video is captured from a stationary camera, the background is expected to be stationary to a degree and an adaptive model can be built to serve as basis for segmentation. For reasons of efficiency most adaptive models are learnt on a per pixel basis [4]. Joint domain-space modelling [3], as well as modelling of the sequence at two different scales, has been suggested as a way to enhance the segmentation and escape the limitations of single-pixel-based models[1]. The improved approaches still fail to consider the texture properties of the objects in the background and suffer from the noise introduced by neighboring objects being modelled as one.

Texture of the background can be used to build models in a more informed way. Large stable regions in the texture of the background allow for larger scale modelling and more precise extraction of texture related features. Conversely, unstable regions in terms of texture correspond to transitions between different objects in the background, and the features should be extracted for each of the objects separately. An approach to background modelling and foreground segmentation derived form these principles is proposed in the paper. In each frame of the sequence stable texture regions are determined, texture descriptor (feature) values extracted using the stable regions information and used to model the background and segment the foreground using standard probabilistic approaches. The process is illustrated in Fig. 1, where the brighter patches in the stability map correspond to more stable regions, in terms of texture.

Once the size of the stable-texture region at each location is determined, arbitrary descriptive features can be extracted from the frames of the sequence. In addition, one can choose any of a number of probabilistic methods [4][2] to learn the statistics of the features.

The size of the stable texture regions can be determined online, using integral images [6]. Observe that:

$$\begin{aligned} \mu(\bar{B}(x,r)) &= 4r^2\bar{I}(x,r) \\ \Rightarrow \log\mu(\bar{B}(x,r)) &= 2\log r + \log 4\bar{I}(x,r) \end{aligned} \quad (1)$$

where $\bar{I}(x,r)$ represents the average intensity of the pixels in a block of size size $2r+1$ centered at pixel x. Bearing the Eq. 1 in mind, it is to be expected that the logarithm of the pixel intensity sum would change somewhat linearly as the block size grows, if the blocks encompass pixels with the same texture. On the other hand, if the texture changes between successive block sizes, this trend would be violated. One can detect the sudden changes in the texture by monitoring how the average intensity changes when the block size is increased. The proposed approach for detecting stable texture regions and subsequently the maximum scale at which a pixel can be effectively modelled proceeds as follows: The logarithms of sums of intensity values for successive blocks of "radius" $r \in \{1,2,\ldots,R\}$ centered at each pixel location x are computed; the derivative of these values across the scales is computed; if the variance of the derivative values for a location is less than the mean variance for the whole frame the region is marked as stable; if not, the maximum derivative value in the sequence is selected as the radius at which the change in the texture is encountered and the radius of the stable region around location x set to the radius of the last stable block.

Figure 1: Stable texture regions based segmentation: frame from a sequence (top-left), region stability map (top-right), extracted texture descriptors (bottom-left), segmented foreground (bottom-right).

Once the stable regions are determined, texture related features can be extracted reliably to be used for segmentation. New texture descriptors (features) are defined in the paper that can be calculated efficiently using integral images, based on observations of Varma and Garg [5]. The feature is the y-intercept of the plot of $\log\bar{I}(x,r)$ across several values of r. Linear least mean squares regression is used to calculate the value of the intercept.

These features are stable, discriminative enough and can be computed efficiently, allowing for real time foreground segmentation, using any of the standard probabilistic approaches. Experiments conducted on diverse sets of sequences show that the widely used MoG [4] segmentation benefits from the introduction of such features.

[1] Yu-Ting Chen, Chu-Song Chen, Chun-Rong Huang, and Yi-Ping Hung. Efficient hierarchical method for background subtraction. *Pattern Recogn.*, 40(10):2706–2715, 2007. ISSN 0031-3203. doi: http://dx.doi.org/10.1016/j.patcog.2006.11.023.

[2] D. Culibrk, O. Marques, D. Socek, H. Kalva, and B. Furht. Neural network approach to background modeling for video object segmentation. In *IEEE Trans. on Neural Networks*, volume 18, pages 1614–1627, 2007.

[3] Yaser Sheikh and Mubarak Shah. Bayesian modeling of dynamic scenes for object detection. *IEEE Trans. Pattern Anal. Mach. Intell.*, 27(11):1778–1792, 2005. ISSN 0162-8828. doi: http://dx.doi.org.ezproxy.fau.edu/10.1109/TPAMI.2005.213.

[4] C. Stauffer and W. Grimson. Learning patterns of activity using real-time tracking. In *IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 22, pp. 747-757*, 2000.

[5] Manik Varma and Rahul Garg. Locally invariant fractal features for statistical texture classification. In *ICCV '07: Proceedings of the IEEE International Conference on Computer Vision*, 2007.

[6] Paul Viola and Michael Jones. Robust real-time object detection. In *International Journal of Computer Vision*, 2001. doi: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.23.2751.