

Direction Selective Contour Detection for Salient Objects

Andrea Manno-Kovacs, *Member, IEEE*

Abstract—The active contour model is a widely used technique for automatic object contour extraction. Existing methods based on this model can perform with high accuracy even in case of complex contours, but challenging issues remain, like the need for precise contour initialization for high curvature boundary segments or the handling of cluttered backgrounds. To deal with such issues, this paper presents a salient object extraction method, the first step of which is the introduction of an improved edge map that incorporates edge direction as a feature. The direction information in the small neighborhoods of image feature points are extracted, and the images' prominent orientations are defined for direction-selective edge extraction. Using such improved edge information, we provide a highly accurate shape contour representation, which we also combine with texture features. The principle of the paper is to interpret an object as the fusion of its components: its extracted contour and its inner texture. Our goal in fusing textural and structural information is twofold: it is applied for automatic contour initialization, and it is also used to establish an improved external force field. This fusion then produces highly accurate salient object extractions. We performed extensive evaluations which confirm that the presented object extraction method outperforms parametric active contour models and achieves higher efficiency than the majority of the evaluated automatic saliency methods.

Index Terms—Direction selectivity, Harris based vector field convolution, active contour, saliency, boundary detection.

I. INTRODUCTION

ACTIVE contour models (snakes) are widely used for automatic object contour extraction. The original model was introduced in [1], where the so called snake energy term includes an internal and an external force, jointly controlling the evolution of the contour curve. The internal force represents the shape of the curve (with elasticity and rigidity terms) and the external force is calculated from the image characteristics. Since the publication of the original theory, several modifications were introduced, which can be divided into two main groups: parametric [2], [3], [4], [5] or region based [6], [7], [8] approaches. These energy minimization models are used to search for contours perceived as closed curves after an automatic initialization of the region of interest, therefore connected and closed loop edge systems are required to form an object boundary. Unlike other contour detections, where all the edges are extracted [9].

The original active contour model has some disadvantages regarding its sensitivity for initialization, parameters and noise. Moreover, it is not able to properly detect concave boundaries.

A. Manno-Kovacs is with the Machine Perception Research Laboratory, Institute for Computer Science and Control, Hungarian Academy of Sciences (MTA SZTAKI), H-1111, Kende u. 13-17, Budapest, Hungary. E-mail: andrea.manno-kovacs@sztaki.mta.hu.

Manuscript received ...

Region based active contours (or level set approaches [10]) related to the Mumford-Shah energy model [11] bring an efficient solution for the initialization sensitivity, and such models (e.g., [7]) have the ability to detect multiple and complex objects in the image. On the other hand, while parametric active contour models are more sensitive to noise, parameter and contour initialization, and have issues with detecting high curvatures, their convergence is much faster and they can handle natural images with varying intra-object textures. Thus, the application of parametric models for automatic natural image segmentation, including salient object extraction [12], [13], [14], [15], [16] is well justified.

Addressing the difficulties of the traditional active contour model [1], a group of parametric approaches were intended to redefine and improve the external energy of the contour [2], [3], [4], [5], [12], [17], [18]. Although these methods perform with high accuracy even in case of complex contours, there are still some challenging issues, even in recent works [18], [12]. E.g., the presence of high curvature boundary segments may lower the detection accuracy. Moreover, cluttered background data is often a source of detection problems and contour initialization anomalies. A precise automatic initialization step could facilitate accurate detection with good localization and it could support the automatic processing of large datasets by eliminating manual intervention. Earlier automatic initializations took shape information into account [19] or required a manually labeled point inside the object region [20]. Another solution for initialization is to define the region of interest by estimating the focus area [21] or by salient region extraction [22], [23], [24], which can be especially efficient in case of cluttered backgrounds. To also handle textural variation, a recent work [12] introduced an automatic initialization for parametric active contours, using texture atoms and their distinctiveness to build a saliency map. However, using only texture atoms can be problematic if the target object contains multiple textures, potentially causing object fragmentation or incomplete detection.

The goal of this paper is to handle complex contours more efficiently, even in the presence of a cluttered background, which is often the case for real life images. To achieve this goal, an improved edge map is introduced, which incorporates direction information. The improved edge information provides a more accurate shape contour representation, which we also combine with texture information. The fusion of textural and structural information is twofold: first, it is applied for an automatic contour initialization and second, an improved external force field is produced by fusing direction, edge and texture data. The principle of the paper is to interpret the object as the fusion of its components: its extracted contour improved

by direction information and its inner texture.

Automatic contour initialization and salient object extraction is based on efficient salient region estimation. With an accurate saliency map, a high quality object representation can be defined for parametric active contour initialization and external force construction, resulting in a more efficient object detection. Therefore, the main contribution of this work also serves as an efficient saliency representation, fusing textural and structural features, even in case of images with cluttered backgrounds. In such cases, using only textural information (as [12]) might falsely emphasize some parts of the background (second column of Fig. 2). To deal with this issue, we propose to apply salient point detection along with the texture feature to integrate the structural information into the process. Besides traditional salient point detectors [25], [26], an enhanced Harris detector [25] was introduced in [27] called the Modified Harris for Edges and Corners (*MHEC*). The MHEC salient point detector was proved to be an efficient tool for object contour detection [28], as it is emphasizing edges and corners in a balanced manner, which makes it a useful tool in the present framework. By integrating the MHEC point set and the textural distinctiveness map we introduce a highly accurate automatic contour initialization step which enables higher detection accuracy and faster convergence as well.

Another contribution of the paper is the integrated orientation feature in the active contour model, fused with structural and textural information. Orientation as a feature has already been used in earlier works in different forms [3], [29], [30]. The Dynamic Directional Gradient Vector Flow (DDGVF) [3] concentrates only on positive and negative step edges, by using the gradient in both x and y directions and dealing with the external force field for the two directions separately. The importance of the direction of the normal is also exploited in [29], by introducing the GVF in the normal's direction (NGVF) instead of the x and y directions. A more recent work [30] introduces a direction vector flow in the external force, resulting in a gradient & direction vector flow model (G&DVF). However, this technique requires manual interaction by drawing a few directional lines after the initialization step to guide the contour towards the correct boundary. This user interaction makes the method inappropriate for automatic detection purposes in huge databases.

In this paper, orientation information is handled in a different manner: the main direction in the small neighborhood of the MHEC points is calculated, then this information is integrated into an improved edge map. Therefore, instead of using histogram bins and manually labeled points, specific orientation information is extracted automatically for the most prominent pixels and extended to the region of interest. Direction information is then integrated into the edge map, introducing an improved linear extension of the Morphological Feature Contrast (MFC) operator [31] which is applied for extracting features only in the defined directions. As a novel contribution, this improved structural information is fused with the texture distinctiveness map and a hybrid, 3-term external force is introduced for iterative active contour detection.

The main steps of the proposed method are shown in Fig. 1: first the texture distinctiveness map is calculated as

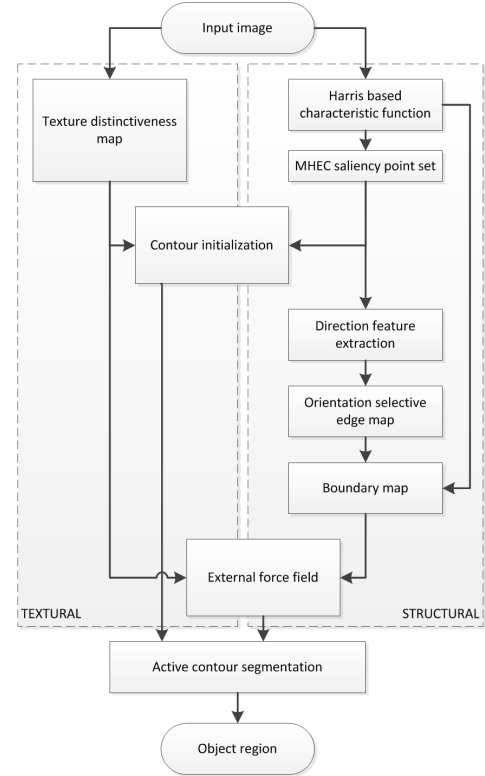


Fig. 1. Main steps of the proposed method.

the textural feature along with the MHEC saliency point set based on the modified Harris characteristic function to perform contour initialization. The direction feature is defined based on the MHEC points and it is integrated into the orientation selective edge map. After constructing the structural feature as a boundary map, then fusing the characteristic function and the improved edge map, the external force is defined. Finally, the iterative active contour algorithm segments the object region.

To support the introduced method, a detailed evaluation is performed on multiple widely known and used public databases: SED and SED2 [32] (100 – 100 images), EPFL [33] (1000 images), MSRA10K [22] (10000 images), SUN [13] (346 images). In the first part of the experiments the proposed technique is compared to other parametric active contour models to confirm the advantages of the automatic initialization and the improved external force. In the second part the proposed method is tested from the point of view of salient object segmentation, and it is compared to different state-of-the-art segmentation techniques. Moreover, with an iterative extension, the introduced algorithm is able to detect multiple objects. The performance of the method is also evaluated quantitatively. The extensive evaluation results confirm that the proposed method is performing with high accuracy and that - using the presented contributions - it outperforms existing parametric active contour models and achieves higher efficiency than the majority of the compared automatic saliency detection techniques.

The paper is organized as follows: Section II is about the related state-of-the-art, Section III introduces the texture and

structure based automatic contour initialization, Section IV gives an introduction on related active contour models, Section V describes the main contribution, the Textural-Directional Harris based Vector Field Convolution (TDHVFC). The experimental validation is carried out in Section VI and the conclusions and future plans are summarized in Section VII.

II. RELATED WORKS

This section concentrates on related methods, used for comparisons during the evaluations (Sec. VI). Since the proposed method is a parametric active contour method, we selected other parametric algorithms with different principles to compare to. Moreover, the introduced method is also compared to salient object extraction techniques to measure segmentation performance.

Motivated by the drawbacks of the original parametric active contour method, a novel external force field called the Gradient Vector Flow (*GVF*) was introduced in [2]. The improved external force was calculated as a diffusion of the gradient vectors, which created the opportunity for a more flexible initialization and convergence to concave boundaries as well. However, the method was still initialization-sensitive and converged slowly, therefore the Vector Field Convolution (*VFC*) was published in [4] for improving the performance of the *GVF*. The *VFC* was a novel vector field kernel, which was convolved with the edge map. However, based on a structural edge map, the *VFC* still had problems with cluttered backgrounds and high curvature boundaries, and required manual initialization.

In the previous works of the author [5], [28], [34] a Harris based improvement was introduced for the *GVF* and the *VFC*, called *HGVF* and *HVFC* respectively. Motivated by the Harris detector's characteristic function [25], an improved feature map was introduced based on the eigenvalues of the Harris matrix, representing the principal curvature. The improved feature map was then applied in the external force instead of the intensity image, which resulted in a more accurate boundary detection process, especially in the case of high curvature, noisy boundary segments. Moreover, the local maxima of the improved feature map served as salient keypoints (MHEC point set), and their convex hull provided an automatic initialization step before the iterative detection process.

In another approach, the hybrid Structure and Texture distinctiveness Vector Field Convolution (*STVFC*) [12] exploited the textural information of the image beside the structural data, and introduced the texture distinctiveness map for contour initialization and hybrid external force calculation. Following a different approach, the same authors recently published the Enhanced Decoupled Active Contour (*EDAC*) model [18], which is decoupling the internal and external forces and the two forces are optimized individually, designed for accurate boundary extraction, despite image noise and complex object geometries.

Visual attention modeling is an active field of research, it is also an important segmentation step in different application field and workflows, such as video analysis [35], sketch-based image retrieval [36] or scene classification in remote sensing [37]. Thus, in the second part of the experiments some prominent and recently published state-of-the-art methods have been

selected, which were included in an extensive evaluation by M. M. Cheng and his colleagues in [38] and the evaluation results are present on his website [39]. The following algorithms are selected (following the abbreviations of [39]): Frequency-Tuned salient region detection (*FT*), introduced in [33], which preserves salient object boundaries by retaining substantially more frequency content from the original image than other earlier techniques. The method exploits luminance and color features, and it publishes the EPFL database. The *SEG* algorithm [40], adapts a saliency measure using a statistical framework and local feature contrast in illumination, color and motion information. The achieved saliency map is then integrated into a conditional random field model to define an energy minimization based segmentation. A Context and shape prior Based segmentation (*CB*) was introduced in [41], where the initial saliency map is computed through a multi-scale superpixel based context analysis followed by an object-level shape prior extraction, achieved as a combination of saliency and object boundary information. Both the saliency map and the shape prior are updated in each iteration. Sparse Salient regions (*SS*) [42] introduced a simple image descriptor, the image signature, and showed that within the theoretical framework of sparse signal mixing it is able to spatially approximate the foreground in an image. The resulting image signature based saliency algorithm can predict human fixation points with high accuracy. Saliency detection via Dense and Sparse Reconstruction (*DSR*) is a method from the perspective of reconstruction errors [43]. The algorithm is based on background templates and a context-based mechanism is designed to propagate reconstruction errors. A pixel-level saliency and an object-based Gaussian model are computed and combined with a Bayesian integration step. Another work in the same year [44] introduced the Bayesian saliency via Low and Mid Level Cues (*LMLC*), which is a bottom-up approach, obtaining a coarse saliency region and analyzing saliency information with mid level visual cues via superpixels. Superpixels are grouped with a Laplacian sparse subspace clustering method and are used to construct a prior saliency map. Finally, low level visual cues are applied to compute an observation likelihood adaptive Bayesian saliency. Global saliency cues, such as global uniqueness and color spatial distribution are integrated in [45] to construct a final Global Cues (*GC*) saliency map. Instead of computing saliency maps by combining feature maps, the discriminative regional feature integration (*DRFI*) [46] method proposed to learn a random forest regressor to map the regional feature vector to a saliency score. Based on higher dimensional features, [47] introduced the High-Dimensional Color Transform (*HDCT*), where the saliency map of the image is calculated as a linear combination of high-dimensional color spaces, including RGB, CIE Lab and HSV with different gamma corrections. In [48], a novel algorithm design was proposed, called *CPMC+GBVS*, along with a new dataset to bridge the gap between the challenges of fixation prediction and salient object segmentation. A very recent work, *RC* [22] introduced global contrast based salient region detection. The proposed algorithm simultaneously evaluates global contrast differences and spatially weighted coherence scores, producing high-quality saliency maps. Moreover, the

novel iterative SaliencyCut algorithm is introduced for salient object segmentation. This multi-pathway-base model is also applied in the latest works, such as [49], which adapts the guided search strategy of biological vision.

The use of machine learning methods for salient object segmentation is a novel trend. Recently, different methods were proposed applying such techniques: Stacked Denoising AutoEncoders (*SDAEs*) were introduced in [14], [50], [51] for building background models with deep learning architectures. Afterwards, salient object extraction was performed as a separation step by measuring the reconstruction residuals of deep autoencoders. In [52], a new loss function was proposed instead of the traditional softmax loss function, called positive-sharing loss. The model partitioned contour (positive) data into subclasses and each subclass shared the loss for the whole positive class in the loss function.

Deep learning based methods usually operate at the patch level, causing blurry saliency maps, especially near the object boundaries. To compensate for these weaknesses, a Multiscale Deep Feature (*MDF*) model was proposed in [53], with fully connected layers on top of convolutional neural networks for feature extraction at three different scales. The aggregation of the multiple saliency maps of different levels resulted in more accurate final saliency maps. In a similar manner, Deep Contrast Learning (*DCL*) [54] was a novel network, consisting of two complementary components, a pixel-level fully convolutional stream and a segment-wise spatial pooling stream. The first component supported pixel-level accuracy, while the second extracted segment-wise features efficiently.

The orientation feature has been used in the literature of object segmentation. An unsupervised object extraction method was published in [13], interesting mainly for its use of Oriented edges for Contour delineation and Texture discrimination (*OCT*). The approach separated oriented edges (horizontal, $+45^\circ$, vertical, -45°) and used the orientation statistics for extracting contour candidates from multiresolution edge maps. Oriented edges also represented texture information, later analyzed by K-means clustering. The object region was defined by merging the results of the two separate analysis channels based on the simple assumption that the object was located centrally in the scene. Although this method handled orientation information differently than the proposed approach both conceptually and practically, we included it in the evaluations.

III. TEXTURE AND STRUCTURE BASED AUTOMATIC CONTOUR INITIALIZATION

A. Texture distinctiveness

Statistical textural distinctiveness is based on a sparse texture model, which was introduced in [23]. In this method rotation-invariant neighborhood-based textural representations are extracted and used to learn a set of representative texture atoms for defining a sparse texture model for the image. After calculating the $t(x, y)$ texture representation for all $U \times V$ pixels in the image, the global texture model of the image is defined: $T = t_1, t_2, \dots, t_{U \times V}$. As textures usually consist of repeating patterns, a smaller number of regions with unique

patterns can be defined as atoms, resulting in a sparse texture model of M textures: $T^r = t_1^r, t_2^r, \dots, t_M^r$. Following the recommendations of [23], [12], we used $M = 20$. Texture distinctiveness means a difference among textured regions that draws a viewer's attention. Statistical texture distinctiveness measures the relationships between atoms of T^r and indicates how unique the texture characteristics of a region are.

The $S(x, y)$ texture distinctiveness map quantifies the expected distinctiveness of each texture atom compared to the others. The calculated distinctiveness value is assigned to all pixels within the corresponding region. The following assumptions are used in the calculation: 1. Higher statistical texture distinctiveness defines more distinct regions. 2. Objects close to the center attract human visual attention more, therefore, increasing a region's distinctiveness. Examples for final S texture distinctiveness maps are in the second column of Fig. 2 and in Fig. 3(b). As the detailed description of texture distinctiveness calculation is out of the scope of this paper, please see [23] and [12] for further information.

B. Structural information for improved contour initialization

A well-performed initialization may result in a more accurate segmentation and also accelerates the convergence, which are both important issues for active contour methods. In case of large databases, automatic initialization is inevitable, as manual intervention is impractical. The generated S texture distinctiveness map can be efficiently used for automatic contour initialization, as it emphasizes the distinct image regions. In [12] the S map is adaptively thresholded by Otsu's method [55], and the resulting binary map contains connected regions of different sizes. The region of maximum area among all regions is selected (ROI_S), shown in the third column of Fig. 2. The initial contour is defined as the convex hull around the ROI_S , calculated using the Quickhull algorithm [56]. Convexity is sometimes also added as a bottom-up cue for salient object detection [57], however in our case we use it as a fast and efficient way for ROI estimation, in a more improved way than it is proposed in [58].

The S texture distinctiveness map works effectively when the salient object differs enough from other background regions of the image. However, in some cases, if multiple distinct textures are present around the center of the image, they all get high values in the S map, which causes the initialization to be far from the real contour outline (third column of Fig. 2). This may result in the false convergence of the contour to local minima. This can be observed typically when objects and their shadows are both present in the image (see first row of Fig. 2), and both are calculated to be highly distinctive, therefore the initialization contains both regions and the active contour might converge to the shadow's contour instead of the real object outline. Structural information can help in these cases, as shadows often have blurry outlines.

To avoid misdetection, this paper proposes an interest point based refinement step using the Modified Harris for Edges and Corners (MHEC) point set [5]. The MHEC point set, introduced as a modification of the original Harris detector [25], emphasizes edges and corners in the image, therefore it

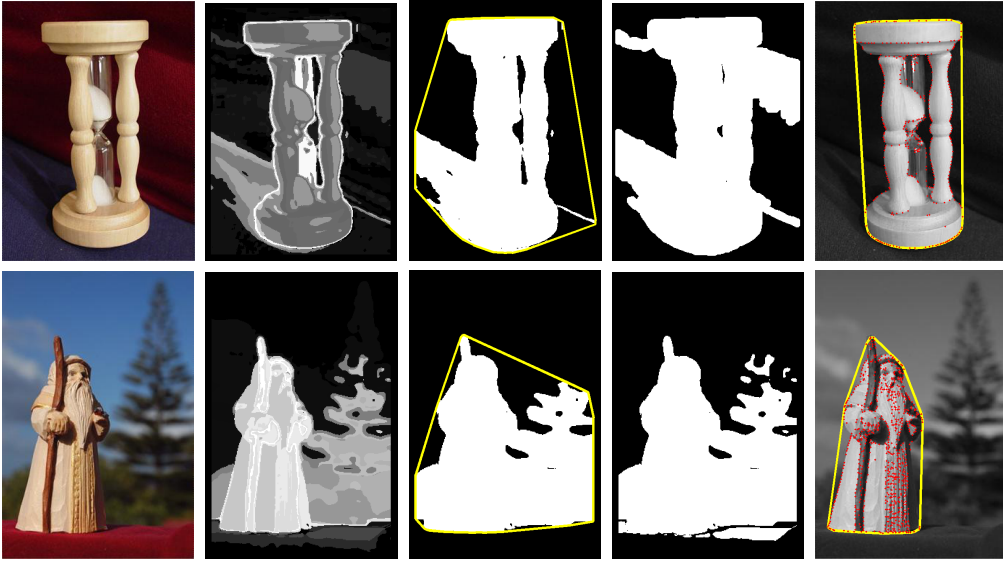


Fig. 2. Contour initialization with improved texture distinctiveness. (First column) Original image. (Second column) Calculated S texture distinctiveness map. (Third column) ROI_S with its convex hull (yellow). (Fourth column) The extended $ROI_{S/2}^D$. (Fifth column) MHEC point set in $ROI_{S/2}^D$ (red) and refined initial contour (yellow).

can represent object contours efficiently and it is applied as a source of structural information in this work.

By considering the behavior of the Harris matrix and its λ_1 and λ_2 eigenvalues (both of them large in corner regions, only one of them large in edge regions and both of them small in flat regions), the following modification of the R characteristic function was proposed in [5] to emphasize edge and corner regions simultaneously:

$$R_{\text{mod}} = \max(\lambda_1, \lambda_2). \quad (1)$$

As both corner and edge regions have at least one large λ component, the R_{mod} function effectively separates flat and non-flat regions.

Since feature points are located on the edges and corners of an object, the MHEC feature point set is calculated as the local maxima of R_{mod} : if a pixel $p_i = (x_i, y_i)$ has the largest $R_{\text{mod}}(p_i)$ value compared to its neighbors in a surrounding $b_i = \{[x_i - 1, x_i + 1] \times [y_i - 1, y_i + 1]\}$ window and its $R_{\text{mod}}(p_i)$ value exceeds a given T_{max} threshold (calculated by Otsu's method [55]), then p_i is included in the MHEC point set:

$$P_{\text{MHEC}} = \left\{ p_i : R_{\text{mod}}(p_i) > T_{\text{max}} \text{ AND } p_i = \underset{r \in b_i}{\operatorname{argmax}} R_{\text{mod}}(r) \right\}. \quad (2)$$

In this paper we propose the application of the MHEC point set combined with texture distinctiveness to extract efficient object features as the fusion of texture and structure. While the S texture distinctiveness map concentrates on the textures of the image and defines the ROI_S based on their distinctiveness, the MHEC focuses on the edges and contours. In order not to miss any important contours, the adaptive Otsu threshold used for S is reduced to its half (marked by $S/2$). Then, the selected $ROI_{S/2}$ is morphologically dilated with a 5-radius disk structural element ($ROI_{S/2}^D$) as important points may be

located on the edges of the region (fourth column of Fig. 2). After these assumptions, a p_i point of P_{MHEC} is kept only if it is located in the dilated $ROI_{S/2}^D$, creating a P_{MHEC}^S point set. As shadow and unfocused regions usually do not contain sharp edges and corners (see the original images in Fig. 2), MHEC points (marked with red in the fifth column of Fig. 2) will not be located in these areas. By calculating the convex hull of P_{MHEC}^S (shown in yellow in the fifth column of Fig. 2), the result is a reduced $ROI_{S+\text{MHEC}}$ area, with superfluous regions omitted. On the other hand, a salient object in the image draws visual attention and is usually in focus, therefore its outline will be effectively featured by the MHEC point set.

IV. ACTIVE CONTOUR MODELS

A. Traditional model

The aim of the traditional active contour model [1] is to find the curve $\mathbf{x}(s) = [x(s), y(s)]$, $s \in [0, 1]$ minimizing the E energy function:

$$E = \int_0^1 \frac{1}{2} (\alpha |\mathbf{x}'(s)|^2 + \beta |\mathbf{x}''(s)|^2) + E_{\text{ext}}(v(s)) ds, \quad (3)$$

where the α elasticity parameter and the β rigidity parameter control the first part of the energy, called the internal energy. This energy part represents the active contour's smoothness and rigidity: $\mathbf{x}'(s)$ and $\mathbf{x}''(s)$ are the first and second order derivatives with respect to arc length s . Following the recommendations of [4], $\alpha = 0.5$ and $\beta = 0.1$ are used throughout the experiments for smooth curves with low rigidity and to also detect corners. E_{ext} is the external energy term which is derived from the image, representing the image characteristics (i. e. edges, ridges). For the improved GVF [2] model it has the following form:

$$E_{\text{ext}} = \int \mu (u_x^2 + u_y^2 + v_x^2 + v_y^2) + |\nabla f|^2 |\mathbf{v} - \nabla f|^2 dx dy, \quad (4)$$

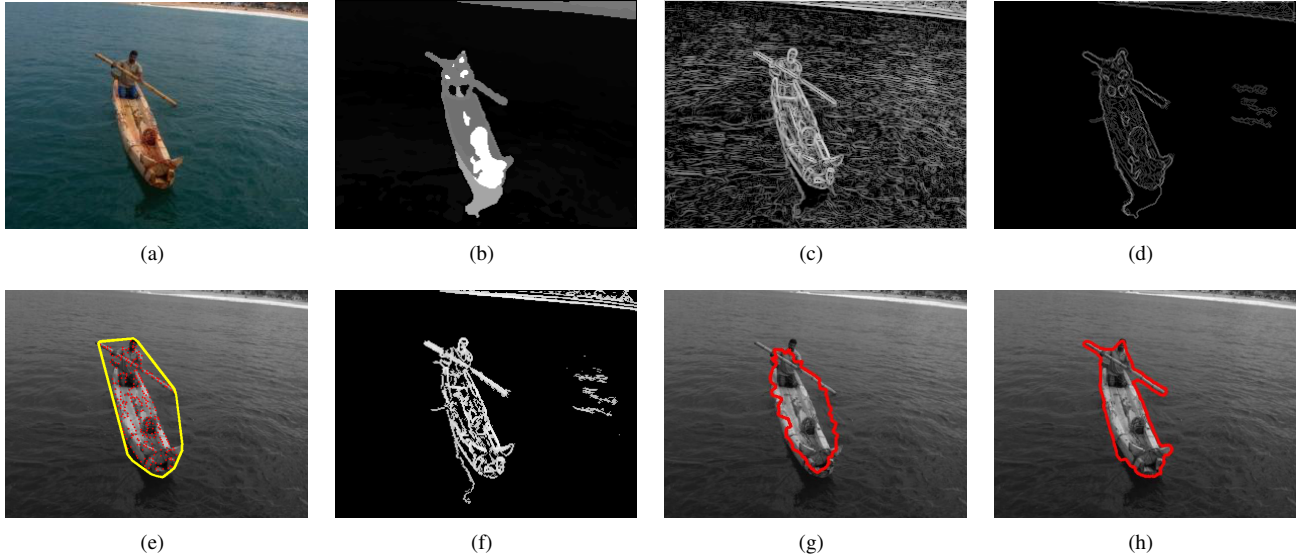


Fig. 3. Main steps of the direction selective contour detection: (a) is the original image from SED dataset [32]; (e) shows the generated MHEC point set in red and the initial contour in yellow; (b) is the S texture distinctiveness map; (f) is the improved R_{MFC} feature map; (c)-(d): the HVFC and TDHVFC feature maps, (g)-(h): the result of the contour detection for HVFC and TDHVFC.

where ∇ is the gradient operator, and μ is a regularization parameter which should be adjusted according to the noise constant of the image. For a higher noise level a higher μ value is selected. For natural images with a lower noise level, $\mu = 0.2$ is used throughout the experimental evaluation. The edge map, denoted by f , is generated from the I image, defined as:

$$f(x, y) = |\nabla(G_\sigma(x, y) * I(x, y))|, \quad (5)$$

where G_σ is a Gaussian function with $\sigma = 0.2$ standard deviation.

The basic f edge map (Eq. 5) had difficulties with detecting sharp or noisy corners and low contrast boundaries. When heading towards such boundary segments in the iterative boundary detection, the f edge map shows low values. This causes the active contour method to fail in accurately converging to these outlines. Cluttered backgrounds also mean a great challenge and often cause the contour to get stuck in a local minimum.

B. Vector Field Convolution model

Vector Field Convolution (VFC) was introduced in [4] as a novel external force term and to compensate for some drawbacks of GVF, like high computational cost, noise sensitivity, parameter sensitivity and the ambiguous relation between the capture range and the algorithm parameters by integrating a vector field kernel into the external force field. Therefore, unlike in the case of the GVF, the external force field becomes dynamic and it is extended to the homogeneous regions of the image. Thus, a free particle is able to move to the desired contour parts from a larger distance, and faster, even in the presence of noise. The novelty of the algorithm was an external force field calculated as the convolution of a vector field kernel and the f edge map (Eq. 5):

$$\mathbf{f}_{VFC}(x, y) = f(x, y) * \mathbf{k}(x, y), \quad (6)$$

where $\mathbf{k}(x, y)$ is the introduced vector field kernel, defined as a multiplication of a magnitude function and a unit vector pointing to the kernel origin, ensuring the force towards the desired image parts:

$$\mathbf{k}(x, y) = m(x, y)\mathbf{n}(x, y), \quad (7)$$

where $m(x, y)$ is the magnitude at (x, y) and $\mathbf{n}(x, y)$ is a unit vector pointing to the kernel origin and calculated as $\mathbf{n}(x, y) = [-\frac{x}{r}, -\frac{y}{r}]$ with $r = \sqrt{x^2 + y^2}$ denoting the distance from the kernel origin. The $m(x, y)$ magnitude is chosen as a decreasing positive function of the distance from the origin to represent a reducing effect:

$$m(x, y) = (r + \epsilon)^{-\delta}, \quad (8)$$

where ϵ and δ are positive parameters, the latter controlling the rate of decrease, the influence of the selected ROI and ϵ preventing division by zero (chosen as a small positive constant). In [4] δ is advised to be chosen as $1.5 \leq \delta \leq 3$; in the experiments $\delta = 3$ is applied, which represents a higher influence of the ROI obtained from the initialization step.

As the VFC external force field was also based on the f edge map, it still has issues in correctly detecting high curvature and noisy, weakly contrasting boundaries.

C. Harris based Vector Field Convolution

As the f edge map (Eq. 5) is not able to emphasize high curvature boundary parts, in our previous work [28] we introduced a modification of the edge map inspired by the Harris corner detector's [25] characteristic function. The same edge map modification was applied for the GVF [2] in an extended work in [5]. The principle behind the idea was that high curvature contour segments can be emphasized by a corner detector's characteristic function efficiently, as such parts cohere with corners. Therefore, if the characteristic

function is integrated into the edge map, then these regions are also emphasized, resulting in a balanced feature map.

The dynamic range of the R_{mod} function (Eq. 1) is very wide, therefore, to produce a steady feature map, we need to compress the dynamics of the R_{mod} function into a balanced distribution, while keeping the main attractors. To fulfill these requirements, we define the R_{logmax} function, where the natural logarithm supports the balanced compression and $\max()$ replaces the negative logarithm values of small R_{mod} values with zeros in the feature map:

$$R_{\text{logmax}} = \max(0, \log(R_{\text{mod}})). \quad (9)$$

Thus, the originally used $I(x, y)$ intensity in Eq. 5 will be replaced by the R_{logmax} modified Harris characteristic function and thus will be able to emphasize complex contours efficiently in the f feature map, proposed in [28]:

$$\mathbf{f}_{\text{HVFC}} = |\nabla(G_\sigma(x, y) * R_{\text{logmax}}(x, y))| * \mathbf{k}(x, y). \quad (10)$$

The drawback of using the R_{logmax} function is that the feature information of the background is also emphasized. Therefore, in case of large background clutter, large feature values appear in the \mathbf{f}_{HVFC} map for the background (Fig. 3(c)), causing the contour to converge incorrectly (Fig. 3(g)). To improve the feature map and eliminate the mentioned disadvantage, we introduce two solutions: 1. integrate the orientation information into the \mathbf{f}_{HVFC} map, for emphasizing relevant contours and preserving the benefits of the R_{logmax} function at the same time; 2. the internal texture should also be represented in the external force term, aiding the accurate convergence of the contour.

V. TEXTURAL-DIRECTIONAL HARRIS BASED VECTOR FIELD CONVOLUTION

The principle of this paper is to represent the object as the unity of its internal texture and its surrounding outline, integrating both these parts into the feature map. By exploiting additional knowledge of directionality, a higher level structural feature can be constructed for more accurate detection.

A. Direction feature extraction

The advantages of using local direction as a feature have already been shown in applications for remote sensing in previous works of the author [27], [59]. As P_{MHEC} points (Eq. 2) are located on important contour segments, the relevant direction information can be extracted from their surroundings. For obtaining such information, we analyze the local gradient orientation density [60], [61] in the small neighborhood of each feature point (shown in red in Figure 3(e)) to find the main direction. Let us denote the gradient vector by ∇g_i with $\|\nabla g_i\|$ magnitude and φ_i^∇ orientation for the i^{th} point. By defining the $n \times n$ neighborhood of the point with $W_n(i)$ (where n depends on the resolution and $n = 7$ is applied for the experiments), the weighted density of φ_i^∇ is as follows:

$$\lambda_i(\varphi) = \frac{1}{N_i} \sum_{r \in W_n(i)} \frac{1}{h} \cdot \|\nabla g_r\| \cdot \kappa\left(\frac{\varphi - \varphi_r^\nabla}{h}\right), \quad (11)$$

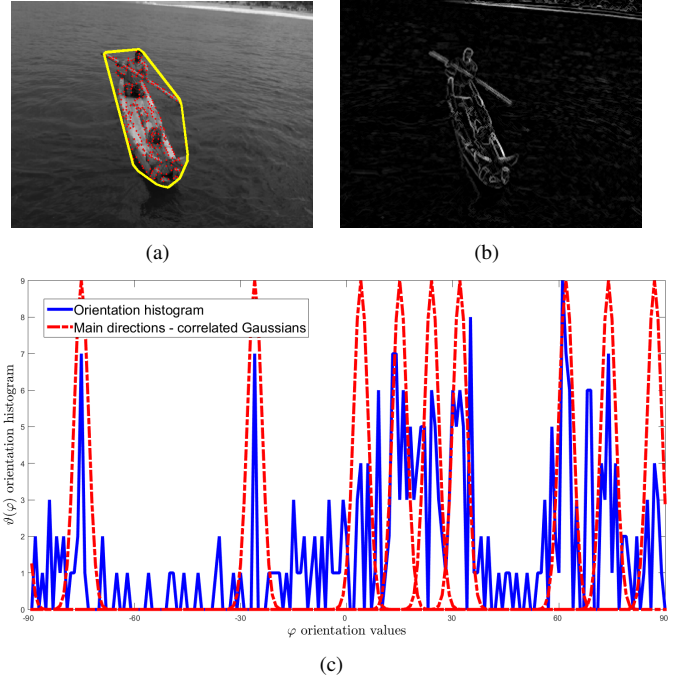


Fig. 4. Calculating the relevant orientations of the ROI: (a) shows the MHEC point set, the φ orientation values are calculated for these pixels, (b) is the result of the improved linear MFC feature detection with (c) showing the $\vartheta(\varphi)$ orientation histogram in blue and the correlated Gaussian functions for the main directions in red.

with $N_i = \sum_{r \in W_n(i)} \|\nabla g_r\|$ and $\kappa(\cdot)$ kernel function with h bandwidth parameter for smoothing the histogram. Following the recommendations of [60], κ should be chosen as a non-negative, symmetric function, therefore a Gaussian smoothing kernel is used with $h = 0.7$. The main orientation for the i^{th} feature point is defined as the most occurring orientation value in the analyzed window:

$$\varphi_i = \underset{\varphi \in [-90, +90]}{\operatorname{argmax}} \{\lambda_i\}. \quad (12)$$

After calculating the direction for all feature points, a $\vartheta(\varphi)$ histogram is obtained (shown in blue in Fig. 4(c)).

To extend the orientation feature for the initialized ROI ($ROI_{\text{S+MHEC}}$ calculated in Sec. III-B), an estimation should be given for all its pixels based on the calculated $\vartheta(\varphi)$ orientation histogram. To calculate the main orientations of the ROI, inspired by [59], a simple Gaussian function is correlated to the $\vartheta(\varphi)$ orientation histogram iteratively to extract the most correlating (most representative) orientation value:

$$\alpha(m) = \int \vartheta(\varphi) \eta(\varphi, m, d_\vartheta) d\varphi, \quad (13)$$

where $\eta(\cdot)$ denotes a Gaussian function, with m mean value and d_ϑ standard deviation.

In every iteration, the most correlating orientation value (maximizing $\alpha(m)$) is selected as the m mean of the Gaussian function (see the correlated Gaussian functions in Fig. 4(c) shown in red). The method stops if: 1. the correlated Gaussians cover a fixed ratio (80%) of the P_{MHEC} points generating the histogram, meaning that the majority of the information

is exploited; 2. the α correlation rate is starting to decrease, meaning that the actual correlation is less representative than the earlier ones (for further details see [59]). The result of the correlation process is a set of orientations $[\theta_1, \dots, \theta_{nr}]$ representing the ROI, where nr is defined by the iterative correlation process adaptively (for example $nr = 9$ for the sample in Fig. 4).

This orientation feature has to be integrated into the structure feature to construct an improved external force instead of the earlier applied $R_{\log\max}$ in f_{HVFC} (Eq. 10).

B. Orientation selective edge map

Different edge extraction approaches can handle directional information in various ways, e.g., the Canny edge detector [62] uses the gradient orientation; [63] introduces shearlets as a theoretically optimal representation of images with edges, capable of fully capturing directional features. However, existing methods cannot handle cases with multiple orientations (corners) [62] and often calculate pixel-level orientation [64] resulting in the loss of the scaling nature of orientation, therefore they are not suitable for edge detection in a higher level interpretation (like object contour detection).

In the present case, contour segments have to be treated as connected pixels with their fixed, calculated main directions. Thus, the selected edge detection method has to be able to handle the extracted orientation values, not simply treating directions as histogram bins [63]. Moreover, as corner points are also involved in the contour, the edge extraction algorithm has to support them as well.

To fulfill these requirements, the Morphological Feature Contrast (MFC) operator [31] is applied, which was introduced for extracting isolated features while not being confused by background texture details. Moreover, an extension of MFC is also presented in the same paper for detecting linear features by handling orientation even in cluttered scenes.

To extract bright and dark individual features, the difference between the original signal and one of its envelopes is used to suppress the background and to enhance features:

$$\psi_{MFC}^+(I) = |I - \rho_{r_2} \tau_{r_1}(I)|^+, \quad (14)$$

$$\psi_{MFC}^-(I) = |\tau_{r_2} \rho_{r_1}(I) - I|^+, \quad (15)$$

where τ is a morphological closing, ρ is a morphological opening, r_1 and r_2 are the size of the square structuring elements (SEs) and I is the image. For the sizes of the r_1 and r_2 structuring elements, the following constraints are defined: $D_1 < r_1 < D_2$, $F_1 < r_1 < F_2$, where D_1 is the maximum distance between texture details, D_2 is the minimum distance between isolated features, F_1 is the maximum size of isolated features and F_2 is the minimum size of texture regions. When choosing the appropriate sizes of the structuring elements, we have to take special care not to miss important features, even if they are very small. Therefore, the applied values are $r_1 = 2$ and $r_2 = 8$, which are determined based on the overall resolutions of the images of the evaluated databases.

To extract both types of individual features (bright and dark) simultaneously, the two operators are summed:

$$\psi_{MFC} = \psi_{MFC}^+ + \psi_{MFC}^-. \quad (16)$$

After extracting the bright and dark features with MFC, specific types of features can be detected by using a sequence of standard morphological transformations. As an extension of MFC a linear feature extraction technique was also introduced in [31], which can be applied for contour detection and has the ability to handle orientation. After removing textures with the ψ_{MFC}^+ and ψ_{MFC}^- operators, a subsequent ρ_{lin} filter is applied in the original algorithm with linear SEs in 12 orientations ranging from 0 to π . The opening is performed for the whole image and the maximum response from all orientations is kept at each pixel. This technique highlights narrow linear features even when longer than the linear structuring element. As an improvement over [31], the ρ_{lin} filter is applied in this paper with linear SEs only in the calculated main directions of the ROI ($[\theta_1, \theta_2, \dots]$), making the linear feature extraction both faster and more accurate. These main directions are characterizing the object and are selected as discrete values, not as histogram bins applied earlier, covering the whole $[0, \pi]$ range. So as not to miss even small linear features, the size of the linear SE is set to 4 in all the experiments, which is determined based on the overall resolution of the images of the evaluated databases. Moreover, the drawback of the original ρ_{lin} filter in also emphasizing texture details, is reduced by the fixed directions and ROI, resulting in a less noisy edge map (Fig. 4(b)).

After applying the MFC operator with the improved linear extension, the resulting feature enhanced (background suppressed) image is denoted by I_{MFC} (see Fig. 4(b)).

C. Feature fusion for external force representation

To represent directional and contour features in a joint feature map, I_{MFC} was rescaled to be fusible with $R_{\log\max}$ in the same way as $R_{\log\max}$ was constructed from R_{mod} (see Eq. 9):

$$I_{MFC, \log\max} = \max(0, \log(I_{MFC})). \quad (17)$$

As small separated features in the ROI might also become emphasized in I_{MFC} , blobs smaller than 100 pixels are eliminated in $I_{MFC, \log\max}$. This threshold is selected according to the resolution of the dataset images.

Finally, the improved edge map, R_{MFC} (Figure 3(f)) combines the advantage of HVFC's high curvature boundary detection ($R_{\log\max}$) with the ability to handle large background clutter by the oriented edge emphasizing feature map ($I_{MFC, \log\max}$) in a pixel level maximization step:

$$R_{MFC} = \max(I_{MFC, \log\max}, R_{\log\max}). \quad (18)$$

At this point, the representation of the boundary outline is provided in R_{MFC} as an improved structural part instead of $R_{\log\max}$ in Eq. 9. As a next step, we also incorporate a region's internal texture into the external force field: the S texture distinctiveness and the R_{MFC} improved edge map are fused into a hybrid feature map, following the recommendations of [12]. By fusing S and R_{MFC} , the improved feature map of

γ value	0.2	0.3	0.4	0.5
Performance	0.8273	0.8299	0.8293	0.8289

TABLE I
AVERAGE F-MEASURE FOR DIFFERENT γ VALUES (EQ. 19).

the Textural-Directional Harris based Vector Field Convolution (TDHVFC) will have the following form:

$$f_{\text{TDHVFC}} = \gamma |\nabla(R_{\text{MFC}}(x, y))| + (1 - \gamma) |\nabla(S(x, y))|, \quad (19)$$

$$\mathbf{f}_{\text{TDHVFC}} = f_{\text{TDHVFC}} * \mathbf{k}(x, y). \quad (20)$$

The f_{TDHVFC} is shown in Figure 3(d). The γ weighting factor balances between the outline and the inner texture. The parameter tuning (Table I) showed that $\gamma = 0.3$ provides the best results. Using the improved external force, the iterative part is the same as for the VFC and HVFC methods [5].

VI. EXPERIMENTAL EVALUATION

A. Comparison with parametric active contour methods

Extensive evaluation has been performed to demonstrate the capabilities of the proposed method for contour detection. Multiple widely used datasets were included in the experimental validation. In the first part of the evaluation process, the proposed method’s reliability was tested on the **SED** single object database [32], which contains 100 images showing a single object over various backgrounds. The database also includes ground truth segmentation, obtained from the manual segmentations into two or three classes by three human subjects. The proposed method was compared to other reviewed parametric active contour techniques, including traditional ones like the GVF [2] and VFC [4], algorithms published previously like the HGVF [5] and HVFC [28], and lately introduced ones as the EDAC [18] and STVFC [12].

Before performing the comparisons, the γ weighting factor, balancing between the outline and the inner texture in Eq. 20, was tested on the SED single object database for different parameter values, which is shown in Table I. Results showed that the $\gamma = 0.3$ selection outperforms other values, which complies with the recommendations of [12]. Therefore, this parameter setting was applied for the whole evaluation process.

SED	F-measure	Recall	Precision
GVF	0.762	0.732	0.795
HGVF	0.784	0.710	0.875
VFC	0.775	0.767	0.784
HVFC	0.795	0.795	0.795
EDAC	0.712	0.675	0.754
STVFC	0.760	0.752	0.769
Proposed	0.830	0.783	0.883

TABLE II
AVERAGE F-MEASURE, RECALL AND PRECISION FOR GVF [2], HGVF [5], VFC [4], HVFC [28], STVFC [12], EDAC [18] AND THE PROPOSED TDHVFC ALGORITHMS FOR THE SED DATABASE [32].

Table II shows the average F-measure, Recall and Precision values calculated for the compared GVF, VFC, HGVF, HVFC, EDAC, STVFC and the proposed TDHVFC methods. The initialization process was the same as explained in Sec. III for all tested algorithms.

The SED database contains foreground objects with a high variety of intra-object textures. This means that there might be objects containing multiple different textures, which makes the initialization more difficult. Although the introduced automatic initialization process was able to work efficiently in most of the cases, it should be mentioned, that it strongly relies on the texture distinctiveness map. I.e., a miscalculated map can lead the initialization to a false segmentation, and initialization problems cannot be fully compensated in later steps.

Applying the same initialization technique for all methods, the performance of the different energy terms can be compared for the active contour algorithms. Moreover, traditional methods do not have their own initialization steps, therefore they do not suffer any disadvantage and the evaluation remains fair. The results in Table II show that the proposed method achieves the highest F-measure value (calculated as in Eq. 21 with $\beta^2 = 0.5$), and also the highest Precision value, which generally has a higher importance for an efficient segmentation algorithm than Recall [22] (however, the Recall of the proposed approach is still the second best). The lower Recall rate is most probably caused by the aforementioned high intra-object textures in the SED database: the initialization might only cover a part of the object with the most distinctive texture, which in turn might distort the feature map. Thus, the algorithm might in some cases only detect a part of the object, resulting in high Precision, but a lower Recall value.

The performance values show that, the direction feature combined with texture and the Harris based gradient (combining the advantages of all three features) is able to represent the object more accurately, than the Harris based gradient alone (HGVF and HVFC methods), or the texture and gradient (STVFC) features. By comparing the performance results of HVFC and the proposed TDHVFC, the performance improvement is achieved by the feature fusion can be measured, ranging from 0.795 to 0.83. On the other hand, when comparing the performance of STVFC and TDHVFC, the improvement is caused by the novel Harris based direction feature (from 0.76 to 0.83). (Please note that Table II contains rounded values, thus the TDHVFC value differs from its value for $\gamma = 0.3$ in Table I.) The four examples in Fig. 5 show the detection results for different active contour algorithms. The sample images were selected to have objects with complex outlines.

B. Salient object segmentation

In the second part of the evaluations, the proposed method’s capabilities are tested from the point of view of salient object segmentation, by performing comparisons with several state-of-the-art techniques (briefly reviewed in Sec. II): SS [42], STVFC [12], FT [33], GC [45], SEG [40], DSR [43], SDAE [14], CB [41], LMLC [44], RC [22], CPMC+GBVS (marked as CPMC) [48], HDCT [47] and OCT [13].

The different algorithms were tested on two major databases: the **MSRA10K** dataset was introduced in [22]

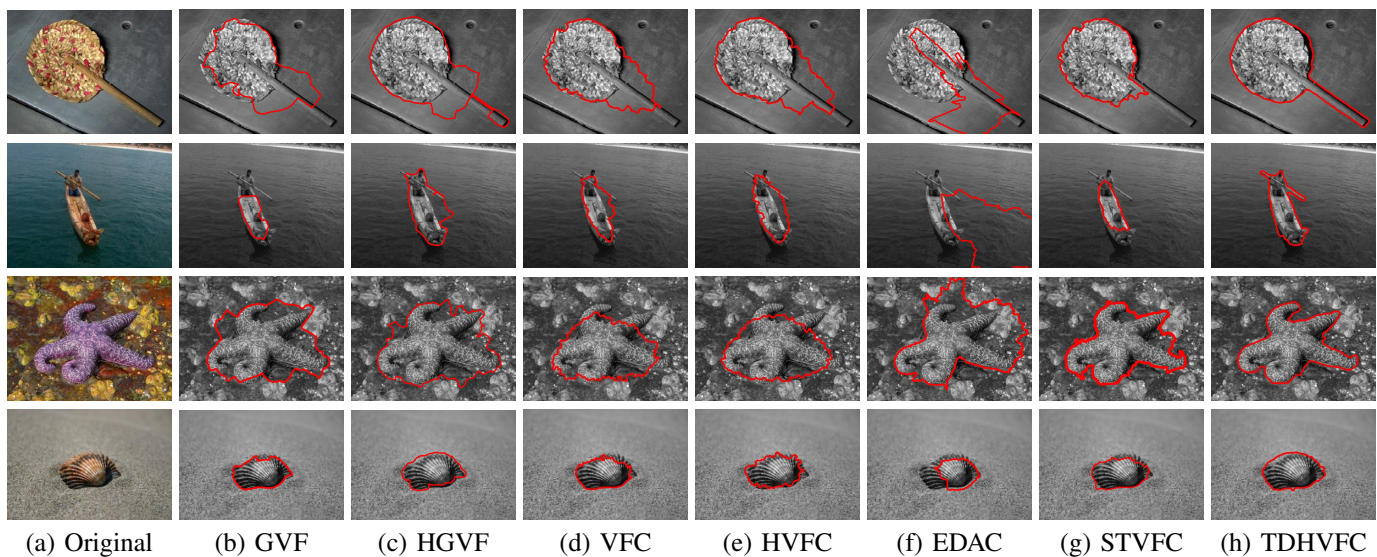


Fig. 5. Segmentation results for sample images of the SED database for different active contour methods: (b) GVF [2], (c) HGVF [5], (d) VFC [4], (e) HVFC [28], (f) STVFC [12], (g) EDAC [18] and (h) the proposed TDHVFC algorithm.

containing 10,000 selected images from the MSRA database [65]. The MSRA10K also contains pixel-level saliency labels, being the largest publicly available database of its kind; the **EPFL** [33] is the second largest such database, containing 1000 images along with their binary segmentation masks.

The compared methods usually produce a grayscale saliency map output. To get a segmented binary result which can be compared to the ground truth mask, we apply SaliencyCut [22] for CB, DRFI, DSR, GC, HDCT, LMLC, RC, SEG and SS methods. CPMC and SDAE approaches have their own binarization technique published in [48] [14] respectively.

As the Precision value is defined to be more representative for segmentation performance evaluation, the F_β measure is calculated for the compared methods for both datasets as follows with $\beta^2 = 0.3$ as in [33]:

$$F_\beta = \frac{(1 + \beta^2)Precision \times Recall}{\beta^2 \times Precision + Recall}. \quad (21)$$

Quantitative segmentation results are shown in Fig. 6 for the EPFL and MSRA10K databases. The results for the SS, FT, SEG, DSR, CB, LMLC, RC methods for both datasets are given in [39], the STVFC segmentation results for the EPFL database is from [12]. The quantitative results of the proposed method are comparable with the most efficient techniques, like the RC and LCMC: besides the relatively high F_β value, the proposed method is able to achieve very high Precision on both datasets. However, its Recall is generally lower, which along with the high Precision value means that the TDHVFC method occasionally detects only portions of the salient objects, a situation that can occur in the case of high variations of intra-object textures.

A third dataset is also evaluated in this section, called the **SUN** dataset, which is based on the MSRA database [65] and includes 346 images with ground truth and it was introduced in [13] for evaluating the OCT method. As this method also uses orientation information - although in a different way -, comparison seemed to be meaningful. For the SUN dataset,

the mean F-measure value is compared for the RC, OCT and the proposed method, the results are shown in Table III.

The method has also been evaluated on the **ECSSD** dataset [66] achieving an average 0.75 F_β score, which is quite good, compared to other state-of-the-art methods based on [39].

Sample images are shown in Fig. 7 together with the results of the FT, SS, GC, SEG, DSR, RC, CB, LMLC, HDCT, CPMC, SDAE, OCT and the proposed TDHVFC methods. The result for the first nine methods are provided by [39], while the results for the SDAE, CPMC and OCT algorithms were provided by the authors of [14], [48], [13].

The first image presents a complex background, where one part (the road) is more similar to the salient object (the occluded cars), than the other part (the grass and trees). The foreground object is also complex and has multiple inner textures (e.g. windshield, headlights). The complex background represents a challenge for some of the algorithms (FT, SS, SEG, CB), and they falsely detect some background regions to be salient. In contrast, other methods (GC, DSR, SDAE, RC, HDCT, OCT) detect only parts of the cars. The LMLC, CPMC and the proposed TDHVFC methods are able to detect the salient object the most accurately.

The second image illustrates the case of a complex foreground object with very different object regions (varying color and pattern), and a cluttered but distinct background region. The complex salient object is only partially detected by some of the segmentation algorithms, and only a few of them (LMLC, HDCT, CPMC, TDHVFC) are able to detect

SUN database	RC	OCT	Proposed
Mean F-measure	0.66	0.71	0.81

TABLE III
MEAN F-MEASURE FOR RC [22], OCT [13] AND THE PROPOSED ALGORITHMS FOR THE SUN DATASET [13].

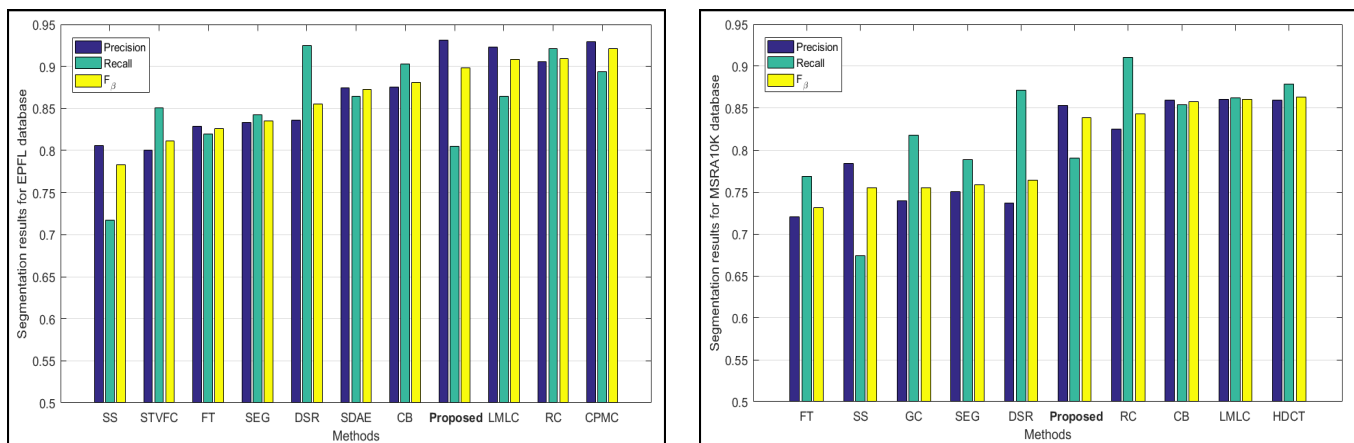


Fig. 6. Region of interest segmentation results for EPFL [33] and MSRA10K [22] datasets.

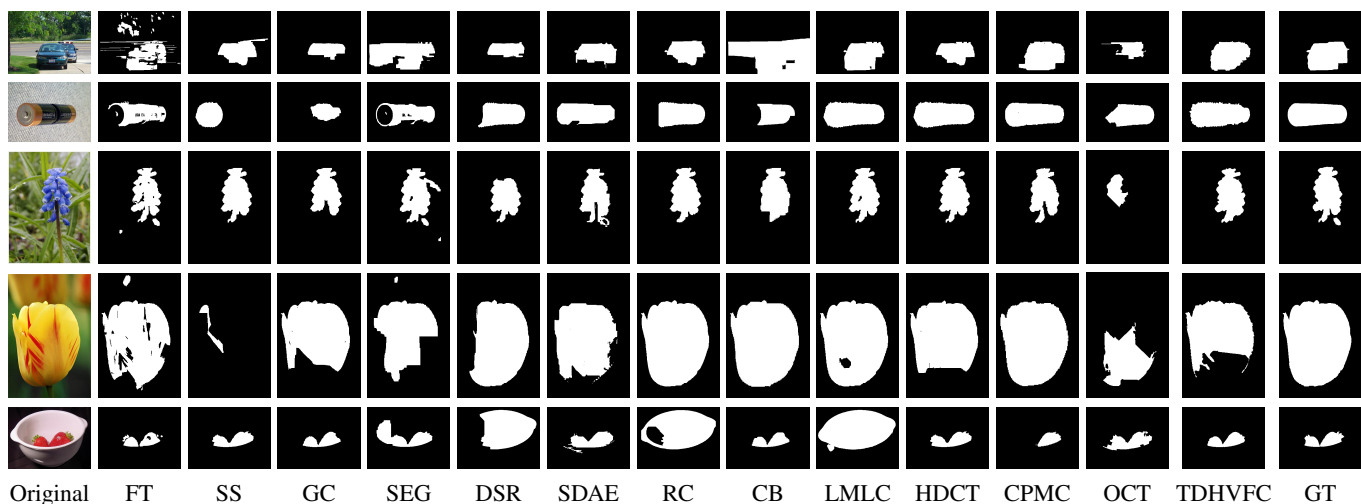


Fig. 7. Segmentation results for selected samples of the MSRA10K and SUN databases for the compared algorithms: FT [33], SS [42], GC [45], SEG [40], DSR [43], SDAE [14], RC [22], CB [41], LMLC [44], HDCT [47], CPMC [48], OCT [13] and the proposed TDHVFC. GT marks the given ground truth.

the whole object. The proper outline is a bit distorted by the background in case of the TDHVFC.

The third image shows a cluttered background with a multi-part object. All the algorithms localize the object well, but the complex shape is only partially detected by most of the methods. It should be mentioned, that this sample shows the drawback of the TDHVFC method: if an object has multiple, disconnected parts, an active contour based method will not treat them as one integral object. Hence, the small separate part of the flower in the lower half of the image is missed. However, an iterative extension can solve this issue, as it will be proposed in Sec. VI-C.

The fourth image shows the disadvantage of the proposed Harris based structural part: on the lower, shady object outline, no MHEC feature points can be found, therefore the initialization contains only a part of the object region, resulting in an inaccurate detection. This sample illustrates again, that initialization problems cannot be fully compensated in later steps. Based on the segmentation results, beside TDHVFC, SS, GC, SEG, SDAE and HDCT methods also have problems with detecting the lower, shady object region part.

The fifth image illustrates the case when the salient object is small and located within a bigger object, which is also in focus. Due to the small size of the salient object, algorithms based on regional statistics might not work. Therefore, some of the methods detect the larger focused object as foreground. The proposed TDHVFC algorithm has the ability to detect even objects of small size in the image.

Overall, based on the above visual and quantitative evaluations, we can state that the proposed method is comparable and competitive with relation to the most efficient salient object segmentation algorithms.

C. Multiple object detection

As the original TDHVFC method is by design only able to handle one object consisting of connected parts, we introduce an extension for detecting multiple objects. By referring back to the beginning of Section III-B, the ROI_S with the maximum area is selected in the initialization. Here, as part of an iterative solution, we propose to repeat the selection process for a defined number of iterations. Thus, not only the largest blob of the binarized S is used for object detection, but also the

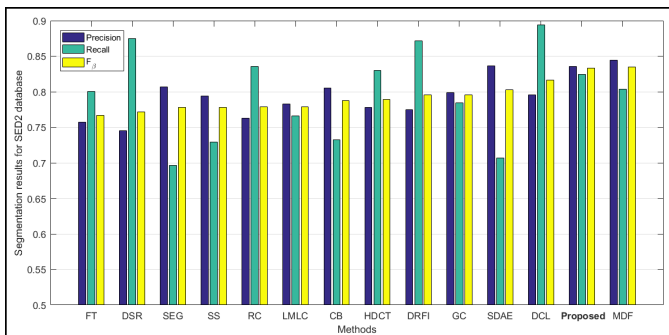


Fig. 8. Multiple object segmentation results for SED2 dataset.

second, third, etc. largest ones. This gives us the opportunity to detect multiple objects in the image. The number of the objects can either be fixed previously, or the iterations can be performed for all the blobs larger than a previously set size.

To test this iterative extension, the **SED2** [32] two objects database was used, which contains 100 images with two separate objects, sometimes only partially visible in front of a background. This dataset, just like the single object one, also includes ground truth, obtained from one or more human subjects. The two foreground objects are varying in size and texture. The TDHVFC detection process is performed in two iterations for detecting the two separate objects. The FT [33], DSR [43], SEG [40], SS [42], RC [22], LMLC [44], CB [41], HDCT [47], DRFI [46], GC [45], SDAE [14], DCL [54] and MDF [53] methods are also evaluated on the SED2 database for comparison. To obtain binary segmentation results, two different segmentation techniques are applied on FT, DSR, SEG, SS, RC, LMLC, CB, HDCT, DRFI and GC approaches: SaliencyCut [22] and a binarization step proposed in [33]. The binarized saliency maps of the different methods were published on [39]. As the evaluation results of the latter binarization technique outperformed the accuracies achieved by the SaliencyCut, the diagram only represents those larger values. The saliency maps of SDAE, DCL and MDF approaches are published on the website of the authors of [14], [54], [53], and the binarization step is performed as it is proposed in the related papers. Fig. 8 shows the segmentation results: the F-measure (Eq. 21 with $\beta^2 = 0.3$) and the corresponding Precision and Recall values for all methods, indicating the superiority of TDHVFC approach over almost all of the compared methods.

To illustrate the processes, Fig. 9 shows visual examples (*cows*, *lamas*, *moon*, *shell*, *wood*) from the SED2 database along with their binary segmentation maps for the compared 14 algorithms. The first *cows* sample shows an image with a salient object quite similar to the background and an other more distinctive foreground object. The bigger part of the tested methods (DSR, SEG, SS, RC, LMLC, HDCT, DRFI, GC, MDF) only emphasizes the more distinctive white animal, while the second is not distinguished from the background. The *lamas* example (second row) shows two similar foreground objects in front of a cluttered background, where some parts of the background is similar to the salient objects. Except the MDF and TDHVFC methods, all other approaches have

some falsely detected background area marked as foreground objects. The *moon* image (third row) shows a challenging case, when the object texture is relatively small compared to the image size and some of the compared algorithms (LMLC, DRFI, GC, MDF) are not able to find the small object, while others, concentrating also on the location of the object (SEG, RC) only find the moon object, but miss the other. When objects are prominent, have a large size, with various intra-object textures and a cluttered background, many algorithms face difficulties in correctly detecting them, including TDHVFC. The next *shell* sample (fourth row) shows the drawback of the proposed TDHVFC method, as it cannot handle the varying intra-object textures efficiently, and only partially segments the left shell. However, it is still able to detect a relevant part of this left object, unlike FT, DSR, SEG, SEG, SS, RC, LMLC, CB, HDCT, DRFI, and GC, which miss the left object almost entirely. Finally, the last image sample *wood* (last row), represents the case, when relevant objects are located on the edge of the image. This is also a very difficult case for almost all of the compared algorithms, most of them only detecting the more distinct, darker object and partially or totally missing the lighter object. Only the LMLC method is able to detect the objects accurately. Due to the location restrictions, TDHVFC gives higher distinctiveness values for the image center, which results in the omission of the outer parts of the objects in the detection step.

D. Computational time

The computational time of the proposed method was compared to other active contour techniques while performing evaluations on the SED database. The computation times are measured on an Intel(R) Core(TM) i7-3770 CPU 3.4GHz with 8 GB RAM in Matlab. The source code for texture distinctiveness map calculation was provided by the authors of [12]. The execution times are shown in Table IV. The common initialization step, and the iterative segmentation steps were measured separately. The number of the maximum iterative segmentation steps was set to 60 for VFC, HVFC and TDHVFC. If convergence was reached earlier, the method was stopped. The computational time for EDAC and STVFC methods were provided in [18] and [12] respectively.

The initialization part consists of the calculation of the S texture distinctiveness map (Sec. III-A) performed in an average of 1.43 s time and the contour initialization part, calculated in 0.45 s. As in [5] it was already shown that GVF and HGVF methods are slower than VFC and HVFC, these methods are not included in Table IV. The iterative segmentation is the fastest with the proposed TDHVFC model, which means that the contour initialization combined with the improved edge map results in a method with faster convergence.

CT	Init.	VFC	HVFC	EDAC	STVFC	TDHVFC
[s]	1.88	15.59	13.71	49.96	6.55	6.32

TABLE IV
AVERAGE COMPUTATIONAL TIME (CT) OF THE SED DATABASE IMAGES FOR DIFFERENT METHODS: VFC [4], HVFC [28], EDAC [18], STVFC [12] AND THE PROPOSED TDHVFC.

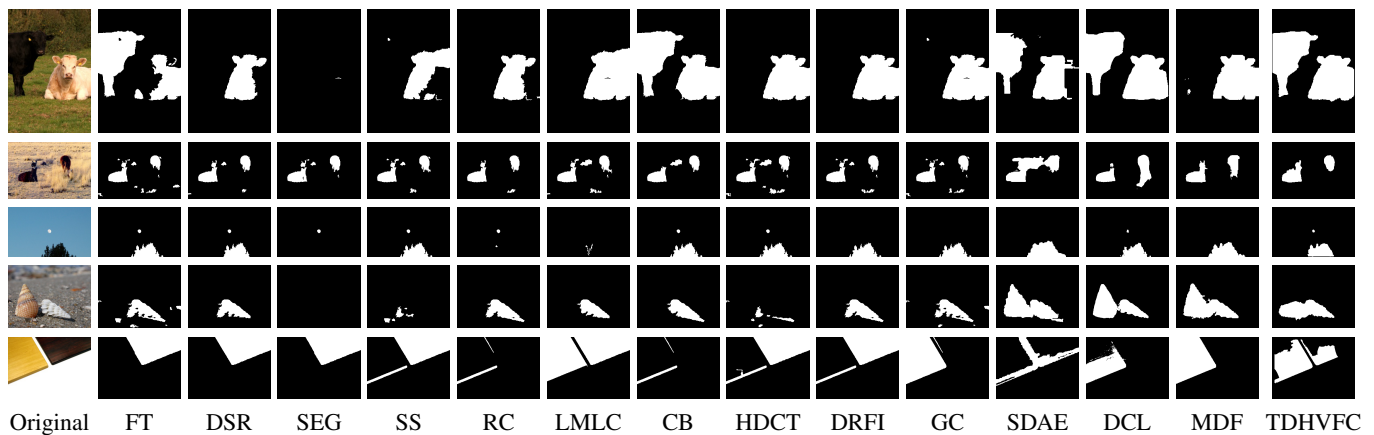


Fig. 9. Segmentation results for sample images of the SED2 database. Images from top to bottom: cows, lamas, moon, shell, wood.

VII. CONCLUSION

This paper presented an improved edge map incorporating direction information, which can be adapted for parametric active contour object detection, with two main steps. First, automatic contour initialization is presented by integrating structural edge information with texture distinctiveness. Second, an improved external force field is presented fusing direction, edge and texture data. The exploitation of direction information creates the opportunity for the model to handle cluttered backgrounds more efficiently. The main idea of the proposed method is to represent the object as the unity of its internal texture and its direction-augmented contour, integrating both into a feature map. Extensive evaluations support that the proposed method performs with higher accuracy than existing parametric active contour models, and that it is more efficient than the majority of the evaluated saliency methods.

The convergence of the presented method is faster than the compared methods', however, iterative active contour evolution is still slower than other saliency-based methods. Therefore, future work will investigate the possibilities for unsupervised segmentation (e.g., with SaliencyCut) of the improved feature map. Nevertheless, contour detection can be an alone refinement step after ROI estimation performed by other state-of-the-art techniques for more accurate detection. Another advantage of the proposed method is its ability for multiple object detection, although the number of objects has to be pre-set. Thus, in the future, we intend to provide a means to estimate the number of salient objects. Finally, as the fusion of direction, edge and texture produces an efficient feature map, interest points extracted from such a map might be useful for content-based retrieval purposes, which is intended to be investigated in the future.

ACKNOWLEDGMENT

The author would like to thank Ming-Ming Cheng for his help and the publicly available extensive evaluation results; the anonymous reviewers for their thorough work. This work was funded by the Hungarian National Research, Development and Innovation Fund (NKFIH), grant nr. NKFIH-KH 126688 and the Hungarian Government, Ministry for National Economy (NGM), grant nr. GINOP-2.2.1-15-2017-00083.

REFERENCES

- [1] M. Kass, A. P. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *Int. J. of Computer Vision*, vol. 1, no. 4, pp. 321–331, 1988.
- [2] C. Xu and J. L. Prince, "Gradient vector flow: A new external force for snakes," in *Proc. of IEEE Conf. on Comp. Vis. and Patt. Rec. (CVPR)*, 1997, pp. 66–71.
- [3] J. Cheng and S. W. Foo, "Dynamic directional gradient vector flow for snakes," *IEEE Tr. on Image Processing*, vol. 15, no. 6, pp. 1563–1571, 2006.
- [4] B. Li and T. Acton, "Active contour external force using vector field convolution for image segmentation," *IEEE Tr. on Image Processing*, vol. 16, no. 8, pp. 2096–2106, 2007.
- [5] A. Kovacs and T. Sziranyi, "Harris function based active contour external force for image segmentation," *Pattern Recognition Letters*, vol. 33, no. 9, pp. 1180–1187, 2012.
- [6] V. Caselles, R. Kimmel, and G. Sapiro, "Geodesic active contours," *Int. J. of Computer Vision*, vol. 22, no. 1, pp. 61–79, 1997.
- [7] T. F. Chan and L. A. Vese, "Active contours without edges," *IEEE Tr. on Image Processing*, vol. 10, no. 2, pp. 266–277, 2001.
- [8] J. Melonakos, E. Pichon, S. Angenent, and A. Tannenbaum, "Finsler active contours," *IEEE Tr. on Pattern Analysis and Machine Intelligence*, vol. 30, no. 3, pp. 412–423, 2008.
- [9] J. Yang, B. Price, S. Cohen, H. Lee, and M.-H. Yang, "Object contour detection with a fully convolutional encoder-decoder network," in *Proc. of the IEEE Conf. on Comp. Vis. and Patt. Rec. (CVPR)*, 2016, pp. 193–202.
- [10] S. Osher and J. A. Sethian, "Fronts propagating with curvature-dependent speed: algorithms based on hamilton-jacobi formulations," *J. of computational physics*, vol. 79, no. 1, pp. 12–49, 1988.
- [11] D. Mumford and J. Shah, "Optimal approximations by piecewise smooth functions and associated variational problems," *Communications on pure and applied mathematics*, vol. 42, no. 5, pp. 577–685, 1989.
- [12] K. Fergani, D. Lui, C. Scharfenberger, A. Wong, and D. Clausi, "Hybrid structural and texture distinctiveness vector field convolution for region segmentation," *Computer Vision and Image Understanding*, vol. 125, pp. 85–96, 2014.
- [13] L. Sun and T. Shibata, "Unsupervised object extraction by contour delineation and texture discrimination based on oriented edge features," *IEEE Tr. on Circuits and Systems for Video Technology*, vol. 24, no. 5, pp. 780–788, 2014.
- [14] J. Han, D. Zhang, X. Hu, L. Guo, J. Ren, and F. Wu, "Background prior-based salient object detection via deep reconstruction residual," *IEEE Tr. on Circuits and Systems for Video Technology*, vol. 25, no. 8, pp. 1309–1321, 2015.
- [15] A. Borji, "What is a salient object? A dataset and a baseline model for salient object detection," *IEEE Tr. on Image Processing*, vol. 24, no. 2, pp. 742–756, 2015.
- [16] J. Yang, G. Zhao, J. Yuan, X. Shen, Z. Lin, B. Price, and J. Brandt, "Discovering primary objects in videos by saliency fusion and iterative appearance estimation," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, no. 6, pp. 1070–1083, June 2016.

- [17] A. K. Mishra, P. W. Fieguth, D. Clausi *et al.*, “Decoupled active contour (dac) for boundary detection,” *IEEE Tr. on Pattern Analysis and Machine Intelligence*, vol. 33, no. 2, pp. 310–324, 2011.
- [18] D. Lui, C. Scharfenberger, K. Fergani, A. Wong, D. Clausi *et al.*, “Enhanced decoupled active contour using structural and textural variation energy functionals,” *IEEE Tr. on Image Processing*, vol. 23, no. 2, pp. 855–869, 2014.
- [19] G. Sundaramoorthi and A. Yezzi, “Global regularizing flows with topology preservation for active contours and polygons,” *IEEE Tr. on Image Processing*, vol. 16, no. 3, pp. 803–812, 2007.
- [20] C. Tauber, H. Batatia, and A. Ayache, “Quasi-automatic initialization for parametric active contours,” *Pattern Recognition Letters*, vol. 31, pp. 83–90, 2010.
- [21] L. Kovacs and T. Sziranyi, “Focus area extraction by blind deconvolution for defining regions of interest,” *IEEE Tr. on Pattern Analysis and Machine Intelligence*, vol. 29, no. 6, pp. 1080–1085, 2007.
- [22] M. Cheng, N. J. Mitra, X. Huang, P. H. Torr, and S. Hu, “Global contrast based salient region detection,” *IEEE Tr. on Pattern Analysis and Machine Intelligence*, vol. 37, no. 3, pp. 569–582, 2015.
- [23] C. Scharfenberger, A. Wong, K. Fergani, J. S. Zelek, D. Clausi *et al.*, “Statistical textural distinctiveness for salient region detection in natural images,” in *2013 Proc. of IEEE Conf. on Comp. Vis. and Patt. Rec. (CVPR)*, 2013, pp. 979–986.
- [24] J. Han, S. He, X. Qian, D. Wang, L. Guo, and T. Liu, “An object-oriented visual saliency detection framework based on sparse coding representations,” *IEEE Tr. on Circuits and Systems for Video Technology*, vol. 23, no. 12, pp. 2009–2021, 2013.
- [25] C. Harris and M. Stephens, “A combined corner and edge detector,” in *Proc. of the 4th Alvey Vision Conf.*, 1988, pp. 147–151.
- [26] D. G. Lowe, “Distinctive image features from scale-invariant keypoints,” *Int. J. of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.
- [27] A. Kovacs and T. Sziranyi, “Improved Harris feature point set for orientation sensitive urban area detection in aerial images,” *IEEE Geoscience and Remote Sensing Letters*, vol. 10, no. 4, pp. 796–800, 2013.
- [28] A. Kovács and T. Szirányi, “Improved force field for vector field convolution method,” in *IEEE Int. Conf. on Image Processing (ICIP)*, sept. 2011, pp. 2853–2856.
- [29] N. Jifeng, W. Chengke, L. Shigang, and Y. Shuqin, “NGVF: An improved external force field for active contour model,” *Pattern Recognition Letters*, vol. 28, no. 1, pp. 58–63, 2007.
- [30] G. Zhu, S. Zhang, Q. Zeng, and C. Wang, “Gradient vector flow active contours with prior directional information,” *Pattern Recognition Letters*, vol. 31, no. 9, pp. 845–856, 2010.
- [31] I. Zingman, D. Saupe, and K. Lambers, “A morphological approach for distinguishing texture and individual features in images,” *Pattern Recognition Letters*, vol. 47, pp. 129–138, 2014.
- [32] S. Alpert, M. Galun, R. Basri, and A. Brandt, “Image segmentation by probabilistic bottom-up aggregation and cue integration,” in *Proc. of the IEEE Conf. on Comp. Vis. and Patt. Rec. (CVPR)*, 2007, pp. 1–8.
- [33] R. Achanta, S. Hemami, F. Estrada, and S. Ssstrunk, “Frequency-tuned Salient Region Detection,” in *Proc. of the IEEE Conf. on Comp. Vis. and Patt. Rec. (CVPR)*, 2009, pp. 1597–1604.
- [34] A. Manno-Kovacs, “Direction selective vector field convolution for contour detection,” in *IEEE Int. Conf. on Image Processing (ICIP)*, 2014, pp. 4722–4726, “Top10%”.
- [35] W. Wang, J. Shen, and L. Shao, “Consistent video saliency using local gradient flow optimization and global refinement,” *IEEE Tr. on Image Processing*, vol. 24, no. 11, pp. 4185–4196, 2015.
- [36] Y. Zhang, X. Qian, X. Tan, J. Han, and Y. Tang, “Sketch-based image retrieval by salient contour reinforcement,” *IEEE Tr. on Multimedia*, vol. 18, no. 8, pp. 1604–1615, 2016.
- [37] F. Zhang, B. Du, and L. Zhang, “Saliency-guided unsupervised feature learning for scene classification,” *IEEE Tr. on Geoscience and Remote Sensing*, vol. 53, no. 4, pp. 2175–2184, 2015.
- [38] A. Borji, M.-M. Cheng, H. Jiang, and J. Li, “Salient object detection: A benchmark,” *IEEE Tr. on Image Processing*, vol. 24, no. 12, pp. 5706–5722, 2015.
- [39] M.-M. Cheng, 2016, <http://mmcheng.net/salobjbenchmark/>.
- [40] E. Rahtu, J. Kannala, M. Salo, and J. Heikkilä, “Segmenting salient objects from images and videos,” in *European Conf. on Computer Vision (ECCV)*. Springer, 2010, pp. 366–379.
- [41] H. Jiang, J. Wang, Z. Yuan, T. Liu, N. Zheng, and S. Li, “Automatic salient object segmentation based on context and shape prior,” in *British Machine Vision Conf.*, vol. 6, no. 7, 2011, p. 9.
- [42] X. Hou, J. Harel, and C. Koch, “Image signature: Highlighting sparse salient regions,” *IEEE Tr. on Pattern Analysis and Machine Intelligence*, vol. 34, no. 1, pp. 194–201, 2012.
- [43] X. Li, H. Lu, L. Zhang, X. Ruan, and M.-H. Yang, “Saliency detection via dense and sparse reconstruction,” in *IEEE Int. Conf. on Computer Vision (ICCV)*. IEEE, 2013, pp. 2976–2983.
- [44] Y. Xie, H. Lu, and M.-H. Yang, “Bayesian saliency via low and mid level cues,” *IEEE Tr. on Image Processing*, vol. 22, no. 5, pp. 1689–1698, 2013.
- [45] M.-M. Cheng, J. Warrell, W.-Y. Lin, S. Zheng, V. Vineet, and N. Crook, “Efficient salient region detection with soft image abstraction,” in *IEEE Int. Conf. on Computer Vision (ICCV)*, 2013, pp. 1529–1536.
- [46] H. Jiang, J. Wang, Z. Yuan, Y. Wu, N. Zheng, and S. Li, “Salient object detection: A discriminative regional feature integration approach,” in *Proc. of the IEEE Conf. on Comp. Vis. and Patt. Rec. (CVPR)*, 2013, pp. 2083–2090.
- [47] J. Kim, D. Han, Y. W. Tai, and J. Kim, “Salient region detection via high-dimensional color transform and local spatial support,” *IEEE Tr. on Image Processing*, vol. 25, no. 1, pp. 9–23, 2016.
- [48] Y. Li, X. Hou, C. Koch, J. M. Rehg, and A. L. Yuille, “The secrets of salient object segmentation,” in *Proc. of the IEEE Conf. on Comp. Vis. and Patt. Rec. (CVPR)*, 2014, pp. 280–287.
- [49] K. F. Yang, H. Li, C. Y. Li, and Y. J. Li, “A unified framework for salient structure detection by contour-guided visual search,” *IEEE Tr. on Image Processing*, vol. 25, no. 8, pp. 3475–3488, 2016.
- [50] J. Han, D. Zhang, S. Wen, L. Guo, T. Liu, and X. Li, “Two-stage learning to predict human eye fixations via SDAEs,” *IEEE Tr. on Cybernetics*, vol. 46, no. 2, pp. 487–498, 2016.
- [51] D. Zhang, J. Han, J. Han, and L. Shao, “Cosaliency detection based on intrasaliency prior transfer and deep intersaliency mining,” *IEEE Tr. on Neural Networks and Learning Systems*, vol. 27, no. 6, pp. 1163–1176, 2016.
- [52] W. Shen, X. Wang, Y. Wang, X. Bai, and Z. Zhang, “DeepContour: A deep convolutional feature learned by positive-sharing loss for contour detection,” in *Proc. of the IEEE Conf. on Comp. Vis. and Patt. Rec. (CVPR)*, 2015, pp. 3982–3991.
- [53] G. Li and Y. Yu, “Visual saliency based on multiscale deep features,” in *Proc. of the IEEE Conf. on Comp. Vis. and Patt. Rec. (CVPR)*, 2015, pp. 5455–5463.
- [54] —, “Deep contrast learning for salient object detection,” in *Proc. of the IEEE Conf. on Comp. Vis. and Patt. Rec. (CVPR)*, 2016, pp. 478–487.
- [55] N. Otsu, “A threshold selection method from gray-level histograms,” *IEEE Tr. on Systems, Man and Cybernetics*, vol. 9, no. 1, pp. 62–66, 1979.
- [56] C. B. Barber, D. P. Dobkin, and H. Huhdanpaa, “The quickhull algorithm for convex hulls,” *ACM Trans. Math. Softw.*, vol. 22, no. 4, pp. 469–483, 1996.
- [57] Y. Lu, W. Zhang, H. Lu, and X. Xue, “Salient object detection using concavity context,” in *IEEE Int. Conf. on Computer Vision (ICCV)*. IEEE, 2011, pp. 233–240.
- [58] C. Yang, L. Zhang, and H. Lu, “Graph-regularized saliency detection with convex-hull-based center prior,” *IEEE Signal Processing Letters*, vol. 20, no. 7, pp. 637–640, 2013.
- [59] A. Manno-Kovacs and T. Sziranyi, “Orientation-selective building detection in aerial images,” *ISPRS J. of Photogrammetry and Remote Sensing*, vol. 108, pp. 94–112, 2015.
- [60] S. Kumar and M. Hebert, “Man-made structure detection in natural images using a causal multiscale random field,” in *Proc. of the IEEE Conf. on Comp. Vis. and Patt. Rec. (CVPR)*, 2003, pp. 119–126.
- [61] C. Benedek, X. Descombes, and J. Zerubia, “Building development monitoring in multitemporal remotely sensed image pairs with stochastic birth-death dynamics,” *IEEE Tr. on Pattern Analysis and Machine Intelligence*, vol. 34, no. 1, pp. 33–50, 2012.
- [62] J. Canny, “A computational approach to edge detection,” *IEEE Tr. on Pattern Analysis and Machine Intelligence*, vol. 8, no. 6, pp. 679–698, 1986.
- [63] S. Yi, D. Labate, G. R. Easley, and H. Krim, “A shearlet approach to edge analysis and detection,” *IEEE Tr. on Image Processing*, vol. 18, no. 5, pp. 929–941, 2009.
- [64] R. Mester, “Orientation estimation: Conventional techniques and a new non-differential approach,” in *Proc. 10th European Signal Processing Conf.*, 2000, pp. 921–924.
- [65] T. Liu, Z. Yuan, J. Sun, J. Wang, N. Zheng, X. Tang, and H.-Y. Shum, “Learning to detect a salient object,” *IEEE Tr. on Pattern Analysis and Machine Intelligence*, vol. 33, no. 2, pp. 353–367, 2011.
- [66] Q. Yan, L. Xu, J. Shi, and J. Jia, “Hierarchical saliency detection,” in *Proc. of the IEEE Conf. on Comp. Vis. and Patt. Rec. (CVPR)*, 2013, pp. 1155–1162.



Andrea Manno-Kovacs received her PhD in image processing in 2013 from the Pazmany Peter Catholic University in Budapest, her MSc in computer science from Budapest University of Technology and Economics. She is currently a research fellow with the Machine Perception Research Laboratory, at the Institute for Computer Science and Control of the Hungarian Academy of Sciences (MTA SZTAKI). She has been the manager of various national and international research projects in the recent years. Beside, she is also supervising BSc and MSc stu-

dents at Pazmany Peter Catholic University. Her main interests include image and video processing, feature extraction, saliency models, boundary recognition and active contours.