Chapter 10
Modelling Sparse Saliency Maps on Manifolds Numerical Results and Applications

Eduardo Alcaín, Ana Isabel Muñoz, Iván Ramírez and Emanuele Schiavi

Abstract Saliency detection is an image processing task which aims at automatically estimating visually salient object regions in a digital image mimicking human visual attention and eyes fixation. A number of different computational approaches for visual saliency estimation has recently appeared in Computer and Artificial Vision. Relevant and new applications can be found everywhere varying from automatic image segmentation and understanding, localization and quantification for biomedical and aerial images to fast video tracking and surveillance. In this contribution, we present a new variational model on finite dimensional manifolds generated by some characteristic features of the data. A Primal-Dual method is implemented for the numerical resolution showing promising preliminary results.

10.1 Introduction

In the last years, there has been a growing interest in the production of computational methods for the detection of saliency objects in an image. A general overview of existing methods, drawbacks and virtues can be found in [5]. Salient (or foreground) objects are those objects which grasp the most interest when an image is considered. This estimation is normally used as a preprocessing step in a pipeline for a computer vision system. The concept of saliency has been applied to adaptive compression of images [7], image retrieval [2] and image cropping [14] to name just a few emerging applications. Saliency methods can be classified into three categories: biological based, purely computational and a mix from both methods.

E. Alcaín · A.I. Muñoz · I. Ramírez · E. Schiavi.
Departamento de Matemática Aplicada, Ciencia e Ingeniería de Materiales y Tecnología Electrónica, Universidad Rey Juan Carlos, ESCET, Móstoles, 28933-Madrid, Spain
e-mail: e.alcain@alumnos.urjc.es, anaisabel.munoz@urjc.es, i.ramirez@alumnos.urjc.es, emanuele.schiavi@urjc.es

Although variational methods have achieved great success when applied in computer vision (denoising, deblurring, inpainting, etc.), there have been few saliency detection models which make use of the power of the variational setting. This introduces the main contribution of the paper which consists of a new variational model for automatic saliency detection in digital images based on Total Variation (TV) minimization on graphs.

We consider as a starting point of our modelling exercise the work in [15] where a non local L_0 minimization problem is proposed for bottom-up pure computationally saliency estimation. Based on the discrete L_0 norm on graphs, this model is appropriate to capture the sparse properties of saliency maps which are vectors, finite dimensional solutions which estimate the saliency of each superpixel in which the image has been previously partitioned. In fact, following the idea in [15] and with a view to fast (real time) video saliency detection, the data image is divided into superpixels for dimensional reduction and a non local graph is constructed in a feature space in which each vertex is a superpixel connected to its k-NN (k-nearest neighbour).

There has been a lot of effort in these years to develop the formalism of Non-local Calculus ([9]) and Non-local Operators so providing a new tool for the mathematical analysis of problems on graphs. This has generated a new kind of models where the influence of any pixel value is extended to all (or part) of the domain through the concept of Non-local derivatives. When a graph structure is generated, by considering some characteristics of a given data image, spatial proximity is lost or attenuated and Non-local operators are necessary to describe the manifold landscape.

In such a framework, our proposal is two-fold. We first generalize the superpixels partition considered in [15] including edges as features in the graph by means of the use of the SCALP algorithm ([12]) where contour adherence is imposed using linear paths. Also, instead of considering a fixed threshold, the final saliency mask is obtained using the method proposed in [13]. Secondly, and more important, the L_0-norm of the non local gradients in the saliency model presented [15] is replaced by the NLTV (non local total variation) semi-norm on graphs defined by the non local total variation operator, which preserves edges and induces the sparsity of the gradients of saliency maps. This step amounts to a convexification of the problem which allows classical variational calculus to be applied. Finally, as in [15], a control map is computed in the graph to drive the solution towards salient regions. With all these ingredients a strictly convex energy functional is considered for minimization and the well-posedness of the model is guaranteed. The numerical resolution is based on a primal-dual algorithm that we designed to solve the associated minimization NLTV problem. The proposed algorithm proves to be faster in convergence to the solution than the one presented in [15] for the non local L_0 minimization problem, while obtaining results at least comparable to the ones in [15].

The paper is organized as follows: First, in section 10.2 we introduce the concept of superpixel (SP) as a guided (informed) partition of the given image through the chosen characteristics. This is a critical step because the produced superpixels partitions (in number and characteristics) shall not be modified along the whole pipeline. In section 10.3 the finite dimensional manifold (weighted graph) generated by the partition of the image into superpixels by using some characteristic features of the data is constructed. Sections 10.4 and 10.5 are devoted to the mathematical definitions of the control map and non-local operators used in the model. The main contribution is presented in section 10.6 where the proposed Non-Local Total Variation Model (NLTVM) and its numerical resolution in terms of a primal-dual algorithm are described. The final saliency map segmentation step is considered in section 10.7. Numerical results and discussion are presented in sections 10.8 and 10.9 in order to illustrate the performance of the proposed model.

10.2 Superpixel segmentation of the image

This section is based on the ideas and methods published in [12,15]. In order to compute the graph structure, we first segment the given image into superpixels which are clusters of pixels partitioning the image following some relevant characteristics of the data. This will reduce the computation time of the saliency map providing a dimensional reduction of the problem and fast implementation. Notice also that the relevant information originally located into pixels is then encoded through the shape and values of the superpixel. The method used in [15] to generate superpixels is the SLIC method (Simple Linear Iterative Method [1]) which has been proven to be very efficient in the sense that it is fast, easy to use and it produces high quality partitions of the image. This method performs a local clustering of pixels taking into account the location and the values of the CIELAB colour space of the pixels of an image. The employed measure makes the superpixels shape to be compact and uniform. As an alternative to SLIC in [15] and in order to impose edges conservation we consider the SCALP method (superpixels with contour adherence using linear path [12]) which takes into consideration a contour prior with the aim of obtaining superpixels constrained by the existing contours, being this aspect not considered with SLIC (see the figure 10.1 for comparison). In the SCALP algorithm, when trying to associate a pixel to a superpixel during clustering, the distance is enhanced by considering the linear path to the superpixel barycenter and a contour prior. The prior contour can be calculated by some recent learning based approaches to detect edges like [8] or more classical variational approaches enhanced by implicit finite differences [3]. In this work we used the approach in [8] suggested in [12]). Different strategies shall be explored in future work.

10.3 Weighted graph

Given an image and the superpixels partition generated by SLIC or SCALP method, we consider, as in [15], an undirected, symmetric and weighted graph G in the space of the superpixels. The weighted graph $G = (V, E, w)$ consists of a finite set of superpixels V, the edges E linking superpixels and their associated weights w_{pq}, $pq \in E$. The weights are given by:

$$w_{pq} = exp\left(-\frac{||\mathbf{f}_p - \mathbf{f}_q||^2}{2\sigma^2}\right),$$

where $\mathbf{f}_p$ is a feature vector at superpixel p defined by $\mathbf{f}_p = (\alpha \mathbf{c}_p, \mathbf{l}_p)$, being $\mathbf{c}_p$ the mean of the superpixels in CIELAB colour space, and $\mathbf{l}_p$, the mean of the coordinates in the spatial space. In this case, $\alpha = 0.9$ is a parameter controlling the balance between the two features.

In order to reduce computational cost and to exploit local relationships in the feature space, once we have computed the weight for each superpixel p with every other superpixel q, we consider its k-nearest neighbours the first k vertices in a decreasing list with respect to weight) and keep these associated weights while setting to zero the remaining ones. Then, a superpixel p is associated with a superpixel q if the weight w_{pq} is not zero. To go further in the suppression of the background, the boundaries are connected together as well those superpixels such that their initial saliency map has a value (score) below a user fixed threshold in the weight matrix. In our case the 25% of the less representative values in the control map are neglected (see the figure 10.2).

10.4 Control map

We still follow the ideas in [15] to compute a saliency control map which we will include in the fidelity term of the variational model here presented. The control map

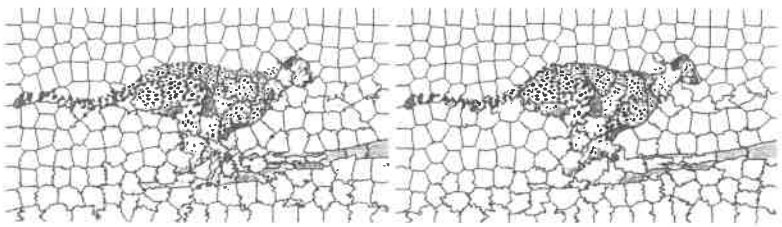

Fig. 10.1 Comparison between SLIC (left) vs SCALP (right) algorithm with 300 superpixels. The edges and contours of the cheetah are better preserved in the partition provided by SCALP.

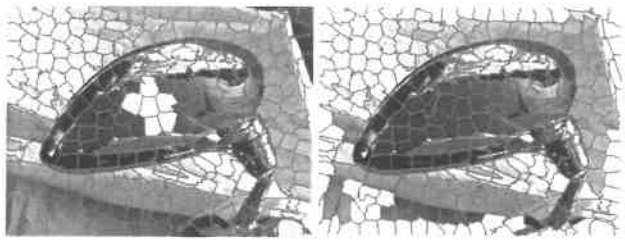

Fig. 10.2 The image was divided by SLIC. On the left, we show a superpixel (yellow) and its k-neighbours. On the right, it is shown a superpixel within less than 25 % in the initial control map is associated to the boundaries as well to its k-neighbours.

v^c is a vector of components $v_c = \{v_p^c,\, p \in G\}$, where v_p^c is the value of the salient map in the superpixel p. This value, v_p^c, is computed as the product of a contrast prior v_p^{con}, which takes into consideration similarity of colours weighted by the distance to each superpixel and an object prior v_p^{obj}, which introduces the assumption of the most likely location of the salient object. Hence, the saliency control map is computed:

$$v_p^c = v_p^{con} \cdot v_p^{obj},$$

where

$$v_p^{con} = \sum_{q,pq \in E} w_{pq}^{(l)} ||c_p - c_q||^2, \quad w_{pq}^{(l)} = exp\left(-\frac{1}{2\sigma^2}||l_p - l_q||^2\right)$$

and

$$v_p^{obj} = exp\left(-\frac{||l_p - \bar{l}||^2}{2\sigma^2}\right).$$

where $\bar{l}$ are the coordinates of the center of the image (where the region of interest is usually located). The parameter σ is empirically set as in [15] where $\sigma^2 = 0.05$.

10.5 Non local operators

In order to deal with variational models in a weighted graph, we need to introduce the notion of non local gradient and divergence operators in a weighted graph G. Details can be found in [9].

Definition 1. Let $p \in G$, $v = (v_p)_{p \in G}$ be a real function defined on G and taking values in $\mathbb{R}$, and $w_{p,q}$, for p and $q \in G$, a nonnegative symmetric weight function. Then, the nonlocal gradient at a superpixel p, $\nabla_w v_p$, is defined as the vector of all partial differences $\nabla_w v_{p,q}$ at p:

$$\nabla_w v_p = \{\nabla_w v_{p,q}, pq \in E\},$$

where

$$\nabla_w v_{p,q} = \sqrt{w_{p,q}}(v_q - v_p).$$

In fact, we are only interested in the components such that $w_{p,q} \neq 0$. Hence, the nonlocal gradient $\nabla_w v$ is defined as $\nabla_w v = (\nabla_w v_p)_{p \in G}$. Analogously, the divergence div_w of a vector d_p, given a function $d : G \times G \to \mathbb{R}$ can be defined as:

$$div_w d_p = \sum_{q, qp \in E} (d_{p,q} - d_{q,p})\sqrt{w_{p,q}}.$$

Notice that, in our particular case, d_p will be taken as $\nabla_w v_p$.

10.5.1 Non local L_0 for saliency detection

In this section, we shall briefly present the model and the method of numerical resolution presented in [15]. This will allow the reader to notice where our contribution resides.

In [15] it is proposed the following nonlocal discrete L_0 minimization model to describe the sparse structure of the saliency maps:

$$\min_v \left(\sum_p ||\nabla_w v_p||_0 + \frac{\lambda}{2}||v - v^c||^2 \right), \tag{10.1}$$

where v_p stands for the value of the saliency map v at the superpixel p, $\nabla_w v_p = \{\nabla_w v_{p,q} : pq \in E\}$, as it was defined in the previous section, and the L_0 norm is given by $||\nabla_w v_p||_0 = \#\{q : \nabla_w v_{p,q} \neq 0, pq \in E\}$ (# is the cardinality of the set). The positive constant λ is a parameter controlling the relative importance given to the fidelity term with respect to the sparsity inducing prior.

In order to solve the minimization problem, the authors in [15] use an Alternated Directions Method (ADM). Setting $d_p = \nabla_w v_p$ the equivalent minimization problem

$$\min_{v,d} \left(\sum_p ||d_p||_0 + \frac{\lambda}{2}(v_p - v_p^c)^2 \right).$$

is solved through the iterative scheme

$$v^{k+1} = \min_v \left(\sum_p \frac{\lambda}{2}(v_p - v_p^c)^2 + \frac{\rho}{2}||d_p^k - \nabla_w v_p + \frac{1}{\rho} y_p^k||^2 \right)$$

$$d^{k+1} = \min_d \left(\sum_p ||d_p||_0 + \frac{\rho}{2}||d_p - \nabla_w v_p^{k+1} + \frac{1}{\rho} y_p^k||^2 \right)$$

where the relaxation variable is given by $y_p^{k+1} = y_p^k + \rho(d_p^k - \nabla_w v_p^{k+1})$. The solution of the problem for v_p^{k+1} can be written as

$$(\lambda + 2\rho \sum_q w_{pq}) v_p^{k+1} = \lambda v_p^c$$

$$-\rho \sum_q \sqrt{w_{pq}} (d_{pq}^k - d_{qp}^k + \frac{1}{\rho}(y_{pq}^k - y_{qp}^k)) + 2\rho \sum_q w_{pq} v_q^{k+1}.$$

Finally, the auxiliary problem for d^{k+1} has a component-wise close solution is given by:

$$d_{pq}^{k+1} = \begin{cases} 0, & |\nabla_w v_{p,q}^{k+1} - \frac{1}{\rho} y_{pq}^k| \leq \sqrt{2}\rho, \\ \nabla_w v_{p,q}^{k+1} - \frac{1}{\rho} y_{pq}^k, & \text{otherwise.} \end{cases}$$

In [15], the iterative procedure is stopped when a fixed number of iterations is reached. This is not a convergence criteria and obscures the behaviour of the algorithm. In order to analyze and compare the results with the ones obtained with our NLTV model, we consider in both cases a typical convergence criteria: the computation is stopped when the difference between two consecutive iterations is less than a small value ε. In our computations we have considered $\varepsilon = 10^{-5}$. We refer the reader to [15] for more details. Notice that this convergence criteria do not reveal if the limit solution is a minimum of the original L_0 model problem (10.1). It just shows that the algorithm stabilizes and converges to a saliency vector which may not be a minimum of functional (10.1). This is due to the fact that, contrary to our model, the energy in (10.1) is not convex. In fact, a careful analysis of the behaviour of the energy functional reveals the existence of two different cases which we show in figures [10.3] and [10.4] where the total energy, the L_0 energy and the fidelity L_2 discrepancy norm are represented. We shall refer to case 1 when the total energy stabilizes to a value but do not converge to a minimum. The L_0 norm is constant and the energy is dominated by the fidelity term. Case 2 refers to convergence to a minimum. An oscillatory form is presented (see figure 10.4) and the energy is dominated by the L_0 norm.

10.6 Non local total variation model

As an alternative to the nonlocal L_0 norm, we propose the following saliency detection model based on the consideration of a non local version of the total variation suitable for weighted graphs, taken into consideration the well known properties of the total variation operator.

In order to do that, we shall first present the notion of NLTV norm of the weighted gradient $\nabla_w v$ introduced in definition 1 (see [10]):

Definition 2. The non local total variation norm in its discrete version can be defined as the following isotropic L_1 norm of the weighted graph gradient $\nabla_w v$:

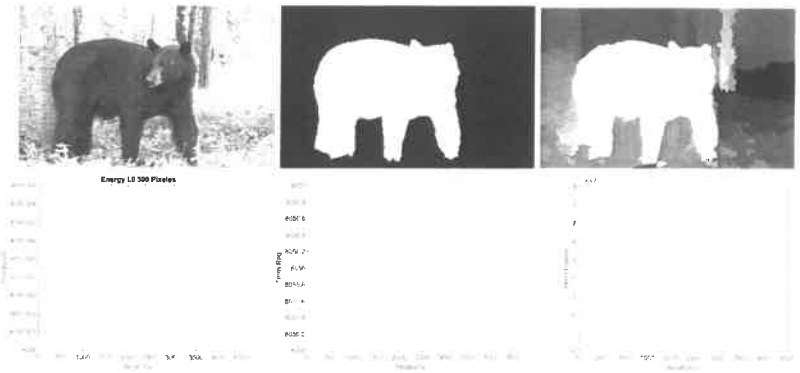

Fig. 10.3 Case 1: From left to right and top to bottom: original image, ground truth, saliency map, energy curve, L_0 term of the energy functional and the fidelity term.

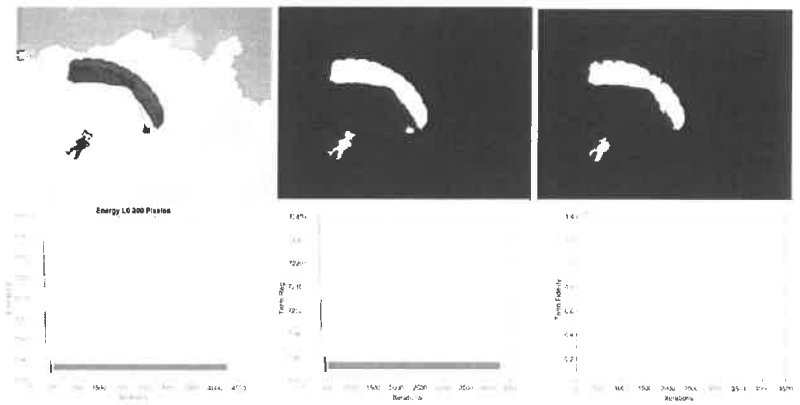

Fig. 10.4 Case 2: From left to right and top to bottom: original image, ground truth, saliency map, energy curve, L_0 term of the energy functional and the fidelity term.

$$J_{NLTV,w}(v) = \sum_{p \in G} \left(\sum_{q,pq \in E} w_{p,q}(v_q - v_p^2) \right)^{1/2}.$$

The minimization problem based on the NLTV for saliency detection we shall consider is:

$$\min_v \left(\alpha \sum_{p \in G} \|\nabla_w v_p\|_2 + \frac{\lambda}{2} \sum_{p \in G} |v_p - v_p^c|^2 \right), \tag{10.2}$$

where

$$\|\nabla_w v_p\|_2 = \left(\sum_{q,pq \in E} w_{p,q}(v_q - v_p)^2 \right)^{1/2}.$$

In order to illustrate the mathematical foundations of our numerical method we rewrite our minimization problem in the general setting of the primal problem:

$$\min_v F(\nabla_w v) + G(v), \tag{10.3}$$

with

$$F(\nabla_w v) = \alpha \sum_{p \in G} \|\nabla_w v_p\|_2 \quad \text{and} \quad G(v) = \frac{\lambda}{2} \sum_{p \in G} |v_p - v_p^c|^2.$$

10.6.1 Numerical resolution: Primal-Dual algorithm

In order to solve numerically the minimization problem (3), we generalize the Primal-Dual algorithm ([4, 11]). We briefly review the mathematical setting. The primal-dual formulation of the nonlinear primal problem in (10.3) is the saddle-point problem

$$\min_v \max_d \langle \nabla_w v, d \rangle + G(v) - F^*(d) \tag{10.4}$$

for the primal variable v, the dual variable d and F^*, the convex conjugate of F.

Assuming that these problems have a solution (v, d), it satisfies

$$\nabla_w v \in \partial F^*(d), \qquad div_w(d) \in \partial G(v)$$

where ∂F^* and ∂G are the subdifferentials of F^* and G. Introducing the resolvent operator defined through

$$\tilde{v} = (I + \tau \partial G)^{-1}(v) = \min_v \left(\frac{||v - \tilde{v}||}{2\tau} + G(v) \right)$$

the final algorithm is:

$$\begin{cases} d^{n+1} = (I + \sigma \partial F^*)^{-1}(d^n + \sigma \nabla_w v^n) \\ v^{n+1} = (I + \tau \partial G)^{-1}(v^n + \tau div_w d^n) \end{cases}$$

The algorithm therefore consists of an alternate minimization and maximization step, where the dual variable d is updated in the maximization step and the solution v is set in the minimization one. These two steps are repeated iteratively until the convergence is reached. Making explicit the above equations we have: Given the $k-$step solution (v^k, d^k):

- Maximization step: For every superpixel q compute

$$d_q^{k+1} = \frac{d_q^k + \tau_d \nabla_w v_q^k}{max(1, |d_q^k + \tau_d \nabla_w v_q^k|_\infty)}.$$

Notice that d_q^k is a vector of components $d_{q,p}^k$ for p superpixel such that $qp \in E$.

- Minimization step: Fixed d^{k+1}, we compute v^{k+1} for every pixel q:

$$v_q^{k+1} = (1 - \tau_p) v_q^k + \tau_p \left(\frac{1}{\lambda} div(d_q^{k+1}) + v_q^c \right),$$

where

$$div(d_q^{k+1}) = \sum_{p,qp \in E} (d_{q,p}^{k+1} - d_{p,q}^{k+1}) \sqrt{w_{p,q}} \tag{10.5}$$

The iterations are stopped when the difference between the values in two consecutive iterations is less than a fixed value ε.

The parameters τ_d and τ_p refers to the gradient descent steps employed to solve the maximization and minimization problems associated to the dual d and the primal variable v respectively. A pseudo code of our primal dual algorithm for saliency is shown in 1.

In figures 10.5 and 10.6, we present the results obtained with our NLTV model for the examples shown in figures 10.3 and 10.4 regarding the L_0-mode. It can be seen that in this case, convergence to a minimum of the functional is achieved, and also that the convergence is faster than in the L_0 case.

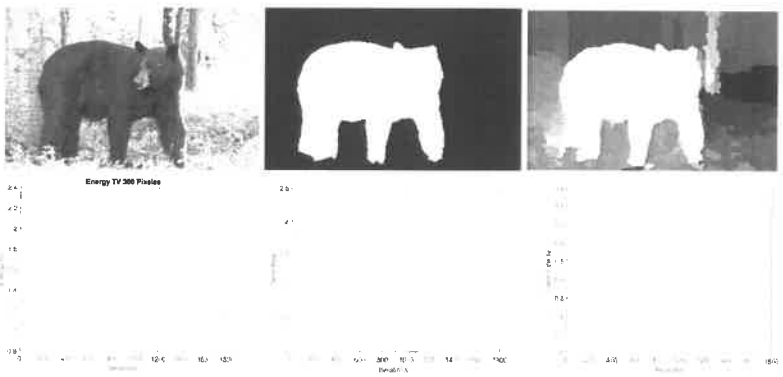

Fig. 10.5 Results obtained with the NLTV model for the image consider in case 1 (figure 10.3): from left to right and top to bottom: original image, ground truth, saliency map, energy curve, the NLTV term of the energy functional and the fidelity term.

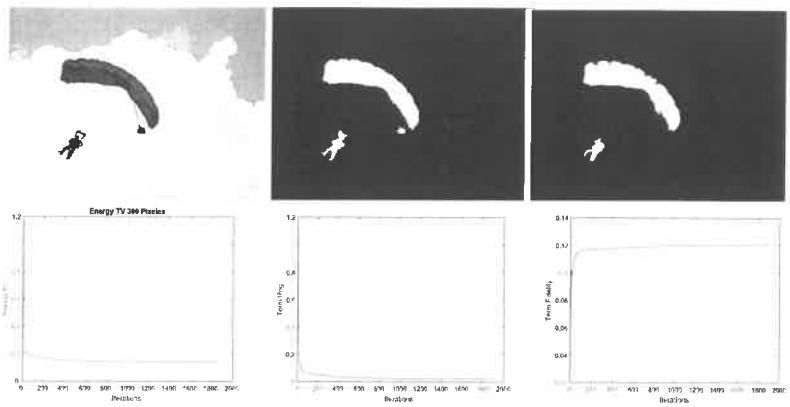

Fig. 10.6 Results obtained with the NLTV model for the image consider in case 2 (figure 10.4): from left to right and top to bottom: original image, ground truth, saliency map, energy curve, the NLTV term of the energy functional and the fidelity term.

Algorithm 1 *Saliency estimation on non local TV*

procedure SALIENCYNLTV(inputImage,parameters)
  Calculate Superpixels
  wknn ← Create knn graph
  controlMap ← Calculate controlMap
  v^k = controlMap;
  d = 0;
  repeat
    d = d + (τ_d · gradNLTV (wknn,v^k,NoSuperpixels) / max(1, | d + τ_d · gradNLTV (wknn,v^k,NoSuperpixels) |$_\infty$)) ;
    div d= (1/λ) · divNLTV(wknn,d,NoSuperpixels);
    $Prev v^k$=v^k;
    v^k = (1 -τ_p) · v^k + τ_p ·(div d + controlMap);
    energyv^k= energyNLTV(wknn,v^k,controlMap ,λ);
    stopCriteria = | Prevv^k-v^k |
    energyPrevv^k=energyv^k ;
    $iter = iter + 1$;
  until stopCriteria $\leq$ tol
  return v^k;
end procedure

end

In section 10.8, we shall show a comparison about the times of computation and also the results obtained for several measures with the L_0 and NLTV models.

10.7 Saliency map segmentation

The segmentation by fixed threshold $T_f \in [0,255]$ is the simplest method to obtain the final saliency map of the input image. Varying T_f provides also a fair methodology to compare other algorithms with the precision vs recall curves and confirm the efficiency. However, there exist methods which provide a final saliency segmentation given a saliency map such as the Saliency Binarization with Mean Shift or the Saliency Cut algorithm [6] to name a few. The Saliency Cut Algorithm is based on the GrabCut algorithm [13] and it can be initialized with the saliency map calculated with the algorithm. For this it was considered here to obtain the final saliency map.

10.8 Numerical results

The results have been carried out on MSRA10K benchmark which has 10000 images and each image has an unambiguous salient object. This benchmark provides the ground truth masks (salient objects) with pixel level accurately in comparison with MSRA where only the bounding box is provided. For L_0 algorithm, we take the parameters proposed by the authors in [15] $N = 300$ (number of superpixels), $k = 5$, $\lambda = 0.001$, $\rho = 0.0001$, $\alpha = 0.9$ and $\sigma^2 = 0.05$ for $NLTV$ algorithm $\tau_p = 0.3$, $\tau_d = 0.03$, $\lambda = 0.1$ and $N = 300$, $k = 5$, $\alpha = 0.9$ for fair comparison. There are three more parameters that have been modified (enable/disable) to see the influence in the variational methods:

- Normal: No boundaries, no location prior.
- Prior: The control map gives more importance to the center position of the superpixels in the image.
- Boundaries: Suppression of the background associating the external borders in the image and the superpixels with less value.

The naming conventions used in the experimental results are as follows:

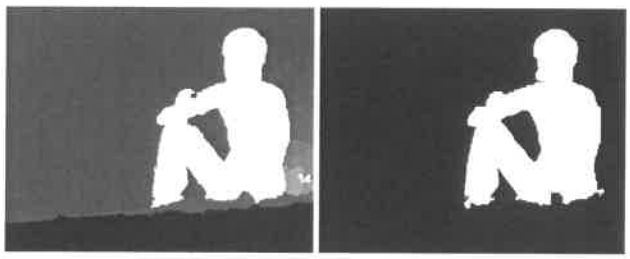

Fig. 10.7 Saliency cut algorithm in an image from MSRA10K benchmark. On the left hand side our NLTV + SCALP + Bon and right hand side the result of applying SaliencyCut to this saliency map.

- Variational method: NLTV and L_0
- Superpixel method: SLIC and SCALP
- Parameters: Loc (location prior is enabled) for the control map and Bon (Location prior is enabled as well as Boundaries) for the weights otherwise these parameters are disabled.

The results of applying the SaliencyCut algorithm to the whole benchmark MSRA10K for both saliency algorithms L_0 and NLTV is shown in the figure 10.8. We used the precision, recall and F_β measurements to compare our results with the ones obtained with the model proposed in [15], and illustrate the fact of including the location prior and the boundaries identification. On one hand, precision (positive predictive value) is the fraction of relevant instances among the retrieves instances. On the other hand, recall (sensitivity) is the fraction of relevant instances that have been retrieved over the total amount of relevant instances. The F_β measure is given by

$$F_\beta = \frac{(1+\beta)^2 Precision \cdot Recall}{\beta^2 \cdot Precision + Recall}, \qquad \beta = 0.3.$$

It can be seen that the results we have obtained are similar confirming the efficiency of our proposal.

The results in the figure 10.9 demonstrate that the quality of the solution in both cases (case 1 and case 2, see figures 10.3, 10.4, 10.5 and 10.6) is good. We compare both methods with the same initial control map enabling a fair comparison. Although the case 2 presents oscillations in the energy functional, it achieves the best performance for these two images. We understand that the oscillations and the way to reach the convergence are not determinant for the quality in the solution. However, the method we proposed achieves a much better performance in computational time while keeping the same grade of quality in terms of precision and recall metrics.

Experimentation was performed on an Intel Xeon E5-1650v3, 3.5GHz hexacore processor, from the 2014 Intel Haswell architecture (Haswell-EP), 1.5MB L2 cache and 15MB L3 cache with 64GB DDR3 RAM as a CPU platform using Microsoft Windows Server 2012R2 as operating system. this procesor can run up to 12 threads among its 6-cores simultaneously due to the Intel Hyper Threading technology (HTT). The Intel Xeon family of microprocessors belongs to the professional line instead of the more consumer oriented Intel Core family.

The algorithm has been implemented in C++ Visual Studio 2015 with Intel tools for compilation. No threads implementation has been used in our code. The time until convergence and the iterations in the case 1 and 2 for L_0 and TV are shown in the following table 10.1.

Table 10.1 Computational time (in seconds) using the two variational methods for each case 1 (Bear) and 2 (Parachute man). As it can be seen NLTV method improves considerably both iterations and time performance.

# Case	Method	Iterations	Time	Size
1	TV	1620	1.38 sec	[400×253]
	L0	4128	9,09 sec	[400×253]
2	TV	1903	1.45 sec	[400×300]
	L0	4266	8.43 sec	[400×300]

10.9 Conclusions

In this work, we have presented a new saliency model as an alternative to the sparse gradient saliency detection model based on L_0 minimization proposed in [15]. An efficient primal dual variational method to obtain the saliency of an input image has also be described and implemented. A numerical comparison is presented based on the MSRA10K benchmark dataset. The results are qualitative and quantitatively comparable to [15], but the numerical resolution is faster opening the way to automatic real time saliency detection in video and multichannel images. We also included the edges and contours of the images to generate high quality superpixels using the SCALP algorithm. Some preliminary results indicate a clear improvement but the results are not conclusive in this benchmark.

Considering the model parameter we observe that the location prior parameter has more influence than the boundaries in the generation of accurate saliency binary partitions similar to the ground truth. From the results we can see that there is a clear improvement when applying the location prior (0.89 vs 0.83 when no location prior is imposed). The initial control map is also a key ingredient to obtain high quality results and machine learning techniques shall be used in future work to learn prior

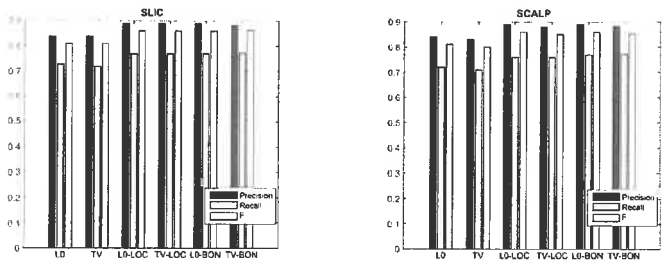

Fig. 10.8 Precision, recall and F-measure for the complete data set MSRA10K using SLIC and SCALP as superpixels method and L_0 and NLTV as the regularization terms and Loc and Bon mean with Location prior and Boundaries background suppression enabled.

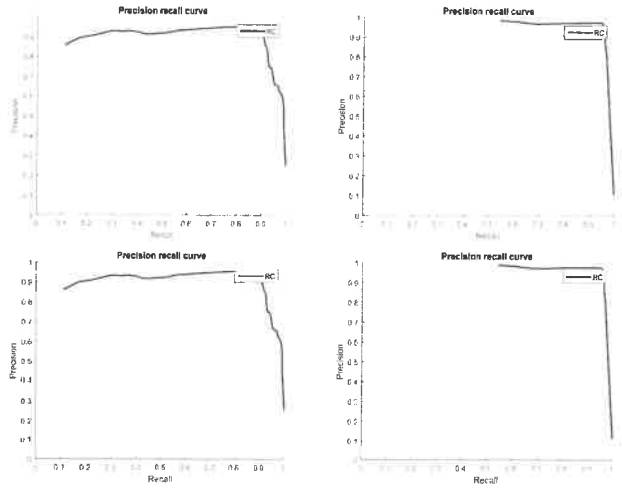

Fig. 10.9 On the top of the image we present from left to right the results obtained in the L_0 model for the precision vs recall curves in the case 1 and case 2 . On the bottom, the ones obtained with our model for the same cases.

information about the object to be detected. This will allow to develop models tailored to specific saliency detection tasks.

The automatic segmentation has proven to be a robust method to efficiently segment the final saliency against the fixed threshold. Regarding the numerical performance the computation time is less than 0.5 seconds per image when the converge criteria is not that strict making this technique promising for real time systems.

In future research, we shall extend this computation to GPGPU[1] to be able to detect saliency in video in real time.

Acknowledgements This research was partially supported by projects TIN2015-69542-C2-1-R and MTM2014-57158-R.

References

1. Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Suesstrunk, S.: SLIC Superpixels Compared to State-of-the-art Superpixel Methods. IEEE Transactions on Pattern Analysis and Machine Intelligence **34**, 2274–2282 (2012)
http://dx.doi.org/10.1109/TPAMI.2012.120

[1] General-purpose computing on graphics processing

2. Belongie, S., Carson, C., Greenspan, H., Malik, J.: Color- and Texture-Based Image Segmentation Using EM and Its Application to Content-Based Image Retrieval. In: IEEE International Conference on Computer Vision (ICCV) (1998)
http://dl.acm.org/citation.cfm?id=938978.939161
3. Belyaev, A.: On Implicit Image Derivatives and Their Applications (2012)
http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.413.634
4. Chambolle, A., Pock, T.: A first order primal-dual algorithm for convex problems with applications to imaging. J. Math. Imaging. Vis. **40**, 120–145 (2011)
5. Cheng, M., Mitra, N., Huang, X., Hu, S.: SalientShape: group saliency in image collections. The Visual Computer **30**, 443–453 (2014)
http://dx.doi.org/10.1007/s00371-013-0867-4
6. Cheng, M., Mitra, N., Huang, X., Torr, P., Hu, S.: Global Contrast based Salient Region Detection. IEEE TPAMI (2015)
http://mmcheng.net/salobj/
7. Christopoulos, C., Skodras, A., Ebrahimi, T.: The JPEG2000 still image coding system: an overview. IEEE Transactions on Consumer Electronics **46**, 1103–1127 (2000)
http://ieeexplore.ieee.org/document/920468/
8. Dollár, P., Lawrence Zitnick, C.: Fast Edge Detection Using Structured Forests, IEEE Trans. Pattern Anal. Mach. Intell. **37**, 1558–1570 (2015)
http://dblp.uni-trier.de/rec/bib/journals/pami/DollarZ15
9. Elmoataz, A., Lezoray, O., Bougleux, S.: Nonlocal discrete regularization on weighted graphs: a framework for image and manifold processing. IEEE Trans. on Image Processing **17**, 1047–1060 (2008)
10. Gilboa, G., Osher, S.: Nonlocal Operators with Applications to Image Processing. SIAM Multiscale Mod. Simul. (MMS) **7**, 1005–1028 (2008)
11. Martín, A., Garamendi, J.F., Schiavi, E.: Two efficient primal-dual algorithms for MRI rician denoising. In: Computational Modelling of Objects Represented in Images III, pp. 291–296 (2013)
http://10.1201/b12753-54
12. Rémi, G., Vinh-Thong, T., Papadakis, N.: SCALP: Superpixels with Contour Adherence using Linear Path. In: 23rd International Conference on Pattern Recognition (ICPR 2016), Cancún, México (2016)
https://hal.archives-ouvertes.fr/hal-01349569
13. Rother, C., Kolmogorov, V., Blake, A.: GrabCut: Interactive Foreground Extraction Using Iterated Graph Cuts. ACM Trans. Graph. (2004)
http://doi.acm.org/10.1145/1015706.1015720
14. Santella, A., Agrawala, M., DeCarlo, D., Salesin, D., Cohen, M.: Gaze-based Interaction for Semi-automatic Photo Cropping. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 771–780 (2006)
http://doi.acm.org/10.1145/1124772.1124886
15. Wang, Y., Liu, R., Song, X., Zhixun, S.: Saliency Detection via Nonlocal L_0 Minimization. In: Computer Vision ACCV 2014, Lecture Notes in Computer Vision **9004**, Springer, pp. 521–535 (2014)