Medial Features for Superpixel Segmentation

David Engel* Luciano Spinello† Rudolph Triebel† Roland Siegwart†
Heinrich H. Bülthoff* Cristóbal Curio*

*Max Planck Institute for Biological Cybernetics
Spemannstr. 38, Tübingen, Germany
firstname.lastname@tuebingen.mpg.de

†Autonomous Systems Lab, ETH Zürich
Tannenstrasse 3, Zürich, Switzerland
firstname.lastname@mavt.ethz.ch

Abstract

Image segmentation plays an important role in computer vision and human scene perception. Image oversegmentation is a common technique to overcome the problem of managing the high number of pixels and the reasoning among them. Specifically, a local and coherent cluster that contains a statistically homogeneous region is denoted as a superpixel. In this paper we propose a novel algorithm that segments an image into superpixels employing a new kind of shape centered feature which serve as a seed points for image segmentation, based on Gradient Vector Flow fields (GVF) [14]. The features are located at image locations with salient symmetry. We compare our algorithm to state-of-the-art superpixel algorithms and demonstrate a performance increase on the standard Berkeley Segmentation Dataset.

1 Introduction

Image segmentation plays an important role in computer vision and human perception. This procedure associates a coherent and meaningful label to each pixel of an image. An image segmentation technique that overcomes two well known issues in the literature is needed: pixels are not natural entities but are a consequence of the quantized representation of images and the number of pixels grows quickly with respect to the resolution. Superpixel techniques are oversegmentation methods that address this problem. A superpixel is a local and coherent cluster that contains a statistically homogeneous image region. Such a method was introduced by Ren and Malik [10] who employ a Normalized Cut criterion [12] to recursively partition an image using contour and texture cues. Another method has been proposed by Felzenszwalb and Huttenlocher [4] using an efficient graph based representation of local neighborhoods. Several applications that use superpixels exist. Recent noteworthy works include depth from single images [11], human pose estimation [8] and general scene understanding [5].

In this paper we present a novel superpixel segmentation methodology that makes use of a new kind of medial feature transform [3]. The transform has maximum responses at image locations of high symmetry and carries the notion of shape-centered medial features (cf. [1]). The novelty of our proposed algorithm is that it exploits the properties of the well known Gradient Vector Flow (GVF) field proposed by Xu and Prince [14]. GVF basically yields long ranging image force vector fields that has been useful in many applications by attracting contours even into areas of strong boundary concavities and in the presence of noise in the original image. We exploit the robustly derived long ranging image forces represented by a vector field, by detecting singularities in them as seeds for image segmentation. The medial features that are retrieved in our context with the help of the GVF field's singularities can be used to compress regional shape information to a few informative image points from which the image even could be reconstructed [13].

To evaluate the performance of the proposed algorithm, we employ the Berkeley Segmentation Dataset [6] and compare the performance of our algorithm to the approaches of Felzenszwalb and Huttenlocher [4] and Ren and Malik [10].

The paper is structured as follows. In Section 2.1 we describe the applied Medial Feature transform and introduce an extension to oversegmentation in Section 2.2. In Section 3 we report on our evaluation and we conclude this work in Section 4.

2 Methods

To give a better intuition on the features used in our segmentation algorithm we first outline the computation of the GVF field that we operate on. The overall pipeline for superpixel segmentation is visualized in Figure 1.

2.1 Medial Features

The GVF [14] is the vector field $V(p) = [u(p), v(p)]^T$ that minimizes the cost function $\mathcal{E}$, where $p = (x, y)$ is a point in the image $\mathcal{I}$:

$$\mathcal{E} = \int\int \underbrace{g\left(|\nabla f|\right)|V - \nabla f|^2}_{\text{data term}} + \underbrace{h\left(|\nabla f|\right)\nabla^2 V}_{\text{smoothing term}}\, dxdy$$

This cost function is subject to the iterative optimization of V until convergence and is obtained with variational calculus. The data term guarantees stability of the vector field $V(p)$ near an edge map f whereas the second term is responsible for the suppression of spurious edges and the propagation of orientation information across the image. The optimization aims to accomplish the two goals of preserving the orientation information at the gradients and creating a smooth flow field across the image. The functions g and h determine the trade off between these two conflicting goals. They are designed to be complementary, enforcing stricter adherence to the underlying edge map at locations of high gradient magnitude and smoothness where the magnitude is low. For our purposes we followed the implementation of [14] using a constant for $h = 0.12$ and for data function $g = |\nabla f|^2$.

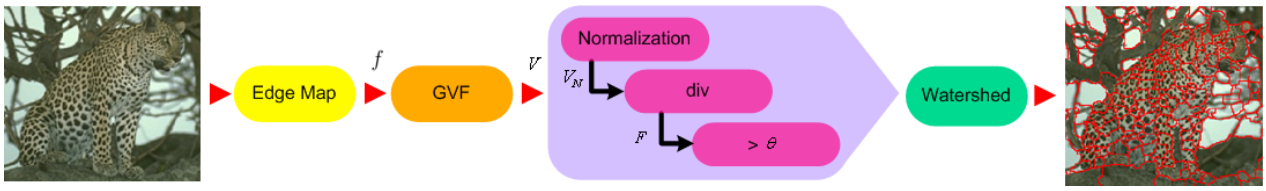

Figure 1: Overview of the processing pipeline for medial feature oversegmentation. The GVF field V is computed on the edge map. After normalization of V and the computation of the flux flow $\mathcal{F}$ the seeds are obtained by thresholding with θ. Starting from the seeds the watershed is computed using $\mathcal{F}$ as a height map.

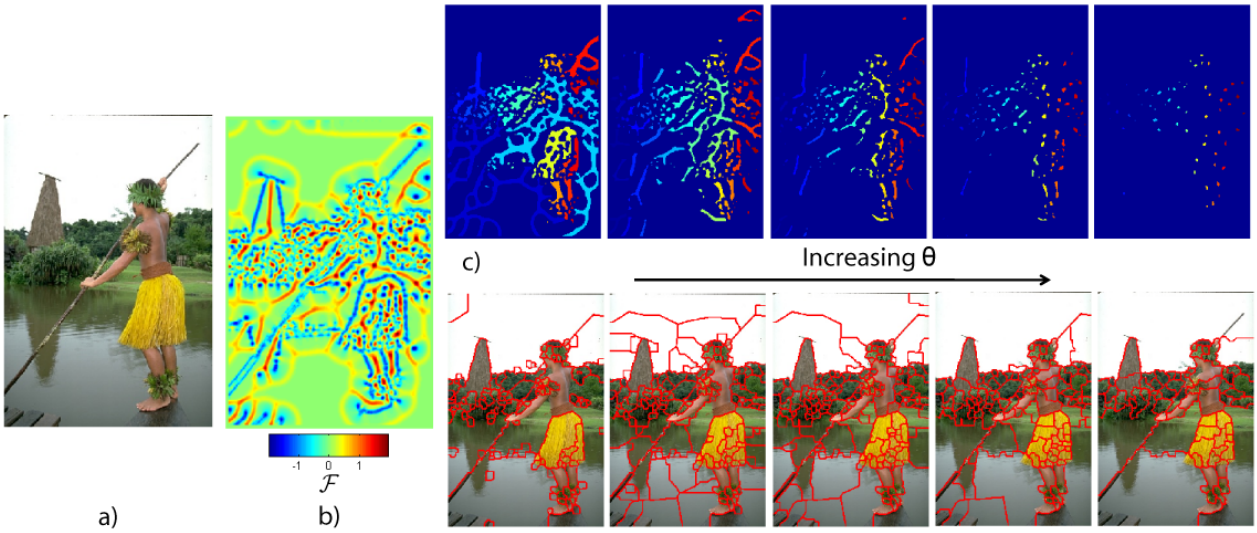

Figure 2: a) Original image b) Flux flow field $\mathcal{F}$ c) Influence of the thresholding parameter θ for the seeding on the resulting superpixel segmentations. From left to right parameter is set to $(0.2, 0.5, 0.8, 1.1, 1.4)$. Top: $\mathcal{F} > \theta$, Bottom: Resulting oversegmentation.

We normalize the solution to $V(p)$ at each image location resulting in $V_N(p)$. Our assumption is that $V_N(p)$ closely approximates the gradient of the L_2-norm distance function, $\nabla D(p)$, with $\nabla D(p) \approx V_N(p) = V(p)/\|V(p)\| \; \forall p$. Given that, we can locate shock loci in $V_N(p)$, which we use as seeds for segmentation as follows. Pizer et al [9] suggest a singularity detection framework based the divergence of V_N yielding the flux flow

$$div\, V_N = \mathcal{F}\left(V_N\left(p\right)\right) = \frac{\oint \langle V_N, \mathcal{N} \rangle \, ds}{Area}, \qquad (1)$$

where $\mathcal{N}$ denote the normals on a ring with diameter of 7 pixels through which the flux flow $\mathcal{F}$ is computed. The computation of this ring integral at each point in the image can be implemented for the two components of V_N as two convolutions with two precomputed kernels containing the two normal vector components of that ring, respectively. This flux is further used as seeds for oversegmentation described as follows.

2.2 Oversegmentation using Medial Features

The idea behind oversegmentation algorithms is to find and group together regions which are uniform in their appearance. The medial features described above provide a means to this end by being formed at the centers of regions of uniform appearance. We take advantage of the points of high symmetry denoted by the medial features as seeds for the oversegmentation. To obtain these seeds we threshold the flux flow field $\mathcal{F}$ with θ and assign unique labels to the connected areas. The GVF field is the first step of calculating the medial features and operates on an edge image f. Choosing the right edge detector to create f is task dependent. Evaluations showed the thresholded Sobel edge operator performs best in the context of oversegmentation. More intricate edge detectors such as Canny [2] suppress fine edge details which is not desirable for our application, since creating too many segments is more desirable than prematurely merging segments belonging to two different regions. Furthermore, the GVF will eliminate spurious edge pixels that are potentially created by the less complex edge detector, allowing a stable formation of the seeds for the oversegmentation. An example of the seed structure after thresholding the medial features is shown in Figure 3 (middle). To avoid problems in large uniform areas the GVF should have fully converged over the image.

To complete the oversegmentation we need to assign the remaining pixels to coherent regions. To this end we apply the watershed algorithm proposed by [7] for which efficient implementations are available. This operates on a height map and simulates successive flooding of the relief, starting from the minima of the image. Borders or 'watersheds' are formed where the rising water of two different basins meet. As a height map we use the flux flow image $\mathcal{F}$ described above (cf. Figure 2b). As an outcome of the GVF optimization process, $\mathcal{F}$ preserves the salient edge information (local minima, negative) complementary to the formed symmetries (local maxima, positive). Thus, this approach

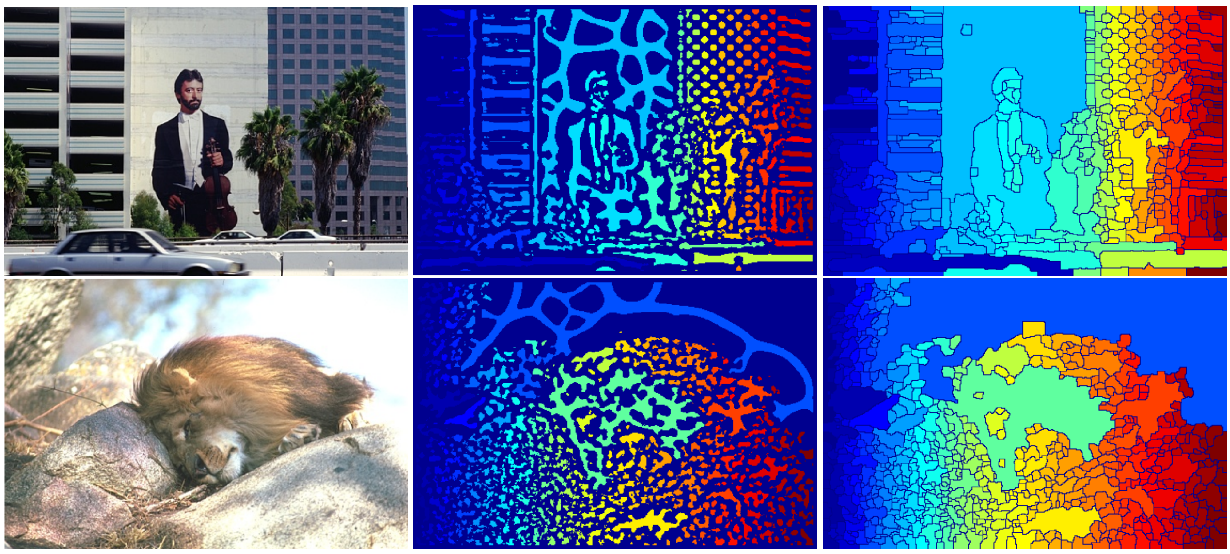

Figure 3: Seeds and segmentation. From left to right: original images, GVF based seeds obtained by thresholding the flux flow $\mathcal{F}$ with θ and the watershed segmentation. The color values denote different labels of the seeds and segments.

preserves the edge structure of the original image which is critical for an oversegmentation algorithm. This preservation of the underlying edge structure is also an advantage over a simple creation of a Voronoi diagram extended from the seeds. A resulting segmentation can be seen in Figure 3 (right).

3 Evaluation

We compare our oversegmentation method with two state-of-the-art approaches. The first one is an extension of the superpixel algorithm proposed by Felzenszwalb and Huttenlocher [4] and the second one is the method proposed by Ren and Malik [10], which is based on normalized graph cuts. As a baseline method we show the performance of a standard watershed method computed on Canny edge maps without initialization. We measure the performance of the oversegmentation algorithms on the Berkeley Segmentation Dataset [6], which contains 300 images (200 in the training data set and 100 in the test set) with several human drawn segmentations for each image. Accordingly, the segmentations are very different among subjects with the number of segments per image reaching from 5 to more than 30. Furthermore, human observers use their vast experience with natural images and their knowledge of the image semantics for segmentation both of which are not available to bottom-up segmentation algorithms. Thus, reconstructing a segmentation by humans is a very difficult challenge.

Superpixel algorithms do not aim at fully explaining human segmentations but provide a starting point for higher level segmentation algorithms. However, should the superpixel segmentation already cross borders between human-made segments an algorithm operating on the oversegmentation would have to fail. Thus, we have to use a performance measure which tells us how well a higher-level algorithm *could* be able to reconstruct the human segmentation. Consequently, our performance measure penalizes segments of the oversegmentation that cross the borders of the target shape. On the other hand, it is desirable to end up with a small number of segments to reduce the complexity of the merging problem. Based on these observations we formulated the following performance measure

Medial Feature Superpixel	0.88
Ren et al. [10]	0.86
Felzenszwalb et al. [4]	0.83
Watershed on Distance Transform	0.79

Table 1: Performance of the algorithms.

$$P = \frac{\sum_{i=1}^{N} \sum_{j=1}^{M_i} \hat{S}_{i,j}}{\sum_{i=1}^{N} M_i S_i}. \qquad (2)$$

This is computed over N images where M_i denotes the number of human segmentations of image i in the dataset. S_i is the number of segments the algorithm produced on image i while $\hat{S}_{i,j}$ is the number of segments produced by the superpixel algorithm that lie inside only one segment, j, of the human created segmentation of image i. To compensate for some noise and uncertainty in the human segmentations the criterion for $\hat{S}_{i,j}$ was relaxed so that only 95% of a superpixel had to be consistent with the human segmentation.

This performance measure indicates the percentage of segments that do not cross a border of the segmentations drawn by the human normalized against the number of segments produced by the oversegmentation algorithm. Taking the mean of these values across all test images and all human segmentations for each of the images gives the resulting performance measures reported in Table 1.

All three algorithms possess free parameters that influence the properties and number of created superpixels (e.g. parameter k from [10] or the threshold θ for the seeding in our algorithm). These parameters were optimized by a grid search for the maximum of the performance measure using the images from the training set of the Berkeley Segmentation Dataset. As a baseline method we employed a standard watershed algorithm applied to the distance transformation of a Canny edge image.

The results of this performance measure indicate that the proposed medial feature segmentation yields better results

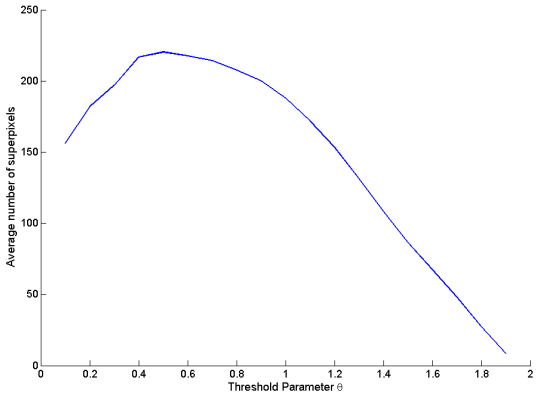

Figure 4: Influence of the thresholding parameter θ on the number of generated superpixels (averaged over all images in the Berkeley Segmentation Dataset).

than the other algorithms and is substantially better than a baseline method. Other performance measures (such as conservation of human segmentation boundaries) are possible and might yield different results. However, we feel that the measure chosen here is appropriate since it provides an indicator for how useful the created superpixels will be for algorithms operating on them. The algorithm by Felzenszwalb and Huttenlocher creates non-smooth borders which can be suboptimal for this performance measure. The superpixel algorithm based on normalized graph cuts produces smooth segment boundaries but as it produces a fixed number of segments (the number of superpixels k was optimized using the training data set) it can not be optimal for a heterogeneous image set.

The average number of superpixels generated by our algorithm depends strongly on the threshold parameter θ as can be seen in Figure 4. For large values of θ only few distinct symmetry points are above the threshold and remain as seeds resulting in a small number of large segments. As the threshold becomes smaller more seeds are generated resulting in higher number of starting areas for the watershed algorithm and consequently in a finer superpixel segmentation. However, the seed regions begin to merge for very small values of θ as can be seen in the leftmost image of Figure 2c, resulting in a decreased number of segments. The location of the maximum depends on the properties of the underlying image.

The average runtime of the medial feature oversegmentation per image is 2.6 seconds in the current Matlab/C-Mex implementation, which is about as fast as the segmentation algorithm proposed by Felzenszwalb and Huttenlocher and much faster than the oversegmentation based on graph cuts. However, the largest portion of this time is taken by the iterative computation of the GVF field. The iterative optimization of the GVF is well parallelizable and there are now GPU implementations available which will further reduce the computation time.

4 Conclusion and Outlook

In this paper we presented a novel way of image oversegmentation based on medial features. The medial features are computed by applying a divergence operator to the GVF field and are formed at points of high symmetry and are therefore well suited as seeds for a segmentation approach based on the watershed algorithm. Using such medial features allows our algorithm to be very efficient and offers many desirable properties such as stability against noise. We compared our algorithm to two state-of-the-art algorithms on the Berkeley Segmentation dataset. We showed that our method can provide a basis for higher level algorithms by producing a high percentage of segments that are consistent with segments found by human observers.

We plan to employ a GPU implementation of the GVF to obtain substantial speedups for the oversegmentation. As any other superpixel algorithm our algorithm offers the opportunity for computer vision frameworks to merge the created segments to arrive at final figure-ground segmentation of objects. Using medial features in large uniform areas can slow the superpixel segmentation down since it takes the GVF many iterations to converge in such regions. To address this problem we plan to integrate medial features extracted along a pyramid of scales to create the seeds for the segmentation.

Acknowledgements

This work was supported by EU-Project BACS FP6-IST-027140 and DFG Perceptual Graphics.

References

[1] H. Blum. A transformation for extracting new descriptors of shape. In W. Wathen-Dunn, editor, *Models for the Perception of Speech and Visual Form*, pages 363–380, Washington, DC, USA, 1967. MIT Press.

[2] F. J. Canny. A Computational Approach to Edge Detection. *IEEE-PAMI*, 8(6):679–698, 1986.

[3] D. Engel and C. Curio. Scale-invariant medial features based on gradient vector flow fields. In *ICPR*, December 2008.

[4] P. Felzenszwalb and D. Huttenlocher. Efficient graph-based image segmentation. *Int. J. of Computer Vision*, 59(2):167–181, 2004.

[5] D. Hoiem, A. Efros, and M. Hebert. Closing the loop on scene interpretation. In *CVPR)*, 2008.

[6] D. Martin, C. Fowlkes, D. Tal, and J. Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *ICCV*, volume 2, pages 416–423, July 2001.

[7] F. Meyer. Topographic distance and watershed lines. *Signal Processing*, 38(1):113–125, 1994.

[8] G. Mori, X. Ren, A. Efros, and J. Malik. Recovering human body configurations: combining segmentation and recognition. In *CVPR*, volume 2, pages 326–333, 2004.

[9] S. Pizer, K. Siddiqi, G. Szekely, and S. Zucker. Multiscale medial loci and their properties. *Int. J. of Computer Vision*, 55(2-3):155–179, 2003.

[10] X. Ren and J. Malik. Learning a classification model for segmentation. In *ICCV*, 2003.

[11] A. Saxena, S. H. Chung, and A. Y. Ng. 3-d depth reconstruction from a single still image. *Int. J. of Computer Vision*, 2007.

[12] J. Shi and J. Malik. Normalized cuts and image segmentation. *PAMI*, 22:888–905, 2000.

[13] A. Tamrakar and B. Kimia. Medial visual fragments as an intermediate image representation for segmentation and perceptual grouping. In *Proc. of the Conference on Computer Vision and Pattern Recognition Workshop*, volume 4, page 47, Washington, DC, USA, 2004. IEEE Computer Society.

[14] C. Xu and J. Prince. Snakes, shapes, and gradient vector flow. *IEEE Transactions on Image Processing*, 7:359–369, 1998.

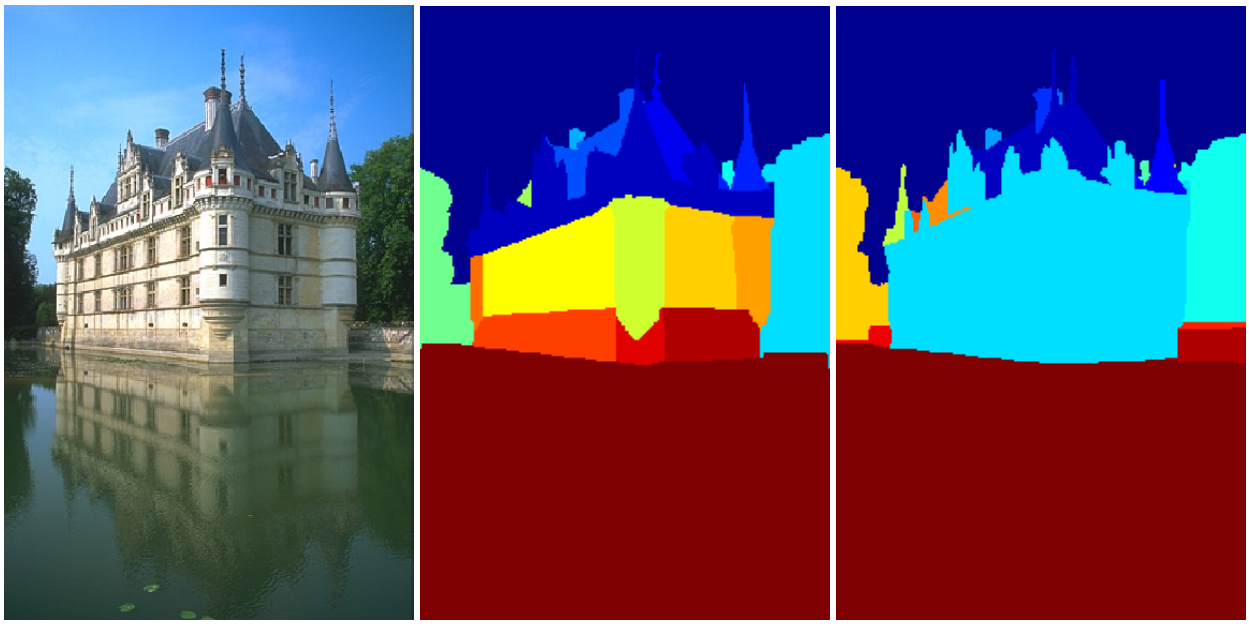

Figure 5: Example image from the Berkeley Segmentation Dataset (left) and two human segmentation results (middle, right).

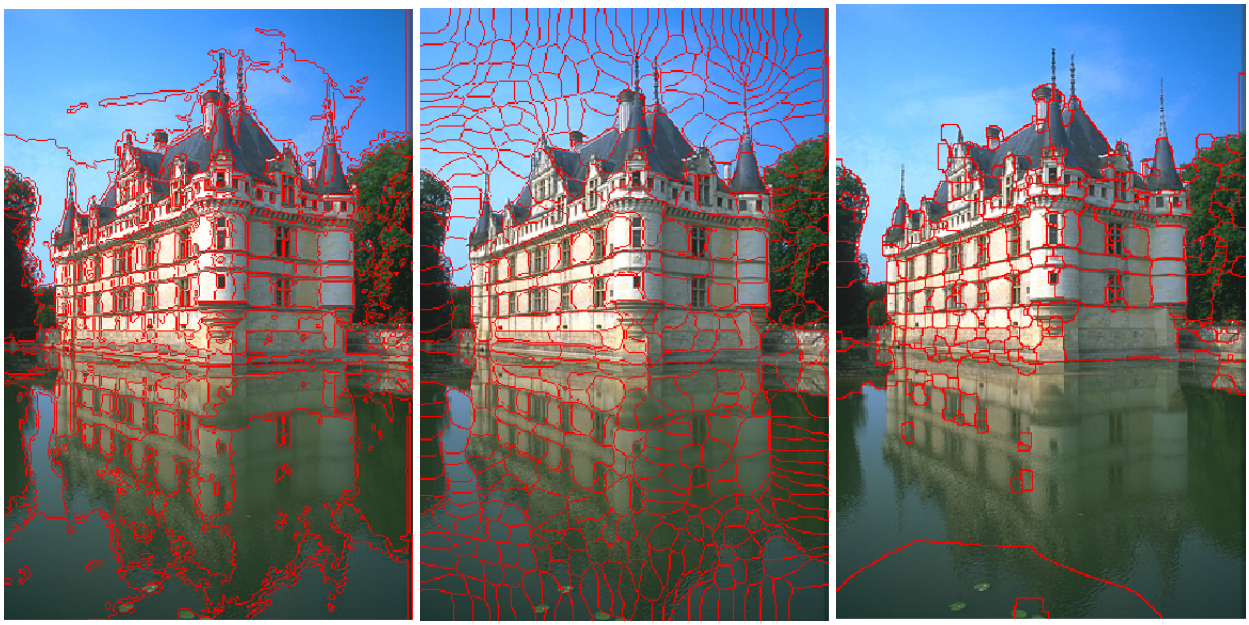

Figure 6: Experimental result on the example image from Figure 5. From left to right: segmentation obtained by superpixel algorithm after [4], segmentation obtained by superpixel algorithm after [10] and the oversegmentation result of our medial feature based approach.