Novel features in object detection for image segmentation tasks

A thesis submitted for the degree of
Doctor of Philosophy

Andrea Manno-Kovács

Scientific advisers:
Tamás Szirányi, D.Sc.
Zoltán Vidnyánszky, D.Sc.

Faculty of Information Technology
Pázmány Péter Catholic University

Computer and Automation Research Institute
Hungarian Academy of Sciences

Budapest, 2013

Acknowledgements

Due to the electrical engineering family background, I have always been keen on mathematics and computer science. This enthusiasm led me since my childhood, cherished by Sándor Dobos, form master of my class majoring in mathematics in Fazekas Mihály Secondary School. However, the past years required continuous effort and commitment. This is why I am deeply grateful to my supervisor Tamás Szirányi, for accepting my supervision and supporting me in so many ways during my studies. His guidance and advices meant a great help in progressing with my work, as well as his invitation to join his research laboratory. I would also like to thank my other supervisor, Zoltán Vidnyánszky for his encouraging and giving the opportunity to deepen my research experiments in medical imaging in cooperation with the Semmelweis University.

I am very thankful to Tamás Roska and Péter Szolgay for providing me the opportunity to spend my Ph.D. years at the Doctoral School of Interdisciplinary Sciences and Technology in Pázmány Péter Catholic University (PPCU), and broadening my view in many fields of interest.

The support of the Computer and Automation Research Institute of the Hungarian Academy of Sciences (MTA SZTAKI) is gratefully acknowledged for employing and awarding me with the Young Researchers' Academic Grant to proceed with my research. I thank my present and former colleagues in MTA SZTAKI, especially to István Méder and the members of the Distributed Events Analysis Research Laboratory headed by Tamás Szirányi for helping me with all of their professional and non-professional advices: Mónika Barti, Csaba

Benedek, László Havasi, Anita Keszler, Ákos Kiss, Levente Kovács, Zoltán Szlávik and Ákos Utasi. Thanks to Eszter Nagy, Virág Stossek and Jenő Breyer for arranging all my scientific visits 'to infinity and beyond'.

I thank the support of Zsuzsa Vágó at PPCU in my teaching and the help of PPCU Students' Office, Financial Department and Viktória Sifter from the Library.

Thanks to former and present, older any younger fellow Ph.D. students, especially to Judit Rónai, Petra Hermann, Ádám Balogh, Ádám Fekete, László Füredi, András Gelencsér, Zoltán Kárász, László Kozák, Vilmos Szabó, Kálmán Tornai, Balázs Varga, András Kiss and Dániel Szolgay.

I would like to thank the financial support of the Hungarian Scientific Research Fund under Grant No. 76159 and 80352. The help of the Hungarian Institute of Geodesy, Cartography and Remote Sensing (FÖMI) and Péter Barsi at the Semmelweis University MR Research Center (MRKK) is acknowledged for providing special images for my research.

I wish to gratefully thank my friends and my family: my grandparents, my brother, my father and especially my mother for giving me a good foundation, teaching me all the qualities and supporting me in all possible ways. I am also thankful to my in-laws I got during my studies, especially to my father-in-law, for his wise and peaceful guidance.

And most of all for my loving, encouraging and patient husband Balázs for believing in me, thank you with all my heart and soul.

Finally, I would like to dedicate this work to my lost Grandma Márta, who left us too soon. I hope that this work makes you proud.

Abstract

In this thesis, novel feature extraction methods are introduced for different object detection tasks in image segmentation, focusing on applications from video surveillance, aerial and medical image analysis. The main contributions are all related to object detection tasks performed by active contour methods, but the primary aim is differing from low dimensional representation of local contours for image matching and change detection to improving feature extraction for contour detection. The developed methods are tested on a wide range of images, including artificial and real world images for confirming that the proposed novelties besides can significantly improve the detection results compared to existing approaches.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Automatic detection is a very important task in several computer vision and image understanding applications. Nowadays, with the widespread availability of affordable digital imaging devices and the presence of high capacity personal computers, significance of digital image processing is increasing substantially. As the amount of digital data is huge and manual administration and operation is unmanageable, automatic processing techniques are continuously improved to solve complex challenges in various fields of interest, like video surveillance [42], change detection [43, 81], medical image analysis [76, 77], force protection and defense applications [83], urban area extraction [67, 87] and building detection [89, 74] in aerial images.

As the large variety of applications shows, there is a wide range of tasks to be resolved: different fields have separate concepts and methodology, therefore context-sensitive solutions should be developed to satisfy the special conditions, cope with the altering challenges and reach the exact goals. The aim of this thesis is to present contributions in three main tasks of automatic detection. Although these tasks are related to each other, and the given contributions of different tasks can be fused to be applied for complex solutions, these tasks should be handled separately. The given solutions can all be labeled as techniques for object featuring, but the distinct aims and applications (like extraction, tracking, change detection) need different tools and developments. The short introduction of the three main tasks referred in this thesis:

- **Task 1:** Giving a low dimensional feature descriptor by exploiting local

structure information around feature points. In this task, the usability of the local shape representation is investigated, where an image series or video is given with a fixed camera position either remaining static or rotating during scanning. The local properties are extracted by generating active contour in the small proximity of the feature point, which is then represented by low dimensional Fourier descriptors. The goal is to match feature points through this descriptor set for further post-processing (tracking, classification and change detection).

- **Task 2:** Improving the detection accuracy of parametric active contour algorithms for high curvature, noisy boundaries and giving an automatic initialization technique on single object images. This involves the analysis of the behavior of existing active contour methods and the development of a new feature map which is able to emphasize complex boundary parts and support initialization by defining feature points simultaneously. Moreover, the proposed feature map and point set is adapted successfully for other change detection and multi-object detection applications in registered aerial and medical image series as well.

- **Task 3:** Detecting built-in areas and building outlines in single airborne images by introducing orientation of the close proximity of the feature points as a novel feature. This task involves the comparison of different feature point detectors for built-in area detection, the statistical analysis of extracted 'high-level' orientation feature to define main directions of the urban area, and a building detection process applying the given directions for extracting the accurate shape of the buildings without any restrictions (like shape template).

The inputs of all the detailed tasks are digital images and the aim is to perform automatic object detection (outlining) which is then improved to reach the specific goals. *Task 1* first outlines the local shape, which is then described by reduced dimensional Fourier descriptors to be compared for different computer vision applications. Besides, the ultimate purpose of *Task 2* and *3* is to detect object contours as precisely as possible. Depending on the application, the detected

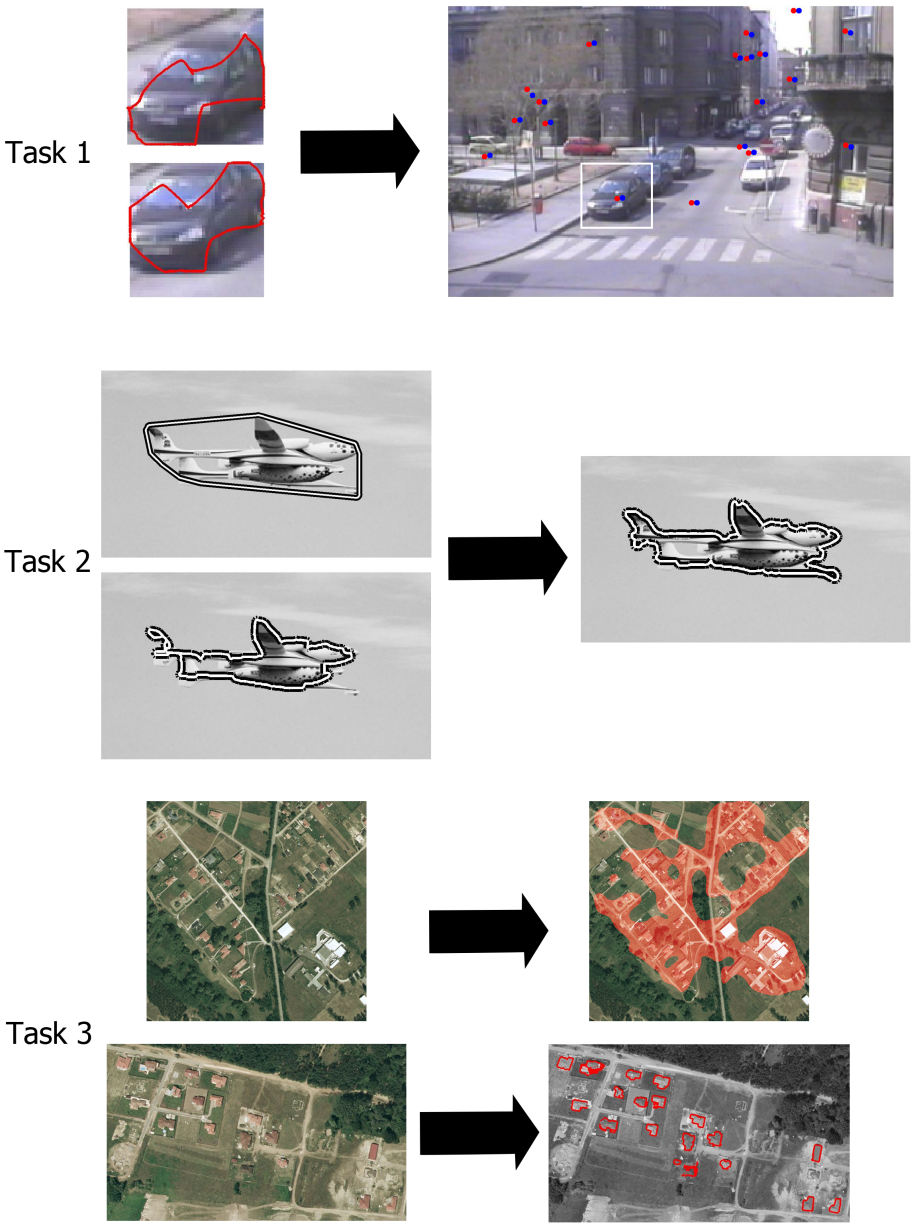


Figure 1.1: Demonstration of the results of the different tasks developed in the dissertation. In *Task 1* the represented local shapes and the matched feature points are shown; in *Task 2* the detection result of the improved parametric active contour method is presented and compared to the original algorithm; in *Task 3* the extracted urban area and building outlines are marked in red.

object can vary from a single object (in *Task 2*) where the point is to recognize the accurate boundary regardless of its complexity, through an appearing Sclerosis Multiplex lesion in a single channel MRI image pair taken by a long time interval to support the radiologist, to an urban area or building in a single airborne image.

Outlining objects is an efficient step in every high-level segmentation process. It can be considered as a preprocessing step for complex applications. When searching for an object contour, low-level tasks, like edge or line detection could be used as steps of bottom-up processes which are then followed by connection of different segments. As the bottom-up approach is a rigidly sequential process, mistakes made at deeper levels are propagated with no chance for correction in higher levels. Therefore, the aim is to use some higher level technique to avoid drawbacks of the sequential method.

Active contour model [17] was introduced for this kind of tasks, using energy minimization as a framework. An energy function is designed by adding suitable energy terms to the minimization, which is able to converge to the desired contour. The model is guided by interactive techniques, developing effective energy functions having local minima in the desired places (on accurate contours) and depending slightly on the starting points.

The novelty of the theory was the management of the contour: handling the whole boundary as a connected system and approximating it in one step, unlike traditional methods performing edge detection first and then linking them in a separate step.

In case of the active model, the connected contour system is represented by the well constructed energy terms, which are also responsible for the smoothness of the controlled curve (internal force), for the convergence of the spline towards salient image features like edges, corners (image force) and minimizing the whole energy function to the desired local minimum (external constraint) at the same time. These forces include user interface, automatic attentional mechanisms and high-level interpretation as well.

The efficiency of the active contour model relies on the definition of the energy terms. Therefore, adaptations of the basic model attempt to design such terms, which are able to represents the aforementioned principles. From the given constraints, this dissertation concentrates on the image characteristics and

tries to improve the image-related energy term. Beside this main concept, it also investigates the significance of the local characteristics applied for higher-level purposes.

When designing an efficient function for representing image characteristics, the given function can also be applied for other aims, such as feature point detection. Due to this advantage, it can also contribute to the initialization process of the active contour model, reducing human interaction.

Moreover, efficient feature maps are constructed with 'higher-level' information interpretation. By exploiting novel sources, like the connected topology of different objects through such channels as their related orientation, accurate object detection task can be raised to even a higher level by invoking an inter-object layer for feature extraction and energy term design.

Regarding the contribution of the thesis in the different tasks: *Task 1* introduces a low dimensional representation of a converged active contour curve representing the local characteristics of the small neighborhood of a feature point; *Task 2* proposes a novel feature map for the image force term of parametric active contour models; *Task 3* investigates the usability of 'higher-level' feature extraction for efficient object detection.

The outline of the thesis is as follows. Chapters 2 – 4 introduce the contributions of the thesis in *Task 1 – 3*, by dedicating each chapter to one task. Chapter 2 concentrates on the challenges of *Task 1* and introduces a low dimensional descriptor representing the local characteristics of feature points. Beside describing the novel descriptor, it also investigates the descriptor's applicability for important computer vision tasks: point matching, texture classification and change detection. Chapter 3 presents an improved feature map and initialization approach which can be successfully adapted for parametric active contour method. The introduced method is then extended for other medical and aerial applications for change detection and multi-object detection. Chapter 4 focuses on novel feature extraction for built-in area and building contour detection in airborne images. Finally, a conclusion and the summary of novel scientific results conclude the dissertation.

Chapter 2

Local Contour Descriptors

When searching for an efficient descriptor, the task is twofold: features must describe the featuring patches at a high efficiency, while the dimensionality should be kept at a manageable low value.

The main assumption in finding local descriptors is the defect of continuity in the discrete neighborhood or the imperfectness of local shape formats. This chapter investigates the potential applicability of methods in which some formal meaning of the local properties can be represented at a reduced dimension. Curve fitting methods for noisy shapes are called here: active contours. A new local feature descriptor is defined by generating active contours around keypoints. Local contours are characterized by a small number of Fourier descriptors, resulting in a new feature set of low dimensionality. Similarity among keypoints in different images can be searched through these descriptor sets.

2.1 Introduction

Processing databases of video frames is an important task in computer vision, therefore describing local characteristics and registering image keypoints through this descriptor set is necessary. It might be very time consuming if the dimensionality of features is high. An early concept was to find specific keypoints of digital images but nowadays we use much more stable features, so called descriptors, to exploit a considerable amount of usable information from an image area. Although now we have very efficient local features, we should continuously search

for more applicable solutions. Task is twofold: features must describe the featuring patches at a high efficiency, while the dimensionality should be kept at a manageable low value.

Presently used descriptors can be grouped into different groups [26]. Distribution based descriptors are using histograms to represent different characteristics of appearance or shape. These histograms may based on purely intensity or a combination with distance or gradient distribution, like in scale invariant feature transform (**SIFT**). Spatial-frequency tehcniques describe frequency content of an image by decomposing the image with Fourier transform into basis functions. Differential descriptors computes a set of image derivatives up to a given order for approximating a point neighborhood.

The main assumption in finding local descriptors is to find unique and robust image features for finding and describing image characteristics. However, the dimensionality of these descriptors (e.g. SIFT) is quite large, resulting in time consuming search process for image/video database. A novel local feature detection method [21] proposed an approach for detecting similarity between a template and a given image by using regression kernels, measuring the likeness of a pixel to its neighborhood. The most distinctive features are filtered from this kernel using Principal Component Analysis (PCA). The extracted features are compared with the analogous features of the given image with Canonical Correlation Analysis (CCA), followed by a 'resemblance map' between the two images to produce the probability of similarity. Although this method models a local shape, it is rather interested in detecting a definite template object (e.g. face), and it is not for point pairing.

There are different dimension reduction techniques for data of SIFT or similar approaches, like PCA-SIFT [25] or Gradient Location and Orientation Histogram (GLOH) [26]. However, the valid interpretation of data content may be lost during compression. To avoid it, such methods are welcomed, where some formal meaning of the local properties can be maintained at a reduced dimension.

The proposed theory [4] is based on the principle that a saliency point (detected corner etc.) can be considered as a peak/valley concerning the intensity landscape. Different feature collection methods scan its neighborhood to describe

the microstructure somehow. The shape of this peak/valley may also be considered in a contour-map description, but in this low-resolution local neighborhood inside a radius of cc. 5–10 pixels a definite shape cannot be found. The idea here is to call for curve fitting methods for noisy shapes: active contours. The limited number of pixel values in a local neighborhood is to be surrounded by a curve. This curve might be distorted due to the transformed image contents, but its main characteristic can remain nearly constant through several frames of videos or through geometrical and illumination transformations of images from similar scenes.

This chapter is about the possible outcome of a hypothetic approach: can active contour be applied instead or together with other local featuring techniques for a better local description of image content? By examining here the possible solutions, regarding the local features, some measurements and adaptations are also given here about this additional approach.

The crucial step of any local featuring process is to find a good salient point. Several solutions are applied, like Harris [24] corner detection or Difference of Gaussians (DoG) [18], [19] in SIFT. If an efficient approach is used to extract the keypoints, the analytical question should be answered about how to characterize a point's neighborhood. As the hypothetical discussion only concentrates on this issue, the very efficient SIFT salient point detector is used for localizing the keypoints and giving a good peak/valley definition method as well.

Shape definition around a small salient region (centered by a keypoint) is not easy because of the limited neighborhood in the raw grid-resolution. Conventional shape definition and comparison methods suffer from the limited amount of information around the point. For example, in [23] the dissimilarity between two shapes is computed as a sum of matching errors between corresponding points, together with a term of measuring the magnitude of the aligning transform given by similar shape contents. Edge detection methods, like Canny edge detector [22] could be adaptable, but as we need a closed curve around the keypoint, active contour method is more efficient as an interpolation-based detector for the given regions.

2.2 Feature description

The main steps of the proof-of-concept algorithm for extracting local curve characteristics are the following:

1. Localizing keypoints as defined for SIFT [19].
2. Generating Local Contour [20] around the given keypoint.
3. Calculating the Fourier Descriptor [29] for the estimated closed curve.
4. Finding similar curves counting a limited set of components of Fourier Descriptors [27].

2.2.1 Scale invariant feature transform

Scale Invariant Feature Transform (SIFT) [18, 19] is a 128-dimensional image descriptor, which gives the opportunity for efficient image matching and view invariant visual object recognition, as it is invariant to scale, rotation, illumination and viewpoint. The method first detects interest points, then generates a descriptor of the local image structure by accumulating statistics of local gradient directions of image intensities. The high dimensionality represents large variation, therefore corresponding points can be efficiently matched using this descriptor between different images.

The four main steps of the descriptor extraction are the following:

1. scale-space extrema detection,
2. keypoint localization,
3. orientation assignment,
4. keypoint descriptor.

In the present approach, SIFT method is used only for localizing the keypoints, therefore just the first two steps are executed to extract the location of the keypoints. However, in the experimental part SIFT is applied for comparison, thus, the complete algorithm with all the four steps is presented in details.

2.2.1.1 Scale-space extrema detection

The first stage of the extraction attempts to find 'characteristic scale' for features, therefore the image is represented by a family of smoothed images known as scale space. The scale space is defined by the function:

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y), \tag{2.1}$$

where $*$ denotes the convolution operator, $I(x, y)$ is the input image and $G(x, y, \sigma)$ is the variable-scaled Gaussian:

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} \exp^{-\frac{x^2+y^2}{2\sigma^2}} . \tag{2.2}$$

Stable keypoint locations of the scale space are then detected by Difference of Gaussians (DoG) function, where $D(x, y, \sigma)$ function is given by computing the difference between two images, one with scale k times by the other:

$$\begin{aligned} D(x, y, \sigma) &= (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) \\ &= L(x, y, k\sigma) - L(x, y, \sigma). \end{aligned} \tag{2.3}$$

The $L(x, y, \sigma)$ convolved images are grouped by octave (an octave corresponds to doubling the value of σ) and k is selected so that a fixed number of convolved images per octave is obtained. DoG images are taken from adjacent Gaussian-blurred images per octave. Figure 2.1 shows the operation of this step. Extrema are then located by scanning each DoG image and identifying local minima and maxima. To detect such locations, each point is compared to its 8 neighbors on the same scale, and 9 neighbors on the higher and lower scale (see Figure 2.2). If the point is the local minimum/maximum of this 26 points, then it is an extremum.

2.2.1.2 Keypoint localization

After extracting extremum points, this step attempts to eliminate points having low contrast or being poorly localized along an edge. For the former case, the scale space value is used. The location of the extremum, $\hat{x}$, is determined by taking the derivative of the Taylor expansion (up to the quadratic terms) of the

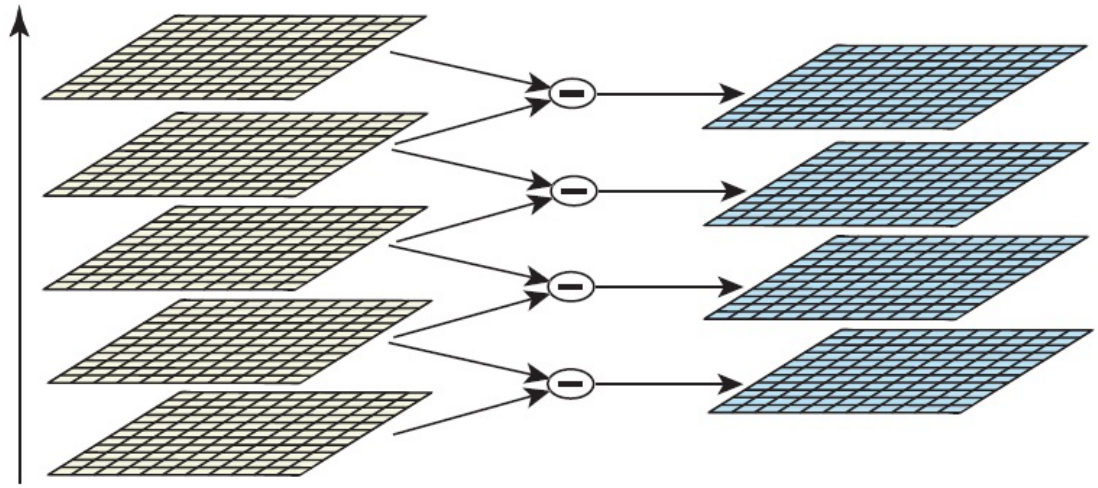


Figure 2.1: An octave of $L(x, y, \sigma)$ images and construction of DoG images [19].

D scale-space function (Eq. 2.3) with respect to $\mathbf{x} = (x, y, \sigma)^T$ and setting it to zero, giving:

$$\hat{\mathbf{x}} = -\frac{\partial^2 D}{\partial \mathbf{x}^2}^{-1} \frac{\partial D}{\partial \mathbf{x}}. \tag{2.4}$$

The function value at the extremum $D(\hat{\mathbf{x}})$ is investigated and eliminated if $|D(\hat{\mathbf{x}})|$ is less then 0.03 (assuming image pixel values in the range $[0, 1]$).

For the latter case (eliminating edge responses), observation about principal curvatures can be exploited: an edge point will have large principal curvature across the edge, but a small one in the perpendicular direction. Principal curvatures can be computed from the H Hessian matrix:

$$\mathbf{H} = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{bmatrix}. \tag{2.5}$$

Eigenvalues of $\mathbf{H}$ (denoted by α_1 and α_2, where $\alpha_1 = r\alpha_2$) are proportional to the principal curvatures of D. As only the ratio of the eigenvalues is required, not their exact values, motivated by [24], the trace (Tr) and determinant (Det) of H is used as follows:

$$\frac{Tr(\mathbf{H})^2}{Det(\mathbf{H})} = \frac{(\alpha_1 + \alpha_2)^2}{\alpha_1 \alpha_2} = \frac{(r\alpha_2 + \alpha_2)^2}{r\alpha_2^2} = \frac{(r+1)^2}{r}. \tag{2.6}$$

If this ratio is over some threshold, the point is taken as an edge point, therefore it is eliminated.

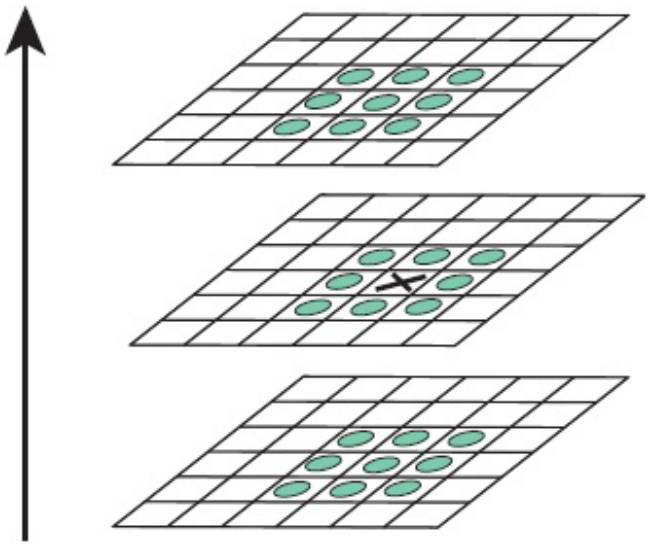


Figure 2.2: Extremum detection: each point is compared with its 26 neighbors on 3 different scales [19].

2.2.1.3 Orientation assignment

After extracting the keypoints, in this step a consistent orientation is assigned to each of them based on local image properties. The keypoint descriptor can later be represented relative to this orientation and achieve invariance to image rotation.

The scale of the keypoint is used to select the related smoothed image L with closest scale, therefore all computations are performed scale-invariantly. Magnitude $m(x, y)$ and orientation $\theta(x, y)$ are calculated using pixel differences:

$$m(x,y) = \sqrt{(L(x+1,y) - L(x-1,y))^2 + (L(x,y+1) - L(x,y-1))^2}, \quad (2.7)$$

$$\theta(x,y) = \tan^{-1}\left(\frac{L(x,y+1) - L(x,y-1)}{L(x+1,y) - L(x-1,y)}\right). \quad (2.8)$$

An orientation histogram with 36-bin is calculated from the gradient orientations of sample points within a region around the keypoint. The highest peak is located and used as the keypoint's orientation. If more local peaks exist with at least the 80% of height of the highest peak, then the orientation of these peaks

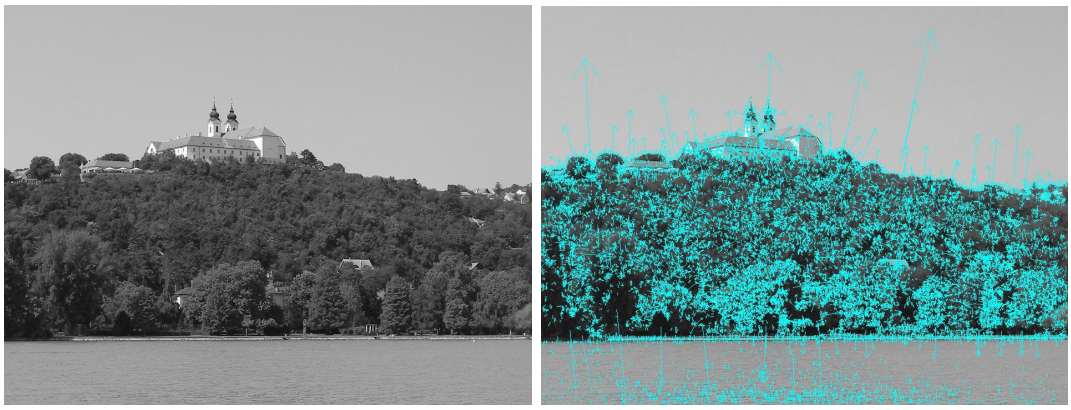


Figure 2.3: The original image is on the left, the calculated keypoints with the assigned orientation can be seen on the right.

are also assigned to the keypoint, resulting in multiple orientations. This step ensures invariance to image location, scale and rotation.

Figure 2.3 shows the result of orientation assignment. The cyan colored arrows indicate the orientation assigned to each keypoint. The size of the arrow is proportional to the magnitude.

2.2.1.4 Keypoint descriptor

The local gradient data, calculated in the previous step is also used when creating keypoint descriptors. Taking this data, the region of the keypoint is separated into subregions of 4×4 pixel neighborhood and an orientation histogram is calculated with 8 bins for each subregion (see Figure 2.4). This time the gradients are rotated by the previously computed orientation and weighted by a Gaussian function with variance of 1.5 times of the keypoint's scale. The descriptor is then becomes a vector of all the values of these histograms, resulting in $4 \times 4 \times 8 = 128$ dimensions. Finally the vector is normalized to unit length, ensuring the invariance to affine changes in illumination. For invariance to non-linear changes as well, the vector elements are capped to 0.2 and then the vector is renormalized.

The disadvantage of SIFT is its high dimensionality. However, this efficient data content cannot be compressed [25] without important information loss. Similar local descriptors [26] also give a batch of collected features. Some of them is

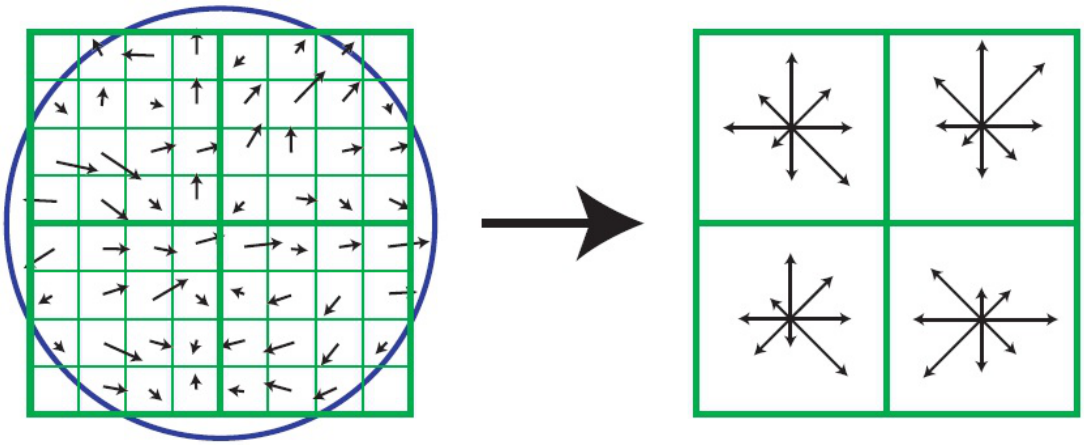


Figure 2.4: Keypoint descriptor extraction: The gradient histogram is on the left, the calculated 128-dimensional descriptor is on the right.

defined as scale-invariant features, where zoomed attributes describe larger scale connectivity (e.g. scaling in [19]).

2.2.2 Local contour

Instead of the batch of features given by SIFT, here the closed curves around localized keypoints are examined with an algorithm resulting in a meaningful low dimensional descriptor. It is used then for the selection of similar keypoints in different images.

The main idea for retrieving further description is analyzing the contour features in the small neighborhood of the keypoint. If significant features are extracted, these can be added later to any keypoint descriptor.

Local Contour (**LC**) is an active contour spline, generated in 11×11 size window with the keypoint in the middle. It is supposed to charactarize the local structure around the keypoint.

Introduced in [17], active contour (AC) is used in computer vision especially for locating object boundaries. The goal of active contour or snake (denoted by $\mathbf{x}(s) = [x(s), y(s)],\ s \in [0, 1]$) is to minimize the following energy:

$$E = \int_0^1 \frac{1}{2}(\alpha \left|\mathbf{x'}(s)\right|^2 + \beta \left|\mathbf{x''}(s)\right|^2) + E_{\text{ext}}(\mathbf{x}(s))\mathrm{d}s, \tag{2.9}$$

where α and β are weighting parameters for the elasticity and rigidity components of the internal energy; $\mathbf{x}'(s)$ and $\mathbf{x}''(s)$ are the first and second order derivatives with respect to s. The rigidity component is responsible for detecting curvature, setting $\beta = 0$ allows the snake to develop a corner. E_{ext} is the external energy derived from the image, representing the image constraints and giving smaller values at features of interest (like edges and ridges), than at homogeneous regions. The traditional Kass-version had limited utility as the initialization had to be near to the real contour of the object. Problems also occurred when detecting concave boundaries.

To compensate these drawbacks, Gradient Vector Flow (GVF) was a new external force used for snakes [20]. It is computed as a diffusion of the gradient vectors of a gray-level or binary edge map derived from the image. The novel E_{ext} energy is as follows:

$$E_{ext} = \int\int \mu(u_x^2 + u_y^2 + v_x^2 + v_y^2) + |\nabla f|^2 \, |v - \nabla f|^2 \, \mathrm{d}x\mathrm{d}y, \tag{2.10}$$

where the $\mathbf{v}(x, y) = (u(x, y), v(x, y))$ is the GVF field that minimizes E_{ext}, μ is a regularization parameter. The f edge map is derived from the image $I(x, y)$. One of the generally used forms is:

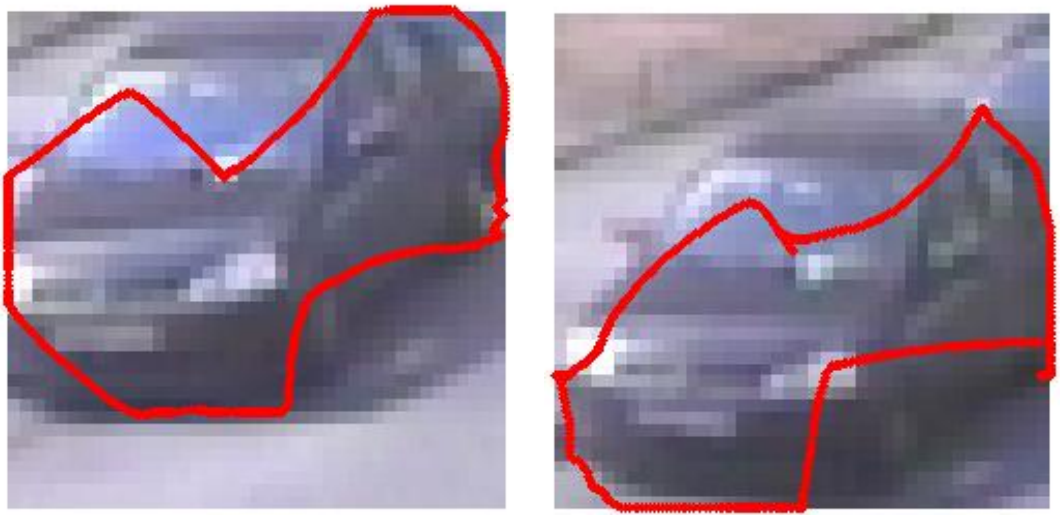
$$f(x, y) = |\nabla(G_\sigma(x, y) * I(x, y))| \, , \tag{2.11}$$

where G_σ is the Gaussian function with σ standard deviation and ∇ is the gradient operator [20].

The resultant field has a large capture range and forces active contours into concave regions. By using GVF, the initial shape of the snake is almost arbitrary.

Although several modification have been developed since the publication of GVF method, this approach was applied to test our theory. Therefore, the complete overview of the active contour theory or the operation of other algorithms is not explained here. Further contributions in the active contour field along with detailed descriptions are presented in the following chapter.

For initialization, a circle-shaped contour was used, where the center of the circle is the middle of the image and the radius is 3 pixel. After the convergence of the iterative process, the result is supposed to characterize the local structures around the keypoint, see Figure 2.5.



(a) (b)

Figure 2.5: Local contour result after the iterative process on different frames for a coherent point: The contour represents the local structure and preserves the main characteristics.

2.2.3 Dimension reduction and matching with Fourier descriptor

The result of the local contour detection is a point series describing the local shape, but the dimension is still high. A dimension reduction technique is needed, which is able to represent the contour efficiently at lower dimension. Fourier descriptors [28] are widely used for shape description, therefore in this step the local contours are represented by the modified Fourier descriptors [29] (MFD) that is invariant to translation, rotation and scaling of shapes. After calculating the low dimensional descriptors, the different images can be matched and compared through these descriptor sets.

The method first calculates the discrete Fourier transform (DFT) of this complex shape described by the point sequence, by measuring magnitude values of the DFT coefficients that are invariant to rotation. Denote the m^{th} pixel of the calculated LC by $(x(m), y(m))$, then a complex number can be formed as

Figure 2.6: Shape description with increasing number of Fourier descriptors [30].

$z[m] = x[m] + jy[m]$ the Fourier descriptor is defined as:

$$F[k] = DFT[z[m]] = \frac{1}{L}\sum_{m=0}^{L-1} z[m]\, e^{-j2\pi mk/L} \quad (k = 0, \ldots, L-1), \tag{2.12}$$

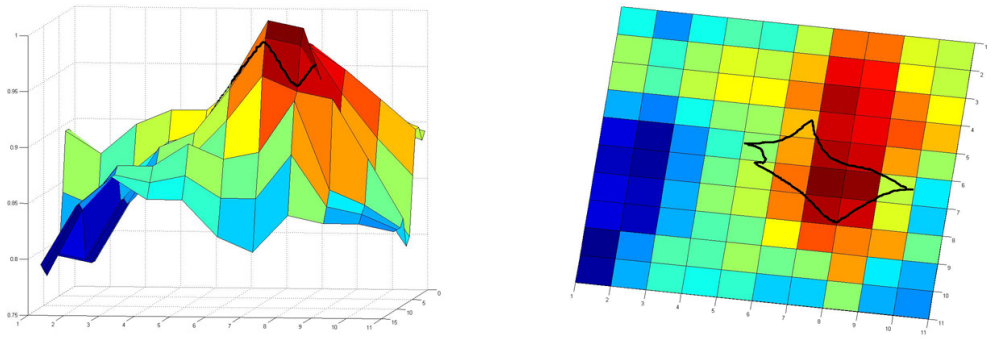
where L is the total number of pixels along the boundary.

If the aim is to reduce the dimension of the local contour descriptor (**LCD**), the first $\mathrm{n} < L$ coefficients $F[k]$ $k = 1, \ldots, n$ can be used to represent the local contour. The 0th component of the LCD is discarded to remove the positional sensitivity. The more coefficients used, the more accurate the local contour described (see Figure 2.6). According to the experiments $\mathrm{n} = 10$ was selected to maximize the recognition accuracies, but keeping the dimensionality low.

After having the LCDs for a keypoint, the possibly non-specific LCDs must be filtered out. Similar selection is also applicable in other descriptor methods, like in the keypoint localization step (see Section 2.2.1.2) of SIFT, where low contrasted and poorly localized extremum points are eliminated. The goal is to filter out the points having less characteristic neighborhood.

LCDs can be quite various and distinctive when searching in a large set of keypoints. On the contrary, when characteristic features cannot be found in the neighborhood, the LC remains similar to the initial configuration, therefore many of them can be easily mixed up when close to an average circle-contoured shape (Figure 2.7 and 2.8). Therefore, a prefilter must be used to exclude keypoints with less characteristic LCDs before the pairing process. Thus, a *Filter* function is defined to filter out the insignificant curves, which are close to the average meaningless cases:

$$Filter(F_{f,i}, \mathrm{n}) = \{Th(\frac{|F_{f,i,1}| - M_{f,1}}{V_{f,1}}) + ... + Th(\frac{|F_{f,i,\mathrm{n}}| - M_{f,\mathrm{n}}}{V_{f,\mathrm{n}}})\}, \tag{2.13}$$



(a) Contour in 3D (b) Contour from top view

Figure 2.7: Characteristic neighborhood and contour

where $F_{f,i}$ denotes the Fourier descriptor of the ith keypoint on frame f, having n coefficients, with $F_{f,i,1...n}$ denoting the DFT coefficients of the curve, $M_{f,j}$ and $V_{f,j}$ are the mean and the variance of the jth component on frame f and $Th(.)$ is a threshold function:

$$Th(x) = \begin{cases} 1 & if\ |x| > t, \\ 0 & otherwise. \end{cases}$$

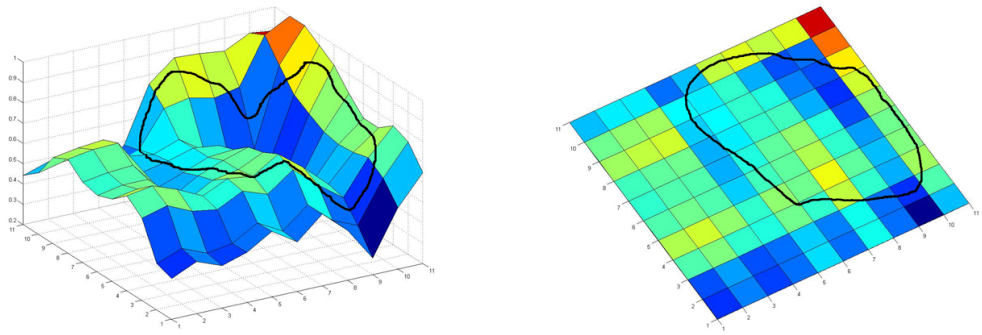
Choosing $t = 1.2$, the less significant keypoints having $Filter(F_{f,i}, \mathrm{n}) < 3$ are eliminated, resulting in a selection of the most representative points, about 30% of the initial set. The remaining point set with representative LCDs are describing an image, these LCDs have to be compared for image matching.

To compare LCDs for different images, symmetric distance computation is introduced and the MFD method has been extended in [27] by proposing the following distance function to compare different descriptors:

$$Dist\left(F_{1,i}, F_{2,i}, n\right) = \sigma\left(\frac{|F_{1,i,1}|}{|F_{2,i,1}|}, ..., \frac{|F_{1,i,\mathrm{n}}|}{|F_{2,i,\mathrm{n}}|}\right) + \sigma\left(\frac{|F_{2,i,1}|}{|F_{1,i,1}|}, ..., \frac{|F_{2,i,\mathrm{n}}|}{|F_{1,i,\mathrm{n}}|}\right), \quad (2.14)$$

where $F_{1,i,1...n}$ and $F_{2,i,1...n}$ denote the DFT coefficients of the compared curves for the ith keypoint on different frames, σ is the standard deviation function and n is the threshold, which defines the number of coefficients.

Therefore, a 10-dimensional LCD has been defined, which represents the local characteristics of a keypoint's neighborhood and a distance function was applied to compare the keypoint descriptors from different images.



(a) Contour in 3D (b) Contour from top view

Figure 2.8: Non-characteristic neighborhood and contour

2.2.4 Experiments

In the following section, the introduced LCDs are evaluated for two, different computer vision tasks. The first is the image matching, where sequential frames of a surveillance video was used, and point matching was performed by comparing the LCDs of the feature points. The second task was the texture classification, where the Brodatz textures were used and LCD method was compared with SIFT method. In this case, 10 training images were selected and the test image was classified after a k-means clustering of the descriptors of the feature points. By comparing the ratio of the points in a given cluster for the training images and the test image, the most similar was selected. As the size of the iterative pattern is varying, dynamic radius was introduced for LC calculation area. An adaptive r value was chosen, based on the variation of FD dimensions.

2.2.4.1 Image matching

Our method was tested on 22 real-life video frames made by an outdoor surveillance camera of a city police central. The algorithm was tested on sequential frame pairs. After extracting the significant keypoints, the LCDs are compared, the distance between them is measured by Equation 2.12. To evaluate results, the ground truth data of matching keypoint pairs can be defined by tracking or SIFT keypoint matching algorithm. Let $p_{1,i}$ denote the ith keypoint on Frame 1

Frame pairs	Nr. of pointpairs	Real pair is closest(%)	Real pair in 5 closest(%)
1-2	10	6 (60%)	8 (80%)
8-9	20	8 (40%)	14 (70%)
18-19	22	7 (32%)	16 (73%)

Table 2.1: The number of matches found in the images. Different columns show the different evaluation results: when the real pair was the closest measured LCD and pair was in the five closest LCDs.

and $p_{2,j}$ its pair on Frame 2, K is the number of significant keypoints on Frame 2. The results were evaluated in two ways:

- The real pair for $p_{1,i}$ is the closest LCD:

$$j = \underset{k\in[1,\dots,K]}{\operatorname{argmin}} \left(Dist(F_{1,i}, F_{2,k}, \mathrm{n})\right)$$

- The real pair is in the five closest LCDs.

As it can be seen in Table 2.1 the real pair was the best match approximately in the 40% of the cases (altogether in 21 out of the 52 cases), and in 73% of the pairs it was in the best five matches. A typical result is on Figure 2.9, where the detected keypoint pairs are shown on one frame. The original keypoint is in red, and its tracked (closest) pair is in blue.

2.2.4.2 Texture classification

The above results showed that LCDs can be comparable features against compressed descriptors. So, due to the promising result for image matching, the method was tested in another challenging field of image segmentation, the texture classification.

A well-known and widely-used dataset for texture classification is the Brodatz-textures, which consists of 112 different natural textures, like brickwall, grass, etc. This dataset was used to compare the SIFT method and the proposed LCD method for classifying textures. The major steps of the classification algorithm are the following:

1. Generating the descriptors (either SIFT or LC) for training and test images.

Figure 2.9: Keypoint pairs (red/blue) on the video frame, plotted to the first frame while the second moved

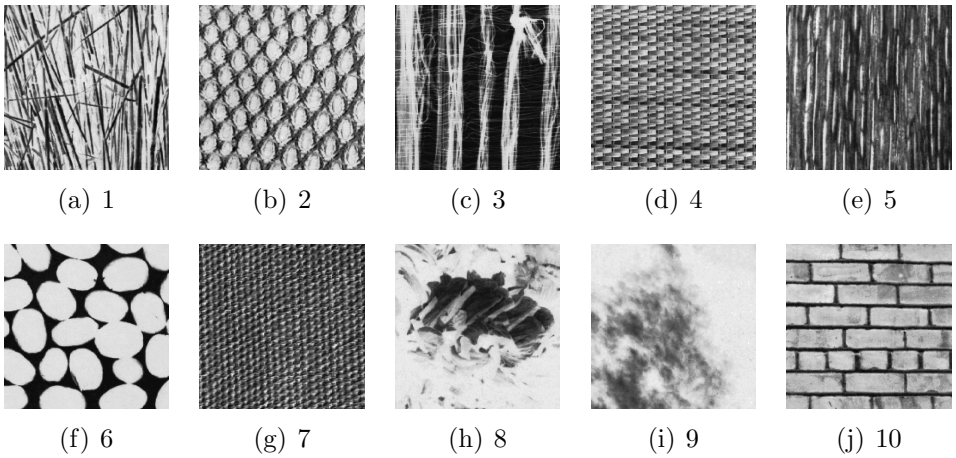
2. Clustering the descriptors of all images with k-means algorithm.

3. Classifying the test image according to the clustering results.

Both for SIFT and LCD, 10 different images for training and 5 images for testing were chosen. These images can be seen in Figure 2.10 and 2.11. In one iteration, all the 10 training images and one test image was used for comparison and classification.

The first step, in case of SIFT, was to generate the 128-long SIFT descriptors (Section 2.2.1). While in case of LCDs, the first three steps, mentioned in the beginning of Section 2.2, were executed.

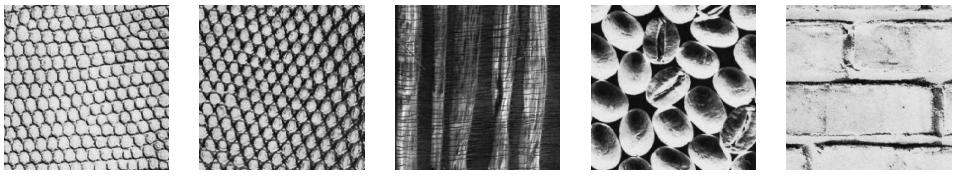
The second step was the clustering, where k-means algorithm, with $k = 15$ was used. The clustered results were summarized and normalized. Therefore, every image had a value for every cluster, which meant the ratio of the descriptors classified into the actual cluster, resulting in a descriptor set: $\{R(i,1), \ldots, R(i,k)\}$



(a) 1 (b) 2 (c) 3 (d) 4 (e) 5

(f) 6 (g) 7 (h) 8 (i) 9 (j) 10

Figure 2.10: Brodatz textures used for training



(a) 1 (b) 2 (c) 3 (d) 4 (e) 5

Figure 2.11: Brodatz textures used for testing

describing the distribution, where i is the image and k is the number of the clusters.

The last step was to compare the $\{R(i,.)\}$ descriptor set and classify the test image. The classification was based on the normalized results of the clustering. For each training image the $R(i,j)$ cluster ratio values were compared with the test image and the difference was summarized. In Equation 2.15, $\mathbf{Diff}(tr,te,j)$ means the difference of training image tr, test image te for the cluster j. Therefore the total difference:

$$\mathbf{Diff}(tr,te) = \sum_{j=1}^{k} \mathbf{Diff}(tr,te,j) = \sum_{j=1}^{k} |R(tr,j) - R(te,j)| \qquad (2.15)$$

After counting the difference for all training images, the lower value means the higher similarity.

Test image	Pair	SIFT	LCD with SR(r=20)	LCD with DR(r)
1	2	4 (80%)	4 (80%)	5 (100%) (20)
2	2	5 (100%)	4 (80%)	4 (80%) (20)
3	3	5 (100%)	5 (100%)	5 (100%) (20)
4	6	0 (0%)	0 (0%)	5 (100%) (25)
5	10	5 (100%)	0 (0%)	5 (100%) (30)

Table 2.2: Result of texture classification: the number and rate of the correct classifications

Results of classification iterations can be seen in Table 2.2. Every test image was classified 5 times. The number of correct classifications are summarized in the table for each case. The number of the most similar pair (used as ground truth) from the training images is also indicated. The third column contains the classification results for the SIFT method.

The fourth column represents the results of the LCD method, where LCDs generated in a static 20×20-size neighborhood of the keypoint (called as static radius (**SR**)). The 4th and 5th test images showed poor results. In these two cases the iterative pattern is larger, than 20×20, so the fixed, $r = 20$ radius is too small and should be extended. To solve this problem, dynamic radius (**DR**) was introduced. DR is counted based on the variation of the FDs. To test this variation, LC and then FD is calculated (see Eq. 2.9 and 2.12) in the $r \times r$ neighborhood around the feature point. If the iterative pattern is large, extending the radius results in the variance of the main FD components. This means, the contour of the pattern is more characteristic. Similarly, if the iterative pattern is small, extending the radius has no effect on the FDs, as the contour of the pattern remains nearly constant. Therefore, the radius should be extended until the variance of the main FD components is not exceeding a threshold and remaining quasi-constant. The last column shows the results of the LCD method with DR. By using the DR, results for the 4th and 5th test images are improving significantly. The calculated DR is also represented for each case.

2.3 Detection of structural changes in long time-span aerial image samples

This section presents an application of local contour descriptors (introduced in the previous section) to find changes in remote sensing image series [6]. Some remotely sensed areas are scanned frequently to spot relevant changes, and several repositories contain multi-temporal image samples from the same area. The proposed method finds changes in images scanned by a long time interval in very different lighting and surface conditions. The presented method is basically an exploitation of Harris saliency function and its derivatives for finding feature points among image samples. In the first step, the Harris corner detection is introduced for difference extraction and localization of change keypoint candidates, instead of the simple edge functions. Change keypoints are then filtered based on local contour descriptors, by comparing the changes in the neighborhood between the older and newer image. The boundary hull of changing objects is estimated by fusing the connectivity information and the change keypoint set with a graph-inspired technique. The method is evaluated on registered remote sensing image pairs taken in 2000 and 2005, therefore having different color and illumination features.

2.3.1 Motivation and related works

Automatic evaluation of aerial photograph repositories is an important field of research since manual administration is time consuming and cumbersome. Long time-span surveillance or recognition about the same area can be crucial for quick and up-to-date content retrieval. The automatic extraction of changes may facilitate applications like urban development analysis, disaster protection, agricultural monitoring, and detection of illegal garbage heaps, or wood cuttings. The obtained change map should provide useful information about size, shape, or quantity of the changed areas, which could be applied directly by higher level object analyzer modules [31], [32]. While numerous state-of-the art approaches in remote sensing deal with multispectral [33], [34], [35], [36] or synthetic aperture radar (SAR) [37], [38] imagery, the significance of handling optical photographs is also increasing [39]. Here, the processing methods should consider that several

optical image collections include partially archive data, where the photographs are either grayscale or contain only poor color information. This section focuses on finding contours of newly appearing/fading out objects in optical aerial images which were taken with several years time differences, in different seasons and in different lighting conditions. In this case, simple techniques like thresholding the difference image [41] or background modeling [42] cannot be adapted efficiently since details are not comparable.

In the literature one main group of approaches is the postclassification comparison, which segments the input images with different land-cover classes, like woodlands, barren lands, and artificial structures [43], obtaining the changes indirectly as regions with different classes in the two image layers [39]. The proposed approach follows another methodology, like direct methods [33], [35], [37], where a similarity-feature map from the input photographs (e.g. a difference image) is derived, then the feature map is separated into changed and unchanged areas.

This direct method does not use any land-cover class models, and attempts to detect changes which can be discriminated by low-level features. However, this approach is not a pixel-neighborhood maximum a priori probability system as in [40], but a connection system of nearby saliency points. These saliency points define connectivity relations by using local graphs for estimating the outline of the objects. Considering this estimated polygon as a starting spline, the method searches for object boundaries by active contour iterations (See Figure 2.12).

The main saliency detector is calculated as a distinguishing metric between the functions of the different layers to extract saliency point set. Harris detector is proved to be an appropriate function for finding the dissimilarities among different layers and handling the varying illuminational circumstances, when comparison of other low-level features is not possible because of the different lighting, color and contrast conditions.

Local structure around keypoints is investigated by local contour descriptors (see Section 2.2.3), representing the local microstructure around keypoints. However, the local contour is generated by edginess in the cost function (gradient information in the external energy), while keypoints of junctions (corners) are characterized in the localization step. To fit together the definition of keypoints and their active contour around them, Harris corner detection was applied as an

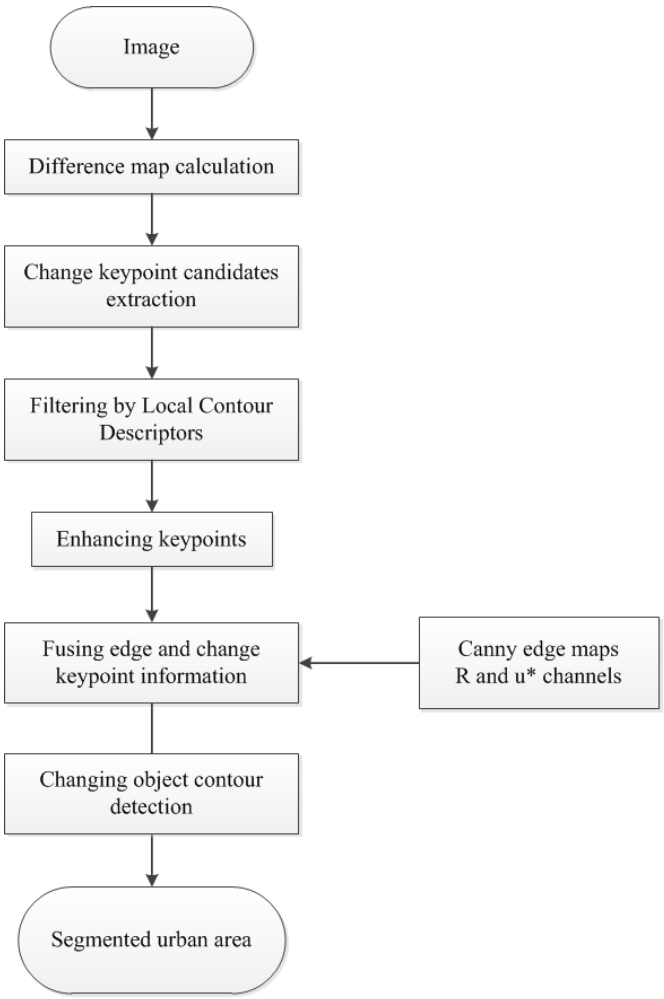


Figure 2.12: Simplified diagram of the workflow of change detection process in aerial images.

outline detector instead of the simple edge functions. This change resulted in a much better characterization of the local structure.

2.3.2 Change detection with Harris keypoints

In case of long time-span aerial image pairs, many changes might have happened during the elapsed time between the scanning of images. Therefore, the image content might have changed significantly. Moreover, the altering weather conditions (different illumination angle, season, etc.) cause varying color, contrast and

shadow information. Due to these differences, SIFT cannot be used for image matching and keypoint localization, unlike in Section 2.2.1.2.

When localizing changes, a robust method is needed which is able to cope with the varying illumination and emphasize the structural changes at the same time. As the available aerial image pairs are already registered manually by the Hungarian Institute of Geodesy, Cartography and Remote Sensing (FÖMI), the idea was to generate a difference image based on some metric, emphasizing changes between images. The applied metric was the Harris corner detector's characteristic function [24], which is proved to be fairly invariant to illumination variation and image noise, moreover the detector itself is reliable and invariant to rotation [44].

2.3.2.1 Harris corner detector

The Harris detector [24] is based on the principle that at corner points intensity values change largely in multiple directions. Change D for a small (x, y) shift is given by the following Taylor expansion:

$$D(x, y) = Ax^2 + 2Cxy + By^2, \tag{2.16}$$

which can be rewritten as

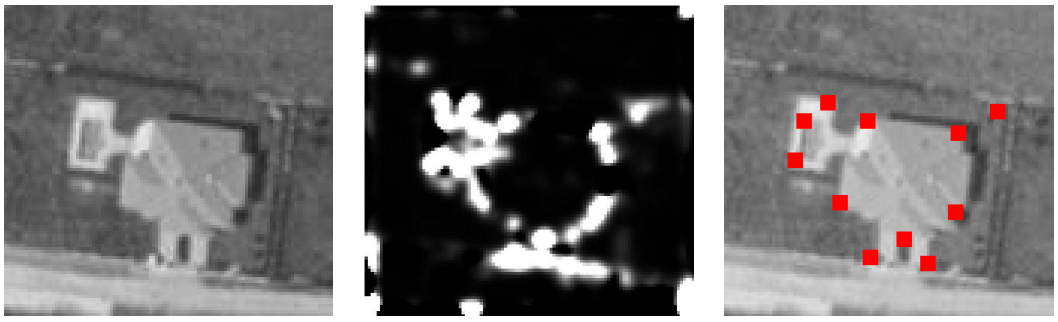
$$D(x, y) = (x, y)M(x, y)^T. \tag{2.17}$$

Here, M is the Harris matrix:

$$M = \begin{bmatrix} A & C \\ C & B \end{bmatrix}, \tag{2.18}$$

where $A = \dot{x}^2 * w$, $B = \dot{y}^2 * w$, $C = \dot{x}\dot{y} * w$. $\dot{x} = \frac{\partial I}{\partial x}$ and $\dot{y} = \frac{\partial I}{\partial y}$ denote the approximation of the first order derivatives of the I image, $*$ is a convolution operator and w is a Gaussian window.

The curvature behavior around an image point can be well described by the Taylor expansion (Eq. 2.16). When D is reformulated by a structure tensor (Eq. 2.17) and becomes closely related to the local autocorrelation function, M describes the shape at the image point. The eigenvalues of M will be proportional to the principal curvatures of the local autocorrelation function and form a



(a) Original (b) R function (c) Keypoints

Figure 2.13: Operation of Harris detector: Corner points are chosen as the local maxima of the R characteristic function

rotationally invariant description of M. This feature is beneficial when external forces are calculated to measure salient points of any boundaries, as the principal curvatures describe well the fine details of shapes.

The Harris corner detector defines the following corner response to select isolated corner pixels:

$$R = \mathrm{Det}(M) - k * \mathrm{Tr}^2(M), \tag{2.19}$$

where Det and Tr denote the determinant and trace and k is a coefficient, usually around 0.04. R is large and positive in corner regions, negative in edge regions and small in the flat regions. By searching for local maxima of R, the Harris keypoints can be found. Figure 2.13 shows the result of Harris keypoint detection. On Figure 2.13(b) light regions shows the larger R values, so keypoints will be detected in these areas (Figure 2.13(c)).

2.3.2.2 Difference image and change candidates calculation

The advantage of Harris detector is its strong invariance to rotation and the R characteristic function's invariance to illumination variation and image noise. Therefore it could be used efficiently for change detection in airborne images. In these kind of images, changes may mean the appearance of new man-made objects, (like buildings or streets), or natural, environmental variations. As image pairs may be taken with large intervals of time, the area might have changed

(a) Older image from 2000.

(b) Newer image from 2005.

Figure 2.14: Original image pairs provided by the Hungarian Institute of Geodesy, Cartography and Remote Sensing (FÖMI).

significantly. In this case the pieces of the image pairs were taken with 5 years difference (2000 and 2005) and registered manually by FÖMI (see Figure 2.14).

In this work the focus is on the detection of newly built objects (buildings, pools, etc.). There are many difficulties when detecting such objects in airborne images

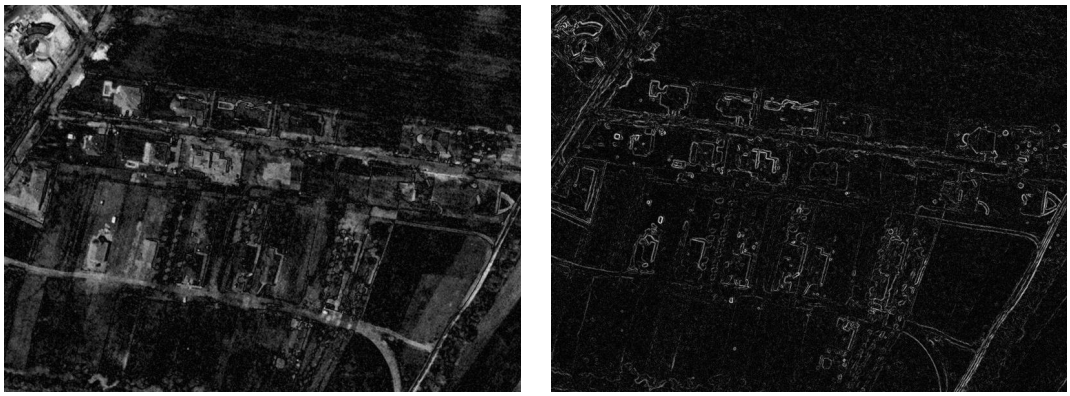
1. The illumination and weather circumstances may vary, resulting in different color, contrast and shadow conditions.

2. The urban area might be imaged from different point of view.

3. Buildings can be hidden by other structures, like trees, shadows or other buildings.

4. The changing objects have quite various shapes, which also makes the detection tough.

To overcome a part of these difficulties, the idea was to use the difference of the image pairs. When searching for newly built objects, the aim is to find buildups, that only exist on the newer image, therefore having large effect both in the difference image and the newer image.



(a) Intensity based difference map (b) Edge based difference map

Figure 2.15: Difference maps calculated based on traditional metrics

First, the usability of traditional metrics was examined, which are usually applied for difference image calculation, like intensity based (Figure 2.15(a)) and edge based difference map (Figure 2.15(b)).

Intensity based and edge based difference maps are calculated as follows:

$$I_{\text{diff}} = |I_{\text{old}} - I_{\text{new}}|, \tag{2.20}$$

$$I^{\text{edge}}_{\text{diff}} = |I^{\text{edge}}_{\text{old}} - I^{\text{edge}}_{\text{new}}|, \tag{2.21}$$

where I_{old} and I_{new} means the older and newer pieces of the image pairs respectively and $I^{\text{edge}} = |\nabla I|$ is the gradient magnitude of the image. The calculated intensity and edge based difference images are shown in Figure 2.15. This figure also shows the drawbacks of such metrics, being too sensitive to illumination changes, thus appearing a lot of false edges. Performing keypoint candidate detection resulted in extracting a lot of false corner points, due to the appearance of false edges in the difference image, causing the unusability of intensity and edginess for constructing difference image.

Therefore, another metric is used instead of intensity and edginess and redefine the difference map according to the new metric. The chosen metric was the Harris R characteristic function, because of the aforementioned advantages and the difference map was calculated as:

$$R_{\text{diff}} = |R_{\text{old}} - R_{\text{new}}|, \tag{2.22}$$

where R is the Harris characteristic function (see Eq. 2.19) calculated for the older and newer image as well. The logarithm of R_{diff} difference map is in Figure 2.16(a), as R-function has a wide range of values, the dynamics of the characteristic function is compressed into a balanced distribution.

For extracting change keypoint candidates, local maxima in R_{diff} and R_{new} simultaneously are detected. A pixel $p_i = (x_i, y_i)$ is the element of the P change keypoint candidate set, if it has the largest $R_{\text{diff}}(p_i)$ and $R_{\text{new}}(p_i)$ values 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 $R_{\text{diff}}(p_i)$ value exceeds a given ϵ_1, $R_{\text{new}}(p_i)$ value exceeds a given ϵ_2 threshold:

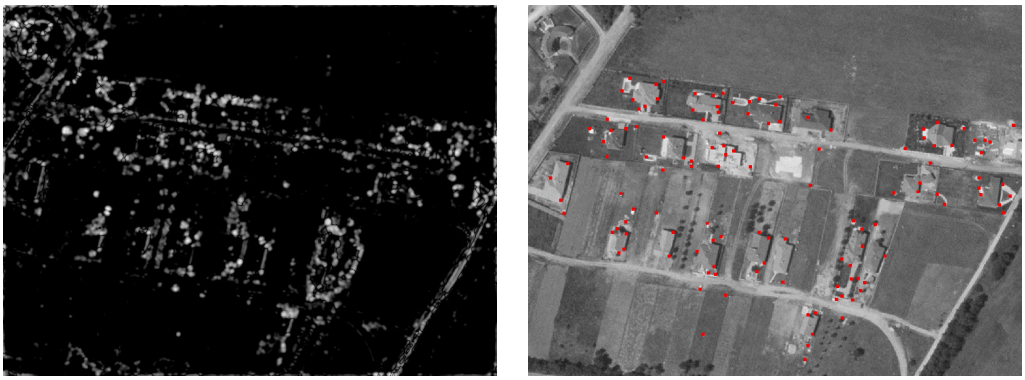
$$\begin{aligned} P = \Big\{ & p_i : R_{\text{diff}}(p_i) > \epsilon_1 \text{ AND } p_i : R_{\text{new}}(p_i) > \epsilon_2 \text{ AND} \\ & p_i = \underset{r \in b_i}{\operatorname{argmax}} R_{\text{diff}}(r) \text{ AND } p_i = \underset{r \in b_i}{\operatorname{argmax}} R_{\text{new}}(r) \Big\}. \end{aligned} \tag{2.23}$$

The ϵ_1 and ϵ_2 thresholds are chosen empirically. It is advised to take higher ϵ_1, than ϵ_2. With this choice the difference map is preferred and has larger weight. Only important corners in the difference map will be marked. Detected change keypoint candidates are in Figure 2.16(b). Keypoint candidates cover all buildings, and only a few points are in false areas. The false candidates have to be filtered out with further techniques, described in the next section.

2.3.2.3 Filtering with local contour descriptors

The assumption in Section 2.2.3 was that after having the local contour descriptors (LCDs) for the keypoints, differences between keypoint surroundings in I_{old} and I_{new} can be searched through this descriptor set and keypoints indicating noise or illumination changes can be filtered out. When generating local contours, the original intensity based image used in the active contour external force (Eq. 2.11) is sensitive to illumination changes. To compensate this drawback, the calculated R characteristic function was applied for contour extraction, the novel feature map looks as follows:

$$f_{|R|}(x, y) = G_\sigma(x, y) * |R(x, y)| \,. \tag{2.24}$$



(a) R-based difference map (b) Keypoint detection

Figure 2.16: Logarithmized difference map and result of change keypoint candidate detection based on the R-function. Detected change keypoints are marked in red.

Detected contours are more stable in case of the $|R|$ function. We benefit from this stability, as contours can be distinguished easier. To extract more information about the characteristics in the keypoint's neighborhood, the contours were calculated in 20×20 window around the keypoint.

The calculated LCs are then represented by LCDs (the method is briefly described in Section 2.2.3), but using the first twenty coefficients of MFD (excluding the DC component) to describe local characteristics more precisely. As the aim of this step is to detect change keypoints, points with further distance between R_{old} and R_{new} are preferred, which means that the point is filtered out if the following condition is not satisfied:

$$Dist\left(F_{\text{old}}, F_{\text{new}}, 20\right) > 3, \tag{2.25}$$

where the definition of $Dist$ is in Eq. 2.14.

The remaining change keypoints after the LCD-based filtering can be seen in Figure 2.17.

2.3.2.4 Enhancing the number of saliency points

As the aim is to outline newly appearing/ fading out objects, the number of keypoints localized on contours, has to be enhanced.

Figure 2.17: Remaining change keypoints after filtering with local contour descriptors.

Therefore, such saliency points are searched which are only present on the newer image and located on a contour in the close proximity of the change keypoints. With this step, information about the objects around the keypoints is extracted to outline structural changes.

Harris corner points are used for the task. By calculating corner points for older and newer image as well, an arbitrary $q_i = (x_i, y_i)$ point is selected if it satisfies all of the following conditions:

1. $q_i \in H_{\text{new}}$,
2. $q_i \notin H_{\text{old}}$,
3. $d(q_i, p_j) < \epsilon_3$.

H_{new} and H_{old} are the sets of Harris keypoints generated in the newer and older image with half of thresholds used in Eq. 2.23, $d(q_i, p_j)$ is the Euclidean distance of q_i and p_j, where p_j denotes the point with smallest Euclidean-distance to q_i selected from P.

New points are searched iteratively, with $\epsilon_3 = 10$ condition. Here, ϵ_3 depends on the resolution of the image and on the size of buildings. In case of the given ϵ_3, the resolution was 0.5 m/pixel. If resolution is smaller, than ϵ_3 has to be chosen as a smaller value.

Figure 2.18 shows the enhanced number of keypoints.

Figure 2.18: Enhanced number of Harris keypoints

2.3.2.5 Fusion of edge and change keypoint information

Now an enhanced set of salient points is given, representing possible areas of changes, which serves as the basis for building detection. Now, redefine the problem in terms of graph theory [67]. A graph G is represented as $G = (V, E)$, where V is the vertex set, E is the edge network. In this case, V is already defined by the enhanced set of Harris points. Therefore, E needs to be formed.

Information about how to link the vertices can be gained from edge maps. These maps can help to discover connectivity relations and only match vertices belonging to the same building.

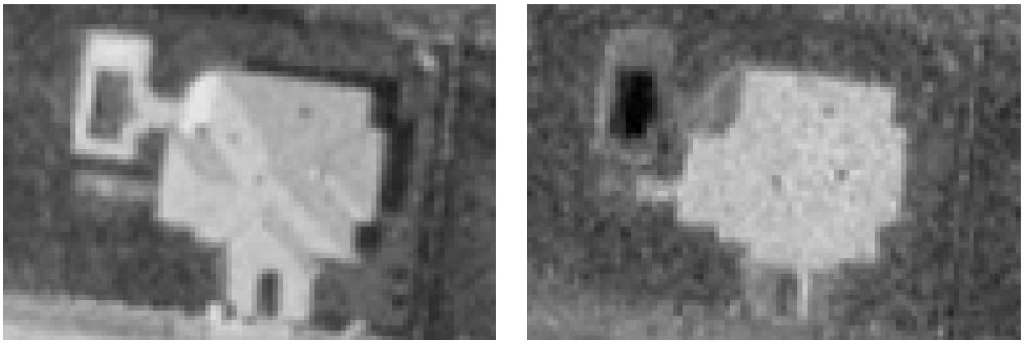


Figure 2.19: Grayscale images generated two different ways: (a) is the R component of the RGB colorspace, (b) is the u^* component of the $L^*u^*v^*$ colorspace.

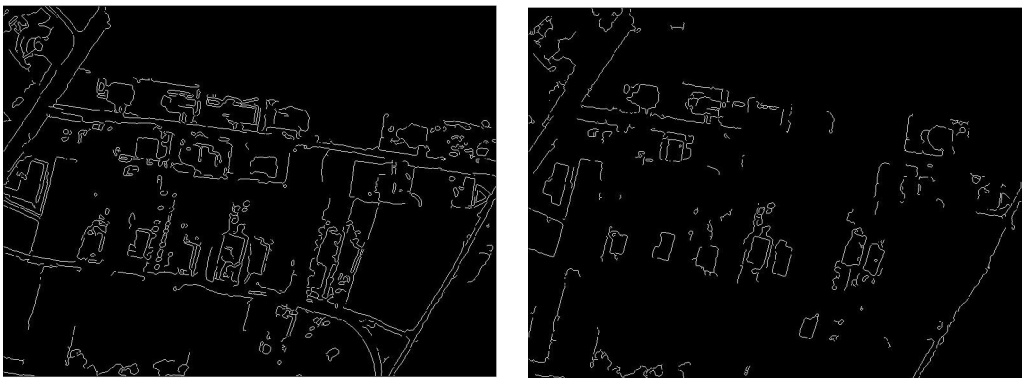
If objects have sharp edges, such image modulations are needed, which emphasize these edges as strong as possible. Figure 2.19(a) and 2.19(b) show that R component of RGB and u^* component of $L^*u^*v^*$ colorspace can intensify building contours [74]. Both of them operates suitably in different cases, therefore we apply both.

By generating the R and u^* components (further on denoted as $I_{\text{new},r}$ and $I_{\text{new},u}$) of the original, newer image, Canny edge detection [22] with large threshold ($Thr = 0.4$) is executed on them. $C_{\text{new},r}$ and $C_{\text{new},u}$ marks the result of Canny detection (Figure 2.20(a) and 2.20(b)).

The process of matching is as follows. Given two vertices: $v_i = (x_i, y_i)$ and $v_j = (x_j, y_j)$. We match them if they satisfy the following conditions:

1. $d(v_i, v_j) = \sqrt{(x_j - x_i)^2 + (y_j - y_i)^2} < \epsilon_4$,
2. $C_{\text{new},\cdot}(x_i, y_i) = \text{true}$,
3. $C_{\text{new},\cdot}(x_j, y_j) = \text{true}$,
4. a finite path exists between v_i and v_j in $C_{\text{new},\cdot}$.

$C_{\text{new},\cdot}$ indicates either $C_{\text{new},r}$ or $C_{\text{new},u}$. ϵ_4 is a tolerance value, which depends on the resolution and average size of the objects. We apply $\epsilon_4 = 30$.



(a) (b)

Figure 2.20: Result of Canny edge detection on different colour components: (a) is for R component of RGB space; (b) is for u^* component of $L^*u^*v^*$ space.

Figure 2.21: Subgraphs given after matching procedure, edges between connected keypoints are shown in white.

These conditions guarantee that only vertices connected in the newer edge map are matched. Like in the lower right part of Figure 2.21 two closely located buildings are separated correctly.

After this procedure a graph composed of many separate subgraphs is obtained, which can be seen in Figure 2.21. Each of these connected subgraph is supposed to represent a building. However, there might be some unmatched keypoints, indicating noise. To discard them, we select subgraphs having at least two vertices.

To determine the contour of the subgraph-represented buildings, the aforementioned GVF snake method (see Section 2.2.2) was applied. The convex hull of the vertices in the subgraphs is applied as the initial contour. (Further discussion about the calculation of the convex hull of a point set is in Section 3.3.2.) Result of the contour detection for the image pair in Figure 2.14 can be seen in Figure 2.22. Further test results are in Figure 2.23 and 2.24.

Figure 2.22: Results of the structural change detection method. Different images shows the result for different building outlines for image pair Figure 2.14.

(a) Original image (b) Detected changes

Figure 2.23: Result of the contour detection for aerial image pair provided by FÖMI.

2.4 Conclusion

This chapter examined the possible outcome of a hypothetic approach, if active contour could be applied instead or together with other local featuring techniques for a better local description of image content. The experiments verified that local structures around the keypoint (local contour descriptors) can be used as an additional feature set for characterizing the neighborhood of a keypoint to register image regions. Moreover, local contour descriptors can be comparable features against compressed descriptors, while the meaningful interpretation can help to design better keypoint descriptors. To show the efficiency of the proposed feature set, an application was presented to detect structural changes between aerial image pairs based on local contour descriptors.

(a) Original image

(b) Detected changes

Figure 2.24: Result of the contour detection for aerial image pair provided by FÖMI.

Chapter 3

Harris Function Based Feature Map for Image Segmentation

Deformable active contour (snake) models are efficient tools for object boundary detection. Existing alterations of the traditional gradient vector flow (GVF) model have reduced sensitivity to noise, parameters and initial location, but high curvatures and noisy, weakly contrasted boundaries cause difficulties for them.

This chapter introduces a Harris based feature map [1, 5, 10], which is applied in the external force of two parametric snake models, Harris based Gradient Vector Flow (HGVF) and Harris based Vector Field Convolution (HVFC). The introduced feature map uses the curvature-sensitive Harris matrix to achieve a balanced, twin-functionality (corner and edge) map. To avoid initial location sensitivity, starting contour is defined as the convex hull of the most attractive points of the map. In the experimental part the introduced method is compared to the traditional external energy-inspired state-of-the-art GVF and VFC; the recently published parametric Decoupled Active Contour (DAC) and the non-parametric Chan-Vese (ACWE) techniques. Results show that the improved methods outperform the classical approaches, when tested on images with high curvature, noisy boundaries.

Moreover, the introduced map and the feature point set - calculated as local maxima of the map-, are also used efficiently in different, complex object detection applications: building localization and outline extraction in aerial images [8]; automatic detection of structural changes (e.g. multiple sclerosis lesions) in single

channel long time-span brain MRI image pairs [12]; and localization of small, flying targets [2, 9].

3.1 Motivation and related works

Object boundary detection is an important field of vision research. The active contour (AC) method (or also called as snake) was introduced in [17], since then deformable models proved to be efficient tools for robust identification of object contours [20, 45, 46, 47, 48, 49, 50, 51, 52]. Snake evolution is controlled by an energy minimizing method based on different energies. Internal energy is responsible for obtaining elastic and rigid curves, while external energy represents the constraints of the image itself and is usually calculated as a function of gradient information over the intensity distribution. This force pushes the snake toward an optimum in the feature space. The traditional snake model has limited utility as the initialization should be close to the real contour of the object. Problems also occur when detecting concave boundaries. To compensate these drawbacks, Gradient Vector Flow (GVF) snake was introduced in [20], which defined a new external force as a diffusion of the gradient vectors of a gray-level or binary edge map derived from the image. Although precision improved, GVF snake was still noise, parameter and initialization sensitive.

Since the publication of the original method [17], several modification have been developed to compensate the drawbacks of the original algorithm, including parametric [20, 46, 51, 52] and non-parametric [45, 47, 48, 49, 50] approaches.

Parametric active contours suffer from weaknesses associated with noise, parameter and initialization sensitivity, topology changes and have difficulties when detecting high curvature boundaries. While non-parametric methods do not depend on initialization and detect complex boundaries with sharp corners and topological variations, they fail when detecting objects with broken edges. Additionally their convergence rate is slower and they are more sensitive to noise than the parametric approaches. Application of non-parametric techniques to images of narrow elongated structures, where intensity contrast may be low and reliable region statistics cannot be computed was independently improved by [48] and

[49]; additionally when using shape priors as [53], non-parametric methods can also cope with broken edges.

One class of the parametric methods tries to redefine the expression of external energy to improve the accuracy of GVF snake [60, 59, 51, 58, 57, 54]. While these approaches reduce the sensitivity in some aspects, they still have difficulties when featuring very sharp and noisy corners. These high curvature, noisy boundary points along with noisy edges are still among the major challenges that existing methods are not able to handle appropriately.

Curve initialization is a challenging task, some representations take shape information into account [55] or extract the focus area to define the region of interest [56], but in case of the detection of randomly shaped objects, the initial outline is usually defined with human interaction. Recently published quasi-automatic method, [61] requires the selection of an arbitrary point in the target region to initialize the curve, but it is not able to segment regions that feature topological changes.

To address the limitation of initialization and curvature sensitivity, this chapter proposes two parametric active contour approaches, introduced as the Harris based GVF (HGVF) snake [5] and the Harris based Vector Field Convolution (HVFC) [10], both use a modified function of Harris corner detector [24] that benefits from the cornerness feature, therefore, it is suitable for emphasizing both corner points and edges, and attains a balanced feature map. The most attractive points of the map are used to initialize a starting curve around the object, while the modified map is applied to determine a new feature map for the external energy expression.

In the experimental part, the performance of the proposed methods have been evaluated on the Weizmann segmentation database [62] and the results have been compared to published techniques, including two external energy-inspired parametric algorithms [20], [51]; a novel parametric method [52] and a non-parametric, region based [47] application. According to the evaluation results, the proposed algorithms perform better in detecting high curvature, noisy object boundaries.

3.2 Active contour model

Section 2.2.2 gave an overview of the traditional active contour theory, introduced in [17]. The energy minimizing method is driven by internal and external energy, representing the snake properties and the image features.

3.2.1 Gradient vector flow

To compensate the drawbacks of the traditional method, [20] defined Gradient Vector Flow (GVF) as external force for active contour methods. The increased capture range of the external force field guides the contour toward the boundary, even into concave regions. The novel E_{ext} energy is as follows:

$$E_{ext} = \int\int \mu(u_x^2 + u_y^2 + v_x^2 + v_y^2) + |\nabla f|^2 \, |v - \nabla f|^2 \, \mathrm{d}x\mathrm{d}y, \tag{3.1}$$

This chapter concentrates on the f edge map in the external energy, which is derived from the image $I(x, y)$. One of the generally used forms is:

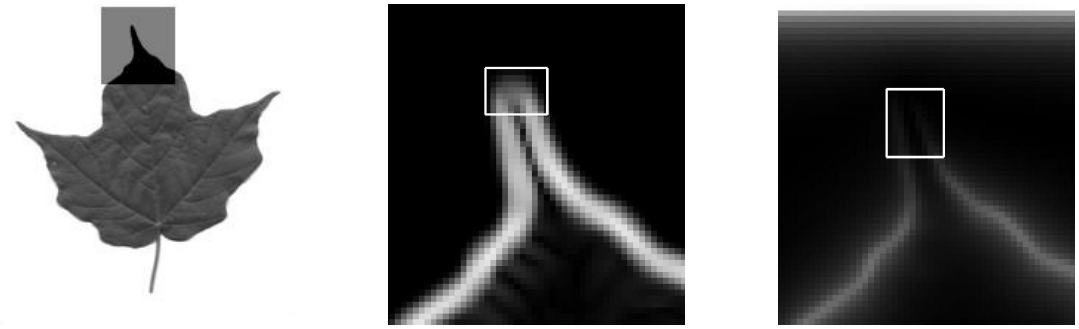
$$f(x, y) = |\nabla(G_\sigma(x, y) * I(x, y))| \, , \tag{3.2}$$

with G_σ denoting the Gaussian function with σ standard deviation and ∇ is the gradient operator.

The original f edge map has difficulties when detecting sharp or noisy corners and low contrast boundaries. When heading towards a sharp corner point in the image, the intensity changes largely from only a few directions. For this reason, these pixels behave similarly like low contrast boundary points: they both have decreased $f(x, y)$ values compared to sharp edges. This effect is shown in Figure 3.1, where decreased f values can be seen around the peak of the leaf inside the white rectangle in Figure 3.1(b). The result of this effect is that the iterative AC method does not detect the peaks accurately (see the third image in the first row of Figure 3.4).

3.2.2 Vector field convolution

Vector Field Convolution (VFC) was introduced in [51] as external force, addressing the disadvantages of GVF, such as high computational cost, noise sensitivity,



(a) (b) (c)

Figure 3.1: The original f edge map for high curvature boundary: (a) shows the original image, region of interest is in gray; (b) is the f edge map of the marked area. The white rectangle indicates the decreased f values of the high curvature boundary; (c) is the $\mathbf{f}_{\mathrm{VFC}}$ map of the marked area. The white rectangle indicates the decreased $\mathbf{f}_{\mathrm{VFC}}$ values of the high curvature boundary.

parameter sensitivity, and the ambiguous relationship between the capture range and parameters. The VFC field is calculated by convolving a vector field kernel with the edge map generated from the image:

$$\mathbf{f}_{\mathrm{VFC}}(x,y) = f(x,y) * \mathbf{k}(x,y), \tag{3.3}$$

where $f(x,y)$ is the original, image based edge map (Eq. 3.2) and $\mathbf{k}(x,y)$ is the vector field kernel, which is defined as:

$$\mathbf{k}(x,y) = m(x,y)\mathbf{n}(x,y), \tag{3.4}$$

where $m(x,y)$ is the magnitude at (x,y) and $\mathbf{n}(x,y)$ is the unit vector pointing to the kernel origin: $\mathbf{n}(x,y) = [-\frac{x}{r}, -\frac{y}{r}]$, r denotes the distance from the origin.

The advantage of VFC representation is that a free particle (i. e. a single contour point) placed in the field is able to move to the feature of interest. The $m(x,y)$ magnitude should be chosen as a decreasing positive function of the distance from the origin, such as:

$$m(x,y) = (r+\epsilon)^{-\gamma}, \tag{3.5}$$

with γ controlling the decrease and ϵ preventing division by zero. As the feature map of VFC (Eq. 3.3) is also based on the $f(x,y)$ original edge map, problems of

detecting high curvatures and noisy, weakly contrasted boundaries still exist (see Figure 3.1(c) and the fifth image in the first row of Figure 3.4).

High curvature corners can be found by corner detectors after generating a characteristic function which emphasizes possible corners in the image. The idea was to apply Harris corner detector as it is reliable and invariant to rotation [44]. The original Harris characteristic function is modified to be able to emphasize the high curvature corners and low contrast edges as well by exploiting the curvature information along the boundary and resulting in a balanced feature map.

3.3 Harris based Gradient Vector Flow (HGVF) and Harris based Vector Field Convolution (HVFC)

The process first generates the main feature points based on the Harris corner detector, then these points are enveloped to get the initialization of the contour. As a novelty, modification of Harris characteristic function is used instead of intensity function in the feature map for generating GVF and VFC snake.

3.3.1 Harris based feature map

A detailed overview about the original Harris detector [24] has already been given in Section 2.3.2.1, therefore only the calculated R characteristic function is introduced here again:

$$R = \mathrm{Det}(M) - k * \mathrm{Tr}^2(M), \tag{3.6}$$

where Det and Tr denote the determinant and trace and k is a coefficient, usually around 0.04 [24]. R is large and positive in corner regions, negative in edge regions and small in the flat regions, therefore, it is inappropriate to emphasize corners and edges equally.

When defining a new feature map for GVF snake, the curvature behavior should be analyzed [5]. Therefore, eigenvalues of M are applied to create a novel characteristic function, emphasizing the curvature around image pixels.

$$R_{\mathrm{logmax}} = \max(0, \log[\max(\lambda_1, \lambda_2)]). \tag{3.7}$$

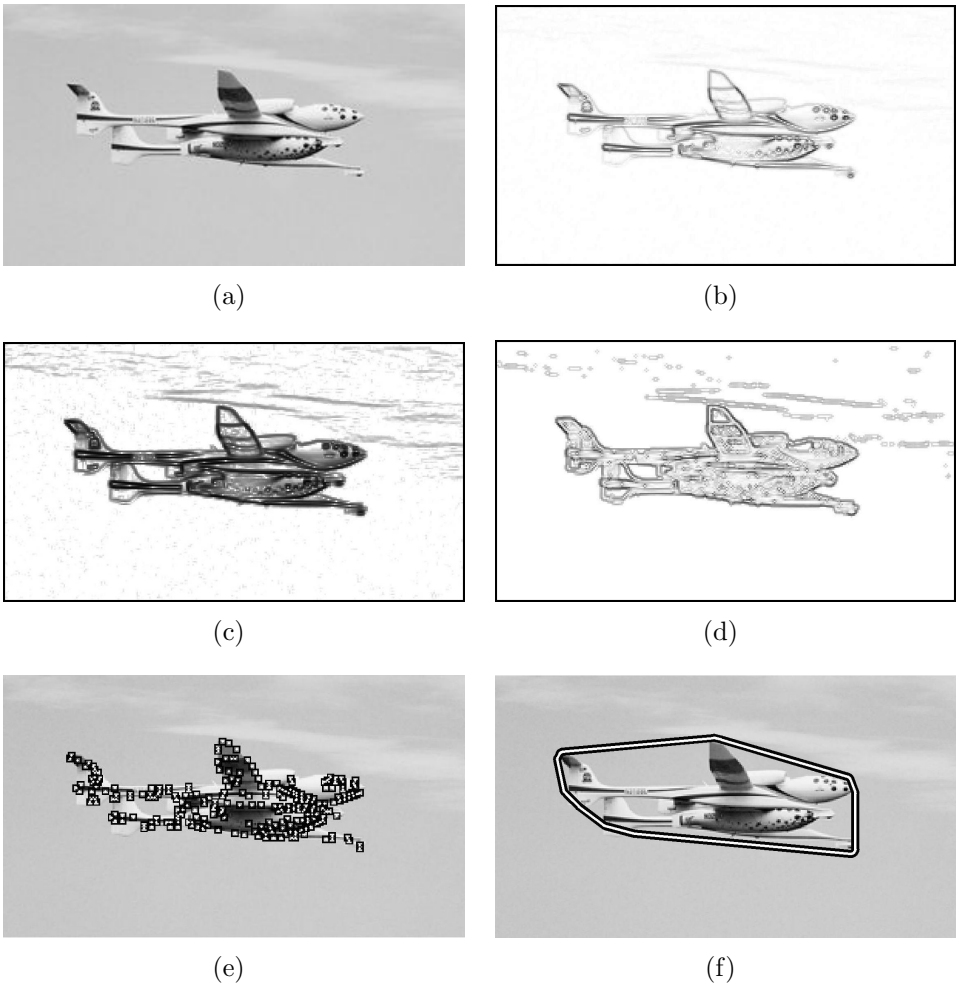


Figure 3.2: Effect of R_{logmax} characteristic function: (a) is the original image; (b) is the original, f intensity based map of GVF; (c) is the generated and inverted R_{logmax} characteristic function; (d) is the proposed f_{HGVF} Harris based map for GVF (Eq. 3.8); (e) shows the generated salient points as the local maxima of the R_{logmax} function (Fig. 3.2(c)); (f) is the initial contour based on the convex hull of the corner points.

Let λ_1 and λ_2 denote the eigenvalues of M (Eq. 2.18). For both edge and corner points $\max(\lambda_1, \lambda_2) >> 1$. The target set of the R_{logmax} is the positive domain (when it is used as a feature map), thus, the outer max function is responsible for replacing negative values of small λs (points in flat regions) with zeros. Example of the proposed R_{logmax} function can be seen in Fig. 3.2(c).

Now, with a balanced output for both corners and edges, R_{logmax} is proposed to be applied in the feature map:

$$f_{\text{HGVF}}(x,y) = \left|\nabla(G_\sigma(x,y) * R_{\text{logmax}}(x,y))\right|. \tag{3.8}$$

This function will be applied in the Harris based Gradient Vector Flow (HGVF) algorithm. Based on f_{HGVF} the new feature map for Harris based Vector Field Convolution (HVFC) method is as follows:

$$\mathbf{f}_{\text{HVFC}} = f_{\text{HGVF}}(x,y) * \mathbf{k}(x,y). \tag{3.9}$$

The comparison of the original and the proposed feature map can be seen in Figure 3.2. While the original, intensity based f edge map [20] loses some low contrast edges on the boundary and fails to emphasize the sharp corners (see Figure 3.2(b)), the proposed f_{HGVF} map (Figure 3.2(d)) is able to feature the real contour more accurately.

3.3.2 Initial contour

Feature (or salient) points are chosen as the local maxima of R_{logmax} (Eq. 3.7), see Figure 3.2(e). Since feature points are located on the edges and corners of the object, they can be used to define an initial contour. To avoid the poor definition of smooth transition or multi-directional saddle effects of edges around corners, the surroundings of the saliency points should be taken into account. Therefore, a local area with a 3 pixel radius around the point should be considered as part of the region of interest (ROI), where relevant local structures can be detected. Points representing the outline of the supported area should be added to the Harris feature point set. After this, the initial contour is defined as the convex hull of the extended set of points denoted by P (see Figure 3.2(f)). Convex hull has already been used for initializing active contour [63], [64], the innovation resides in the extended Harris based contour point set for localizing the object.

The convex hull of P set of points is the smallest convex set that contains all the points. It can be characterized as the set of all of the convex combinations of

finite subsets of points from P and looks as follows:

$$H_{\text{convex}}(P) = \Big\{ \sum_{i=1}^{k} \alpha_i p_i : p_i \in P,\ \alpha_i \in \mathbb{N},\ \alpha_i \geq 0, \sum_{i=1}^{k} \alpha_i = 1,\ k = 1, 2, \dots \Big\}. \tag{3.10}$$

A wide range of algorithms is known for constructing the convex hull for a finite point set with various computational complexities. In this implementation the built-in function of MATLAB was applied, which uses the 'Qhull' algorithm [65]. Let n denote the number of points in P and h the number of points in the hull, then the computational complexity of the calculation of the hull is $O(n \log h)$.

The algorithms follow the operation of traditional GVF and VFC methods, but proceed from the $H_{\text{convex}}(P)$ convex hull of the feature points and utilize the novel f_{HGVF} and $\mathbf{f}_{\text{HVFC}}$ maps instead of simple f in the external force component. (See Equation 3.1.) Therefore, the behavior of the proposed algorithms is similar to the traditional approaches, including parameter settings. When constructing the novel feature maps, the Gaussian window (see Section 3.2.1) have to be chosen, determining the smoothness of the map. In the experimental part, the w Gaussian window was used with $\sigma = 0.2$ for images without noise and $\sigma = 1.5$ for all the noisy images. Apart from this parameter, the proposed algorithms uses the parameters of the traditional active contour approaches, thus, the sensitivity to parameter tuning is analyzed in [20] and [51].

3.4 Experimental results and discussions

In this experimental part the proposed methods were evaluated quantitatively using the Weizmann segmentation evaluation database [62] and qualitatively on specific images, representing various characteristics: objects with weak edges or high curvature boundary parts and noisy images (Figure 3.4). The performance of the proposed methods was compared to traditional parametric and non-parametric active contour algorithms. Gradient vector flow (GVF) [20] and vector field convolution (VFC) [51] were selected to show the improvement of effectiveness due

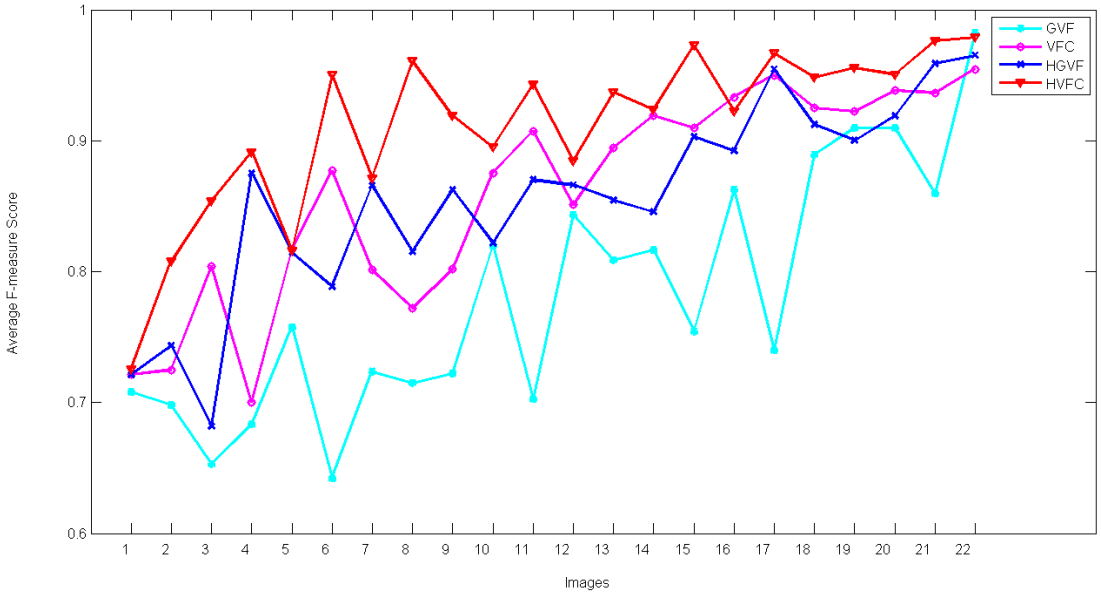


Figure 3.3: Detailed evaluation results. Vertical axis shows the achieved average F-measure score for each test image separately. Horizontal axis shows the numbered images from the Weizmann database [62] used for the evaluation set. Separate bars indicate the results of different methods: light gray is GVF [20], white is VFC [51], dark gray is HGVF (proposed) and black is HVFC (proposed).

to the introduced feature map. Region based methods can detect high curvature boundaries accurately, therefore, the non-parametric active contour without edges (ACWE) [47], based on the Chan-Vese model have also been evaluated. Decoupled active contour (DAC) [52] method has recently been introduced and was found to be able to detect regions of high curvatures and converging rapidly, thus, an image from the referenced paper was used for comparing the proposed method to DAC.

For evaluation, the published MATLAB source code of the compared algorithms was applied without optimization. Parameter settings for the compared methods were chosen according to the mentioned references.

3.4.1 Quantitative evaluation using the Weizmann database

In the quantitative part of the evaluation 23 test images (with ground truth data), having sharp edges with high curvature or weak contrast with unsettled

Table 3.1: Average F-measure Score (mean ± standard deviation) for GVF [20], VFC [51] and the proposed HGVF and HVFC algorithms for 23 images [62]. Bold text indicates the highest achievement.

Algorithm	Average F-measure Score
GVF	0.785 ± 0.0923
HGVF	0.865 ± 0.0807
VFC	0.858 ± 0.0734
HVFC	$\mathbf{0.913 \pm 0.0632}$

edge-transition, have been chosen for a numerical evaluation process from the Weizmann dataset [62] (see Figure 3.3). For each algorithm, initial contours are calculated as described in Section 3.3.2. Parameter settings for the compared methods were applied according to the mentioned references to achieve the best performance. Like it was also mentioned in [52], due to the intensity variation inside the object, this database is not suitable for ACWE method. Therefore, results were evaluated only for GVF, HGVF, VFC and HVFC algorithms.

For quantitative results the traditional F-measure score is used, which is the weighted, harmonic mean of precision and recall values:

$$F = \frac{2 \cdot \text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}. \tag{3.11}$$

According to the summarized results of 23 images, the average F-measure score (mean ± standard deviation) for each algorithm can be seen in Table 3.1. HGVF and HVFC provide higher F-measure score relative to the classical approaches.

Figure 3.3 shows the detailed evaluation results for the 23 images separately. It is important to note that Harris based methods (HGVF and HVFC) outperform their traditional corresponding in nearly all of the cases.

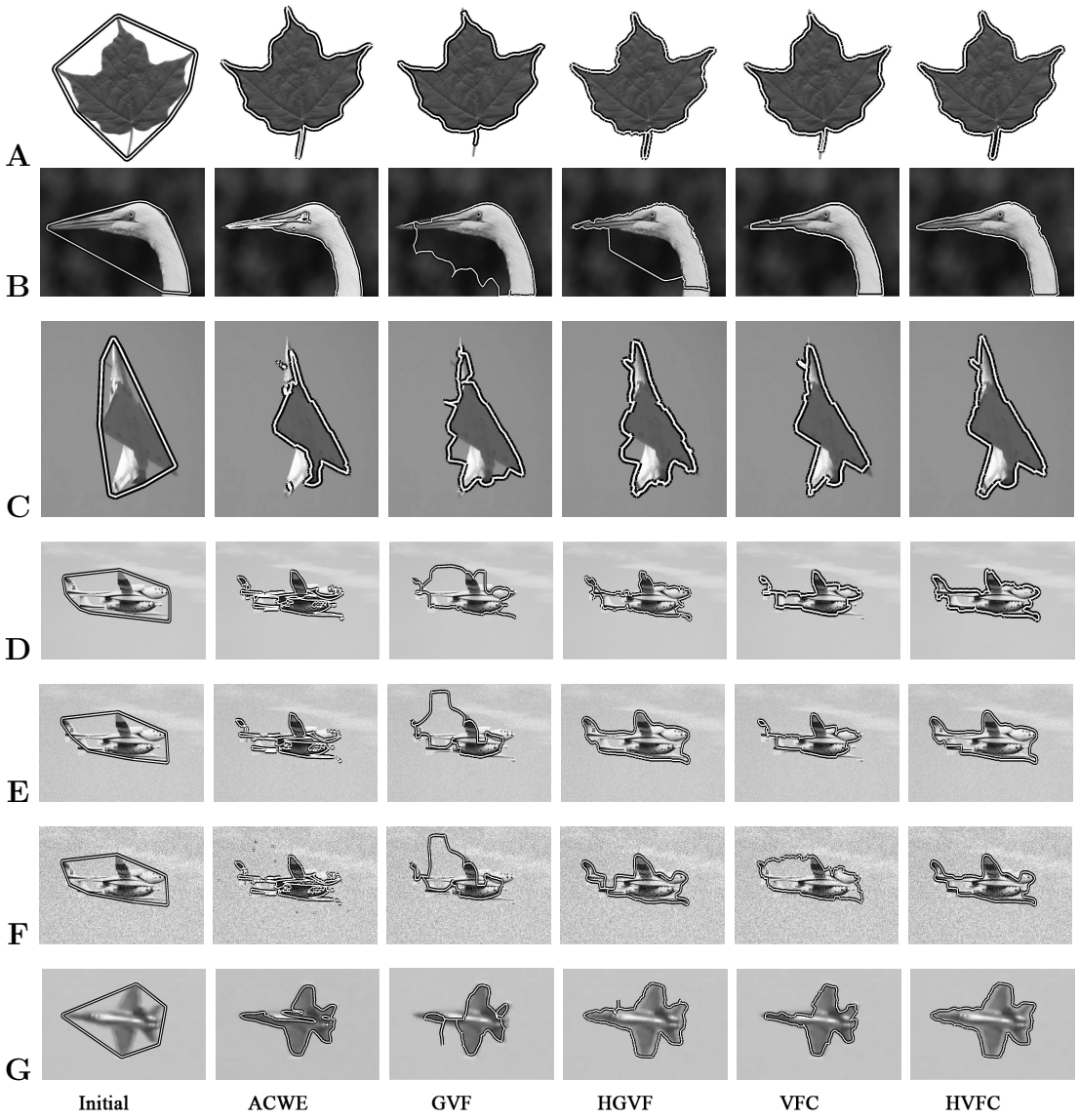


Figure 3.4: Examples of contour detection: The first column shows the calculated initial contour (see Section 3.3.2). Second, third, fourth, fifth and sixth columns present the results for ACWE [47], GVF [20], HGVF (proposed), VFC [51] and HVFC (proposed) methods.

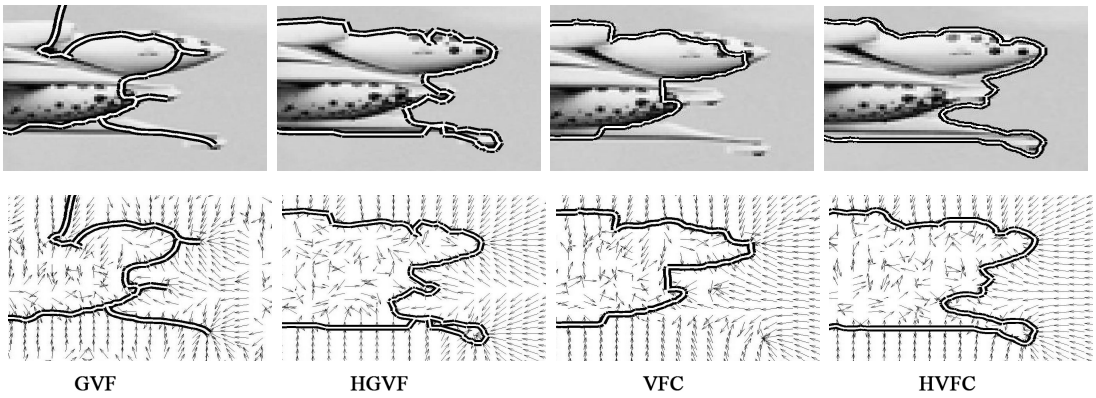


Figure 3.5: Improvement of the different feature maps in image **D** achieved by modified Harris based characteristic function: the first row shows the image part with the detected outline, the second row is the corresponding force field and the contour. Columns show the results for GVF [20]; HGVF (proposed); VFC [51] and HVFC (proposed) methods.

3.4.2 Qualitative results for boundary accuracy

In the next part of the experiments, specific images were selected to show qualitative results (Figures 3.4, 3.5 and 3.6). The execution time of different methods for images **A** – **D** (with sizes 200×216, 300×200, 335×364 and 300×170) in Figure 3.4 can be seen in Table 3.2(a). These experimental results were obtained on an Intel(R) Core(TM) i7 CPU with 4 GB RAM and MATLAB R2010b.

ACWE [47] method follows a level-set representation; therefore, it uses intensity homogeneity constraints instead of gradient based edge map. ACWE can successfully identify object boundaries even with high curvature parts, if intensity is homogeneous inside the object and the contour is closed properly (see image **A**, first row of Figure 3.4). Else, ACWE converges to object parts representing homogeneous regions which differ largely from the estimated background (**B**, **C**, **D** images in Figure 3.4).

Beside suffering from high curvature and low contrasted boundaries, as the detailed explanation showed in Section 3.2.1, GVF [20] also fails when the initial contour is further from the real boundary. In case of larger concavities of the object boundaries the convexity feature of the contour initialization step results in a distant initial contour, therefore, the method is trapped in local minima. The

Table 3.2: Performance of different active contour algorithms, including execution time for images without noise (a) and robustness to increasing Gaussian noise (b) for ACWE [47], GVF [20], HGVF (proposed), VFC [51] and HVFC (proposed) methods.

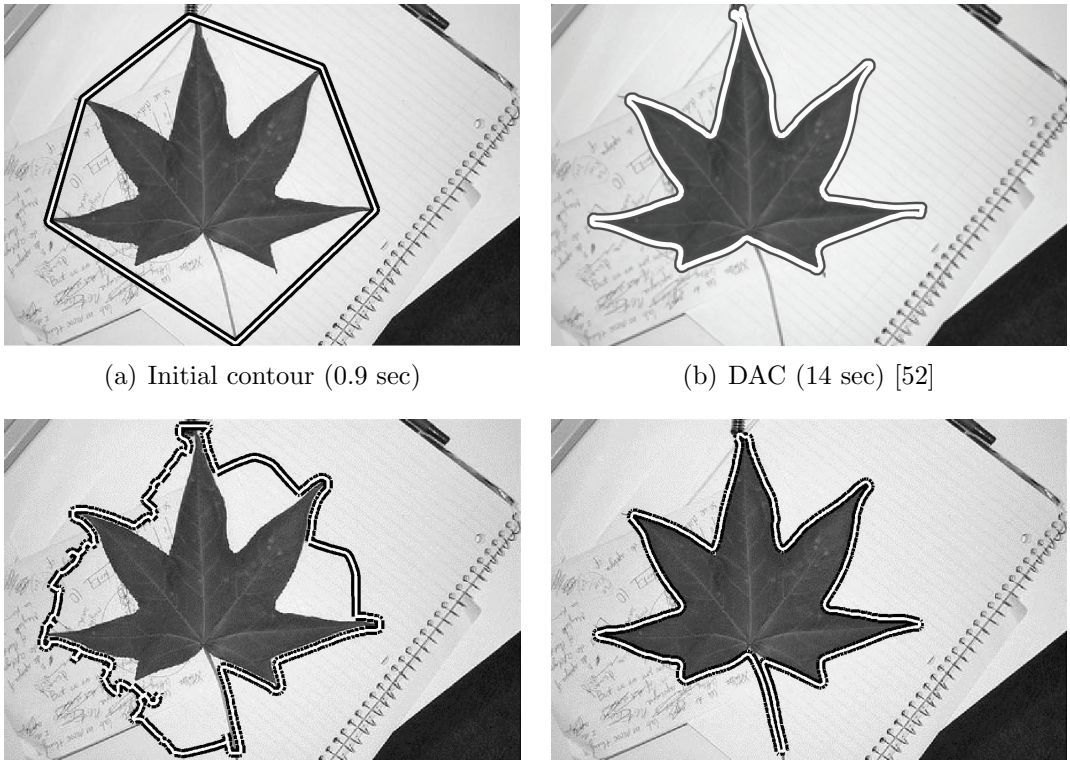
(a) Execution times for images **A** – **D** in Figure 3.4, IC indicates the initial contour.

	Execution Time [seconds]					
Images	IC	ACWE	GVF	HGVF	VFC	HVFC
A	0.36	13	5.8	5.6	3.7	4.8
B	0.44	66	6.2	6.7	4.2	5.1
C	0.85	68	9.4	11	6.3	6.9
D	0.38	12	3.9	4.3	3.2	3.7

(b) Robustness to noise for images **D**, **E** and **F** with increasing Gaussian noise, SNR is the signal-to-noise ratio, F-measure is given in Eq. 3.11.

	F-measure				
SNR	ACWE	GVF	HGVF	VFC	HVFC
∞ (Image **D**)	0.66	0.64	0.89	0.79	0.93
5 dB (Image **E**)	0.58	0.53	0.88	0.78	0.87
0.5 dB (Image **F**)	0.60	0.52	0.86	0.74	0.87

proposed HGVF improves the performance of the original method, by emphasizing low contrasted and high curvatures parts, therefore, it detects the peaks and the stem accurately in image **A**, and converges to the desired outline in image **D**. By enlarging a small part of image **D** having high curvature corners with weak edges, first and second row of Figure 3.5 show the detailed detection result and the corresponding magnified force field of GVF and HGVF methods. Unlike the edge map of GVF, the proposed feature map of HGVF is able to lead the contour to the real boundary. However, when the concavity is larger (like in images **B** and **C**), HGVF still fails.



(a) Initial contour (0.9 sec)

(b) DAC (14 sec) [52]

(c) HGVF (23 sec)

(d) HVFC (14 sec)

Figure 3.6: Comparison with Decoupled Active Contour (DAC) method [52] with the execution time in brackets. (Image size: 450×297.)

VFC [51] has the advantage to be less sensitive to initialization than GVF, due to the calculated vector field kernel. Therefore, large concave outlines do not cause challenge (like image **B**), but the method fails to detect high curvature and low contrasted boundary parts due to the existing problems of the feature map (see Section 3.2.2 for further details).

The proposed HVFC method benefits from the advantages of traditional VFC algorithm and the introduced feature map of modified Harris function and detects the aforementioned complex boundaries accurately (see last column of Figure 3.4 and 3.5).

Images **E, F** and **G** (with sizes 300×170, 300×170 and 282×191) in Figure 3.4 show images containing Gaussian noise with different signal-to-noise ratio (SNR) and an originally blurred image to evaluate the robustness of the different

methods against noise and blur.

For quantitative evaluation, **D, E** and **F** images were compared. The same parameter setting was applied in every case; therefore, the robustness to noise could be tested. Image **D** was the original image without any additive noise, Image **E** with 5 dB SNR, Image **F** with 0.5 dB SNR. In Figure 3.4, the results show that the proposed approaches are able to keep the main characteristics of the object outline, in Table 3.2(b) the corresponding quantitative F-measure scores can be seen.

In case of the blurred image **G**, HGVF and HVFC are able to converge to the real boundary more accurately than the compared GVF and VFC methods without skipping any object part with differing intensity as ACWE.

Figure 3.6 shows a comparison with the novel Decoupled Active Contour method [52] for an image with high curvature boundary parts. DAC addresses the limitations of traditional active contours (slow convergence rate and misconverging in the presence of noise or complex contours) by applying internal and external image forces independently. It consists of a measurement update step, employing a Hidden Markov Model (HMM) and Viterbi search, and then a separate prior step, which modifies the updated curve based on the relative strengths of the measurement uncertainty and the non-stationary prior. Since DAC method had just been published before the work of this chapter and no source code could be reached, DAC was compared with the proposed Harris based methods for image in Figure 3.6, which was also used in [52]. Result of the comparison showed that DAC misses the stem and some peaks of the leaf. Although the proposed HGVF algorithm is not able to converge into large concavities (due to the aforementioned drawbacks), the HVFC approach detects the contour almost accurately. By considering the execution time of the different algorithms, HVFC is able to converge as fast as DAC.

3.5 Applications of the introduced feature map and point set

Due to the promising results of the proposed feature map on specific test images, further applications of the formerly introduced Harris based feature map and

point set were developed. Although the detailed tasks are all related to complex object detection, they largely differ from each other: the first processes aerial images and extracts building outlines, by refining the results of the preliminary rectangle mask [8]; the second one is a medical application for single channel long time-span brain MRI image pairs to detect structural changes (multiple sclerosis (MS) lesions) using saliency map and active contour methods and the aim is to focus the radiologist's attention by providing a list of hypothetical lesions, that corresponds to significant, but not anatomical changes [12]. This work was a collaboration together with the Semmelweis University MR Research Center (MRKK), providing the MRI image pairs and supporting the medical background. The third application was a flying target detection and recognition [9, 2], which used the novel feature map and point set as a preprocessing step to localize and extract small flying targets (like airplanes or helicopters).

3.5.1 A joint approach of MPP based building localization and outline extraction

3.5.1.1 Introduction

Building extraction is a key task in several aerial surveillance, exploration and recognition applications. In the present section a single optical image input (an aerial photo or a satellite image) is processed, where building identification is a hard monocular object recognition problem. Previous methods can be divided into three different groups based on the levels of solving the task. First, some techniques address purely detection and counting the houses [67], by extracting the estimated center points of the buildings without any geometrical information. Although these approaches are efficient in computational complexity and less sensitive for parameter tuning, they can hardly adopted for building classification or change detection which require more detailed shape, size and spectral characterization of the entities. A second group of the existing approaches uses simplified geometric approximations of the buildings or building parts, such as rectangles [66] or hexagons [69]. These models are suitable for a more comprehensive interpretation of the scenes, however fine details and the irregular patterns can not be retrieved accurately. As the resolution of the available remotely sensed images

increases rapidly, such rough modeling of the building shapes yields a notable amount of lost information for the higher level interpretation modules. These effects increase the significance of a third group of approaches, which aim to describe the accurate building outlines by general parametric curves [70]. Variational techniques are efficient tools to address contour approximation, meanwhile they can embed prior information for the building outlines. However, contour initialization is challenging, especially, if the scene can not be described by a few template building shapes as in [70]. To achieve an efficient outline estimation, corner detectors [24] can be applied to extract contour points of the objects. Based on these contour points, active contour algorithms [20, 50, 51, 71] are able to detect the outline of the buildings.

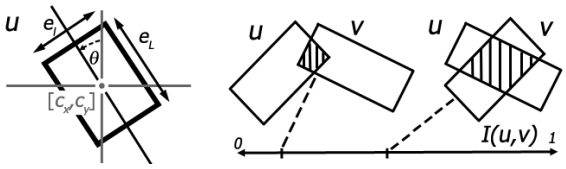
3.5.1.2 Proposed approach

The proposed approach consists in two steps. *First*, the buildings are localized in the scene and the shape of each rooftop is roughly estimated. This task is fulfilled by a Marked Point Process (MPP) model (Section 3.5.1.3–3.5.1.6) introduced in [72, 74], which fits rectangles to the observed building candidates by a Markov Chain Monte Carlo method, similar to rendering in [68]. *Second*, using the coarse location and shape information feature points are identified inside and around the rectangular rooftop masks, and based on the selected key points an active contour is initialized to obtain the building outlines [5]. The main contribution of this work lies in the linking process between the two steps: a novel key point selection technique is proposed for the initial contour, which jointly exploits low level features from the original image and structural descriptors from the MPP output mask. Finally the efficiency of the cascade procedure is tested on real aerial images and discussion is given about the results.

3.5.1.3 Preliminary building mask estimation with MPP model

The goal of this model part [72, 74] is to assign to each building segment of the scene a unique rectangle which estimates the mask of the observed rooftop.

Let S denote the 2-D pixel lattice of the input aerial image, and $s \in S$ a single pixel. A rectangular building segment candidate u is described by four



(a) (b)

Figure 3.7: Demonstration of the (a) object rectangle parameters and (b) calculation of the interaction potentials [72].

parameters: $s_c = [c_x, c_y] \in S$ center pixel, e_L, e_l side lengths and $\theta \in [-90°, +90°]$ orientation [see Figure 3.7(a)].

An aerial image is modeled as composition of an arbitrary number of u rectangular Building Segments (BS) whose positions and geometric parameters are realizations of a MPP [73]. Let Ω denote the space of all configurations of a finite number of BS objects. A given object configuration in Ω will be referred by ω, where $\omega = \emptyset$ or $\omega = \{u_1, \ldots, u_n\}$ for an arbitrary positive integer n. Moreover, a neighborhood relation is defined in the BS object space: $u \sim v$ if their rectangles intersect.

Next, a $\Phi(\omega)$ energy function is introduced on the Ω configuration space, to fulfill that the optimal configuration exhibits the minimal energy. The energy function takes into account the interactions between the geometric objects (the prior energy $\Phi_p(\omega)$), and the way they fit to the image data (the data energy $\Phi_d(\omega)$):

$$\Phi(\omega) = \Phi_p(\omega) + \Phi_d(\omega).$$

The optimal BS population $\hat{\omega}$ is obtained as the Maximum Likelihood (ML) configuration estimate:

$$\hat{\omega} = \underset{\omega \in \Omega}{\operatorname{argmin}}\, \Phi(\omega). \tag{3.12}$$

3.5.1.4 Prior energy

The first is the $\Phi_p(\omega)$ energy part [72, 74], which describes *prior geometric* constraints in the model. Since the aim is to extract individual building entities, overlapping between different BS rectangles must be penalized. Therefore, $I(u, v)$

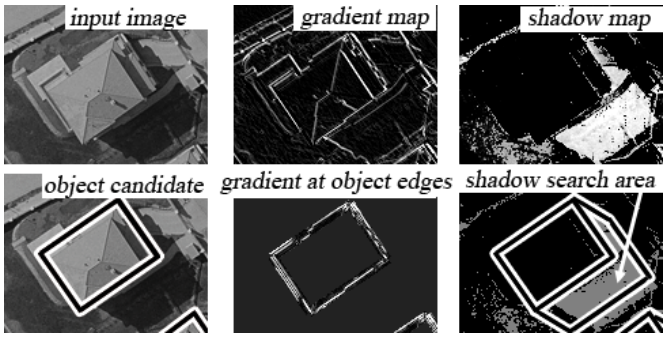


Figure 3.8: Edge and shadow features [8].

interaction potentials are defined, which penalize object pairs according to the $I(u, v)$ normalized intersection area [Figure 3.7(b)]:

$$I(u, v) = \frac{\text{Area}\{u \cap v\}}{\text{Area}\{u\} + \text{Area}\{v\}},$$

where $u \cap v$ means the intersection of objects u and v.

Finally, the complete prior energy term of the population is calculated as:

$$\Phi_p(\omega) = \sum_{\substack{u,v \in \omega \\ u \sim v}} I(u, v)$$

.

3.5.1.5 Data energy

The *data energy* [72, 74], $A_{\mathcal{D}}(u)$, associated with object u, characterizes a proposed building segment $u = \{c_x, c_y, e_L, e_l, \theta\}$ depending on the local image data, but independent of other objects of the population. The data term of the whole configuration is obtained as:

$$\Phi_d(\omega) = \sum_{u \in \omega} A_{\mathcal{D}}(u)$$

A BS u with $A_{\mathcal{D}}(u) < 0$ is called *attractive object*. Since according to Section 3.5.1.4 all prior terms are non-negative, the optimal population (3.12) should consist of attractive objects exclusively: if $A_{\mathcal{D}}(u) > 0$, removing u from the configuration results in a lower $\Phi(\omega)$ global energy.

To fit the above framework to the building detection task, the $A_{\mathcal{D}}(u)$ data term should be defined appropriately. This term must assigns to each proposed object candidate an energy value (i.e. negative fitness value), which evaluates the hypothesis that u is a building in the image. The proposed energy function, detailed in [72], can integrate different feature-information such as roof color, roof edge and shadow. On one hand, *red* roofs [69] can be detected in color images using the u* and v* color components of the corresponding pixel values in CIE L*u*v* color space representation (See Figure 2.19). Thereafter with notation R_u for the mask and $\tilde{R}_u$ for the dilated mask of u's rectangle, the color term prefers objects which contain in majority roof colored pixels inside R_u and background pixels in $\tilde{R}_u \backslash R_u$. For non-red roofs the gradient and shadow maps could be used, exploiting that under the roof edges strong intensity changes should be observed in the images, while in sunny weather dark shadow blobs are present next to the buildings in the shadow direction [66] (See Figure 3.8).

3.5.1.6 Optimization

For the estimation of the optimal object configuration (3.12) the Multiple Birth and Death (MBD) algorithm [73] is applied [72, 74]. The steps are as follows:

Initialization: start with an empty population $\omega = \emptyset$.

Main program: initialize the inverse temperature parameter $\beta = \beta_0$ and the discretization step $\delta = \delta_0$ and a constant basic birth frequency b_0. Alternate the following three steps, till convergence is reached

1) *Birth step*: Visit all pixels of the S lattice one after the other. At each pixel $s \in S$ call *Object Generation* procedure with probability $\delta \cdot b_0$: create a new BS object u with center s; set u's side length and orientation parameters randomly following prior size distributions and add u to the current configuration ω.

2) *Death step*: Consider the current configuration ω and create a list of the $u \in \omega$ objects sorted from the highest to the lowest $A_{\mathcal{D}}(u)$ values. For each object u taken in this order, compute the cost of deleting u from ω w.r.t. the global configuration energy:

$$\Delta\Phi_{\omega}^{u} = \Phi(\omega/\{u\}) - \Phi(\omega) = -A_{\mathcal{D}}(u) - \sum_{\substack{v\in\omega \\ v\sim u}} I(u,v)$$

Next, the $p_d(u)$ *death rate* is derived as follows:

$$p_d(u) = f\left(\Delta\Phi_{\omega}^{u}\right) = \frac{\delta \cdot \exp\left(-\beta \cdot \Delta\Phi_{\omega}^{u}\right)}{1 + \delta \cdot \exp\left(-\beta \cdot \Delta\Phi_{\omega}^{u}\right)} \tag{3.13}$$

Finally, u is removed from ω with a probability $p_d(u)$.

3) *Convergence test*: if the process has not converged yet, increase the inverse temperature β and decrease the discretization step δ with a geometric scheme, and go back to the *Birth step*. The convergence is obtained when all the BS objects added during the birth step, and only these ones, have been killed during the death step.

Figure 3.9: Rectangular footprint results, obtained by the MPP based detector module [72].

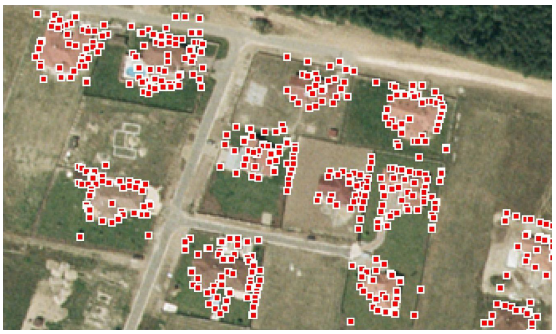


Figure 3.10: Extracted feature points around the object locations estimated by the MPP based detector.

3.5.1.7 Discussion of the MPP detector results

The above detailed MPP detector results in a set of oriented rectangles which estimate the building footprints. As seen in Figure 3.9, most building segments are correctly detected, however some roofs are only partially covered by the extracted masks, and connected roof parts may break up into many pieces (see the building in the bottom left corner). Moreover, since the roofs in the scene have various and irregular shapes, the proposed rectangular outline approximation proves to be often coarse. For this reason, a second step, detailed in the next section, is performed, which is responsible for the extraction of more faithful contours.

3.5.1.8 Refinement of the MPP Detection

Based on the contribution of this section, Harris based feature map and point set is able to enhance edges and corners in the image, even in the presence of noise and low contrasted objects. Therefore, the feature map introduced in Section 3.3.1 is used to extract feature points as the local maxima of this map in the close proximity of the estimated rectangle building masks, which can be seen in Figure 3.10. To separate the feature points of closely located buildings edge information is also exploited in a graph based representation like introduced in Section 2.3.2.5, Figure 3.11 shows the result of this step. After defining the initial contours as convex hulls of the separated point subsets (Section 3.3.2), Gradient Vector Flow (GVF) [20] active contour method detects the accurate object boundaries. To

Figure 3.11: Subgraphs given after matching procedure.

achieve the best performance, the initial contour of the GVF has to be defined closely to the object boundaries. Two cases were defined:

1. if the calculated convex hull surrounds an s_c center pixel of the u building candidate estimated by the MPP detector, then it was used as initial contour;

2. else the outline of the rectangle mask of u was used as initial contour.

After the GVF iterative process the detected contours are checked. If the detection was successful, the obtained boundary should contain the center of the given rectangle mask. If this condition is not satisfied, then the detection has to be reconsidered. Depending on the initial contour, two cases were separated again:

1. if the convex hull was used as initial contour, the GVF method is repeated with the rectangle mask used as initial contour and the resulted contour is checked again with the same procedure;

(a) (b) (c) (d) (e) (f)

Figure 3.12: Results of the joint building localization and outline extraction approach. In the first column the original images can be seen, the second column shows the detected buildings.

2. if the outline of the rectangle mask was used as initial contour, then the unsatisfied condition shows that the active contour method could not refine the result of the Marked Point Process. In this case the given rectangle mask will be the detection result.

Typical results of the refinement process are shown in Figure 3.12. The advantage of the MPP detector (see Sec. 3.5.1.3) is the accurate localization of buildings, but the rectangular contour estimation should be refined. By comparing Fig. 3.9 and Fig. 3.12 the improvement of the joint approach can be observed. The various outline of buildings were detected more accurately and broken building segments were merged into one contour (building in the bottom left corner).

3.5.2 Automatic detection of structural changes in single channel long time-span brain MRI images using saliency map and active contour methods

Detection of changes between medical image frames can be applied in the recognition and the monitoring of the procession of many diseases, from which this work concentrates on Multiple Sclerosis (MS). The main steps of a change detection method [75] are: (i) the geometric registration of images; (ii) the comparison, which may include preprocessing steps like intensity adjustment, difference image calculation.

Lately, Magnetic Resonance Imaging (MRI) is often applied for imaging different structures (such as the brain and muscles), as it provides better contrast between the different soft tissues than computed tomography (CT) and X-rays and does not use ionizing radiation. Clinical diagnosis of MS strongly relies on multispectral MRI scans, to detect typical lesions, to count them, to decide if standard MRI criteria of the disease are met and to detect new lesions during the course of the disease. The importance of disease progression measured by MRI is to classify the type of MS and to measure the efficiency of the therapy.

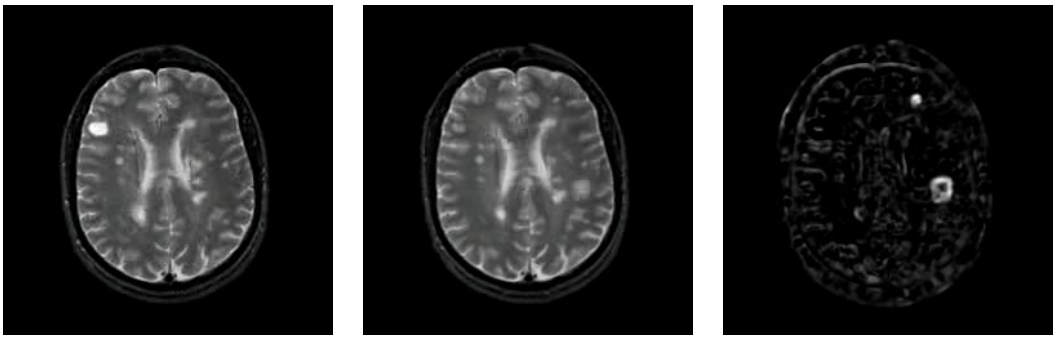
One main group of approaches applies statistical technique, like [76], where distribution of each brain tissue is modeled using a Gaussian Mixture Model and MS lesions are detected as outliers with respect to this model. However, in some cases only one modality of MRI images is collected, since the scanning is costly or the duration of the examination might be cut down due to the patient's condition. [77] detects infarct lesions from single MRI modality, the approach based on the theory, that, when lesions are present, the voxel-intensity-based segmentation and the spatial-location-based tissue distribution should be inconsistent in lesion

regions. In the absence of prior knowledge, [78] applies local regression kernels to measure the likeness of a pixel to its surroundings, resulting in a dissimilarity map between reference and target image. Dissimilarity map values exceeding a threshold then identified relevant changes.

In this section, an automatic structural change detection method is introduced for single channel MRI image pairs, which does not require any prior knowledge (brain atlas or tissue probability maps). This method misses the segmentation of different tissues, only concentrates on the appearance of MS lesions. The primary aim of the method is not to classify and localize regions, but to focus the radiologist's attention by providing a list of hypothetical lesions, that corresponds to significant, but not anatomical changes. After a coarse registration of the images, a difference image is calculated based on a modified Harris characteristic function [24] to define change candidates. Features from the neighborhood of the candidates, called local contour descriptors [4], are compared in the reference and target image to filter out false hits. The result of the algorithm has been tested on simulated and real data, including registered, longitudinal images from [79] and non-registered image pairs provided by Semmelweis University MR Research Center (MRKK).

3.5.2.1 Image registration

Although there is an internationally accepted MRI protocol, image series taken at different examinations can largely vary, as they may take place at various institutions with different MRI machines using different settings and head position. Moreover, in spite of the thorough stabilization during the experiments, the patient's head might move, due to the heart-beat and body movements. These circumstances can all lead to altered image characteristics. Additionally, the progression of MS may cause varying lesions, meaning topological (structural) changes, making the registration step more complicated (see Fig. 3.13). Both [80] and [81] collected many ways to precise registration, even with deformations, however, these methods are not able to handle all the aforementioned changes. Thus, instead of constructing an improved registration step, only a coarse, SIFT point based matching [19] was performed and then a difference map was constructed, which is able to handle possible misregistration errors.



(a) (b) (c)

Figure 3.13: (a): Reference image (I_r); (b): Target image (I_t); (c): Probability map $P^{(i)} = R_t^{(i)} \cdot R_d^{(i)}$.

3.5.2.2 Difference image calculation

Appearing lesions can be emphasized and then searched by a well-defined difference map. When constructing such map, the fact that different examinations may cause altering contrast, noise conditions and variance in anatomical structures should be taken into account. Thus, simple techniques like intensity or edge difference and thresholding cannot be adopted. In Section 2.3.2.2 Harris detector's characteristic function was used to construct the difference image, as aerial change detection also suffers from the heavy changes of the background (illumination changes, weather circumstances, etc.). Therefore, it could be used efficiently for change detection in MRI images as well.

In case of MRI images pairs, when searching for round-shaped MS lesions, the original Harris characteristic function has to be modified, to emphasize not only corners, but edges and corners equally. Therefore, the proposed R_{logmax} modification (Equation 3.7 in Section 3.3.1) was applied for enhancing structural changes.

3.5.2.3 Change detection

After calculating the R_{logmax} function for both the reference (R_r) and the registered target image (R_t), a difference image is counted as the absolute value of their difference ($R_d = |R_t - R_r|$). For localizing appearing lesions between the

reference and target images, the following assumption was formulated: Changes have high effect both on the Harris-modified target (R_t) and the difference image (R_d), so joint local maxima in both images have to be searched. Therefore, by performing a pixel-wise multiplication as:

$$P^{(i)} = R_t^{(i)} \cdot R_d^{(i)}, \tag{3.14}$$

a probability map is given for change candidates. By thresholding the P map (Figure 3.13(c)) with Otsu's method [82] (resulting in a binary B edgemap), the set of local maxima (denoted by H) define the location of the change candidates (see Fig. 3.14(a)).

Otsu's method was introduced for automatic image thresholding, by separating the image to two (or more) classes by minimizing the weighted intra-class variance. Otsu showed that minimizing the intra-class variance is the same as maximizing the inter-class variance. The method first computes a histogram and probabilities of each intensity level and sets up an initial class probabilities and class means. Then it steps through all possible thresholds, updates class probablities and means and computes the inter-class variance. The desired threshold corresponds to the maximum inter-class variance.

H may contain false candidates due to differing anatomical structures on the target image, which have to be filtered out. Local contour descriptors (LCDs, introduced in Chapter 2) were applied to compare the neighborhood of the candidates in R_r and R_t. To handle various lesion sizes, multi scale LCs were generated, one for a smaller (c_1) and one for a larger (c_2) radius. As the size of lesions cannot grow arbitrarily and the diagnostic criteria labels a mutation as lesion if bigger than 3 mm, c_1 and c_2 values are selected depending on the resolution of the image.

After calculating the distance metric (Equation 2.14 in Section 2.2.3) for both sizes, dissimilarity metric for the ith candidate (x_i, y_i) looks as:

$$DS_i = \min(D_{i,c_1}, D_{i,c_2}). \tag{3.15}$$

DS_i shows the change rate of the neighborhood of the candidate (Figure 3.14(b)): if it is small, the region has not changed; if it is large, the region

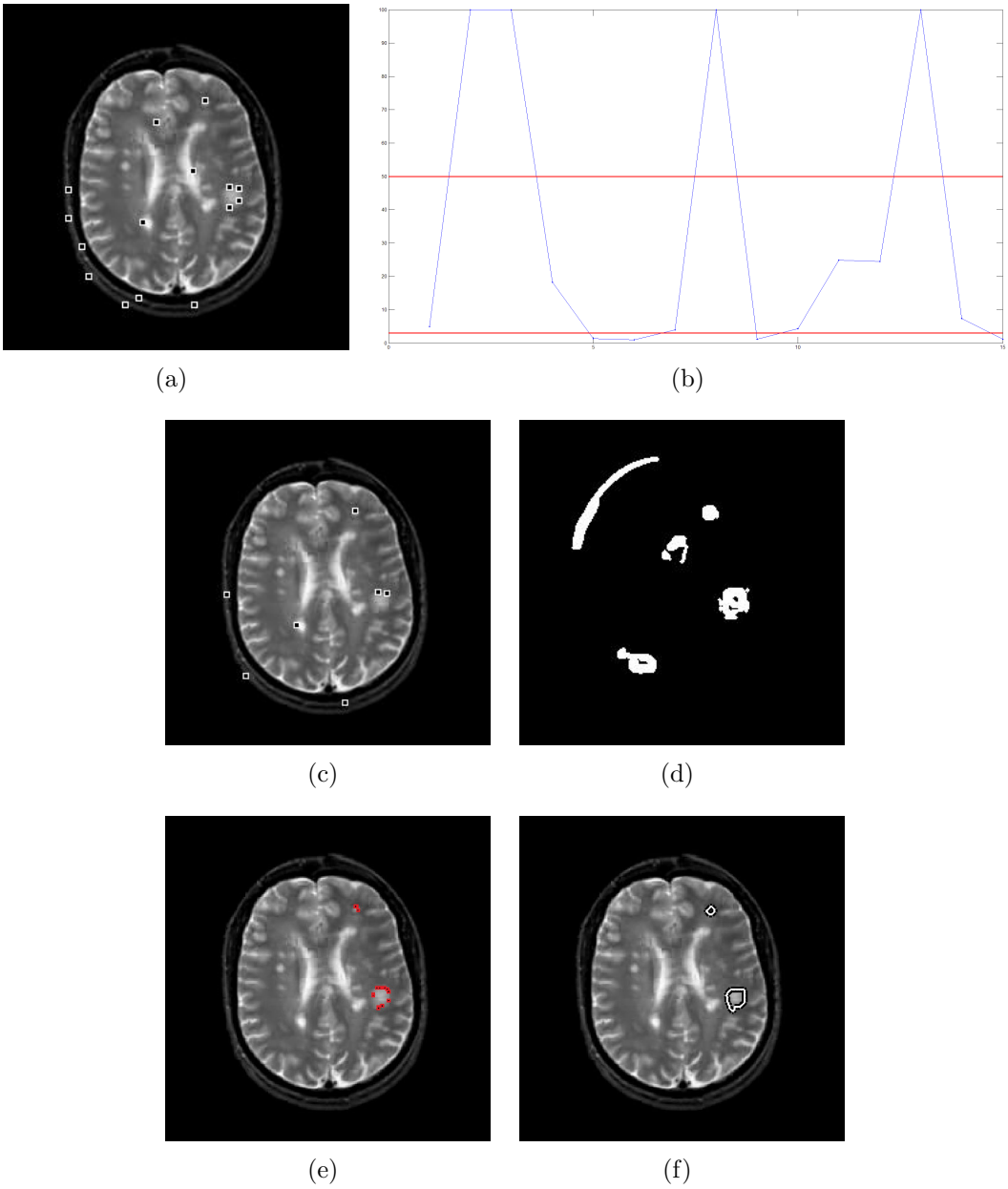


Figure 3.14: Steps of the algorithm. 1. Change detection: (a) Change candidates, (b) The DS_i values for candidates in blue scaled into $(0, 100)$ range with $\epsilon_1 = 3$, $\epsilon_2 = 50$ thresholds marked by the red lines, (c) Localization of changes; 2. Lesion boundary recognition: (d) B map, (e) Lesion contour point extraction; (f) The detected lesion boundary.

has significantly changed. In case of an appearing lesion, the surrounding has changed, but not as much as in case of an anatomical structure variance caused by registration errors. Therefore, after calculating DS_i for every candidate, a distribution function is given from the values, containing ranges of different orders, which can be simply separated automatically. The preassumption for selecting lesion candidates is that only a few pixels will represent significant changes in the image, thus most of the candidates will be grouped either in the non-changing (left tail of the distribution function) or the misregistration error class (right tail of the distribution function). By defining a lower and higher threshold automatically, containing a few points between them, the remaining set of change keypoints (Fig. 3.14(c)) is as follows:

$$H' = \{(x_i, y_i) \in P : \epsilon_1 \leq S_i \leq \epsilon_2\}, \tag{3.16}$$

where ϵ_1 and ϵ_2 are the lower and higher threshold, determined based on the distribution function automatically, cutting down both tails of the value range. It should be mentioned that in case of registered image pairs and simulated lesions (like Fig. 3.15), images can be matched precisely and misregistration errors are not represented (DS values of high order are missing), therefore only the non-changing regions have to be filtered out. This means that only a lower threshold is needed, cutting of the left tail of the distribution function.

3.5.2.4 Lesion boundary recognition

The change detection step localized the lesions, now the accurate boundary needs to be detected. Therefore, first contour points on the boundary of the lesions are extracted, and then iterative curve evolution method is applied by using the extracted contour points as initialization.

For the extraction of contour points, P (Equation 3.14) is thresholded with Otsu's method (Thr), a binary edge map (B) is given. This edge map contains the points of H' as previously H was chosen as the set of local maxima over Thr (see Section 3.5.2.3). Contour points are extracted as the points connected to the element of H' in B (Fig. 3.14(d)).

After the extraction step, a subset is given for every $h_i \in H'$, including contour points of the object and h_i itself. By calculating the convex hull of the elements in

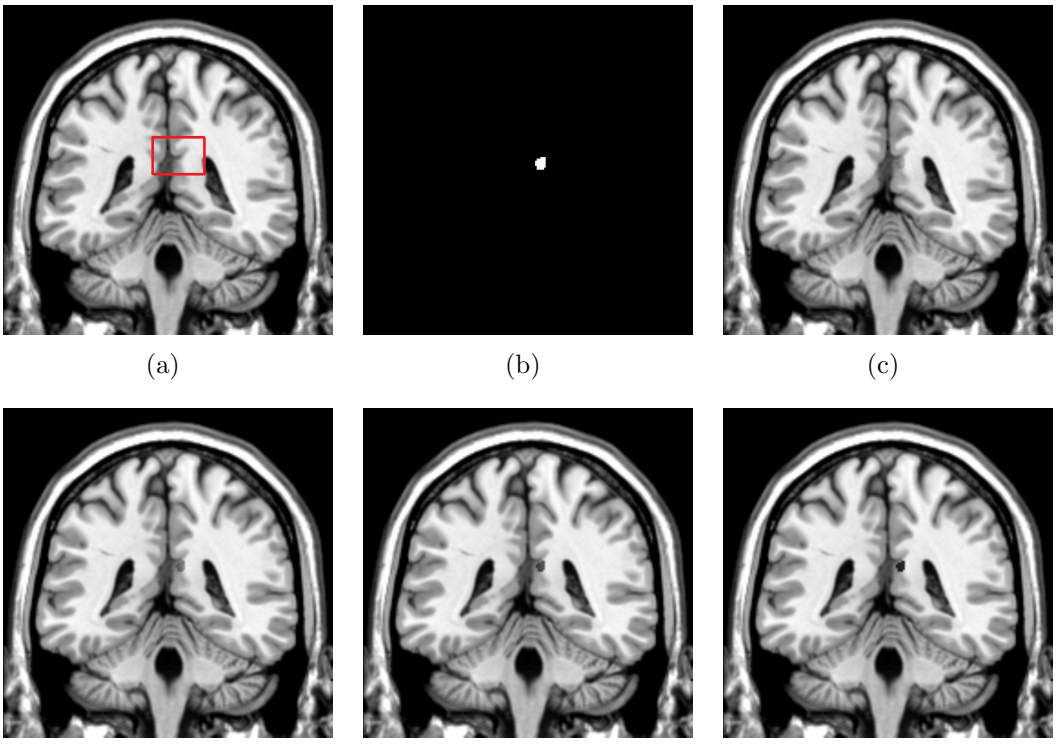


Figure 3.15: Example for simulated lesions: (a) is the original image without lesion (red rectangle indicates the location of the lesion in the next images); (b) is the mask of the simulated lesion; (c)–(f) are the images with simulated lesion with 20%, 40%, 60% and 80 % intensity reduction respectively. Image is taken from the MRIcro software.

the subset, an initialization for curve evolution is achieved. (If convex hull cannot be constructed because of the small number of points, the subset may mark noise and is eliminated.) The region-based Chan-Vese method [47] is applied (with $\mu = 0.1$ and $\lambda_1 = \lambda_2 = 1$ parameter settings for 500 iterations) to refine lesion boundary with the counted convex hull as initial contour (Fig.3.14(f)).

3.5.2.5 Experimental results

To validate the method quantitatively, random shaped lesions were simulated in normal MRI scans with MRIcro software and they were treated as ground truth. To model different cases, lesions with 20%, 40%, 60% and 80% intensity change

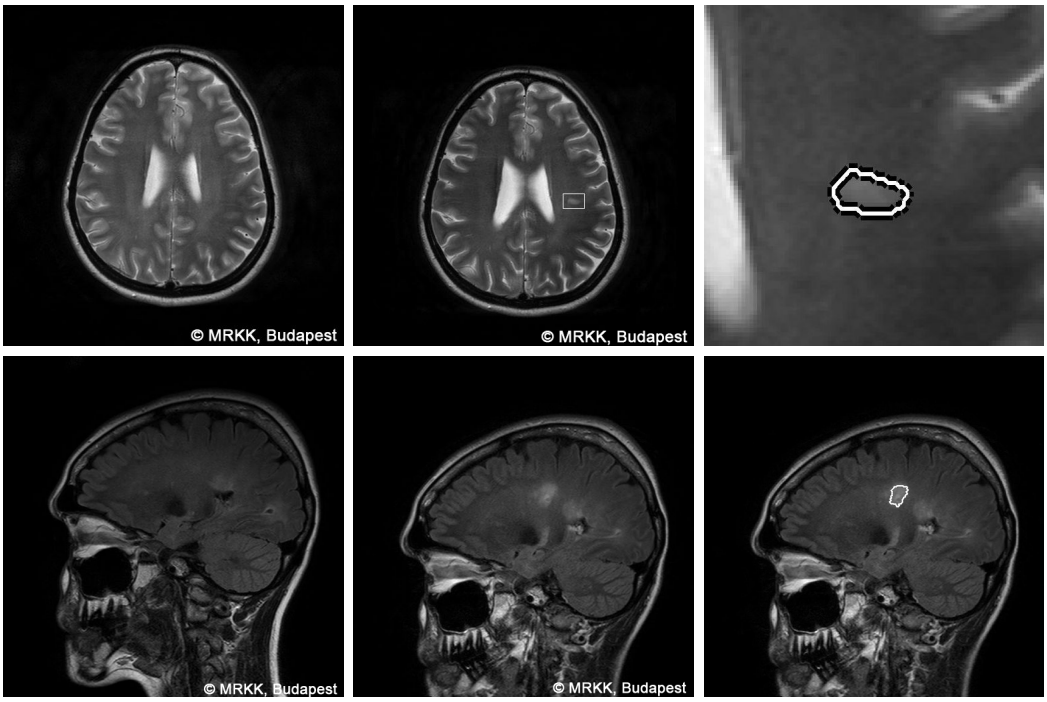


Figure 3.16: Reference image, target image and result of detecting appearing lesions for images provided by MRKK.

(Fig.3.15) were constructed and the intensity range of the reference and target images was also modified. Same evaluation technique was applied in state-of-the-art approaches [78], [76]. 21 simulated lesions were used in different images and views (axial, coronal and sagittal) to test the method, which meant altogether $21 \times 4 = 84$ test cases. In all cases the same parameter setting ($c_1 = 10$; $c_2 = 15$)

Table 3.3: Average SI scores on simulated cases.

Algorithm	SI Score
Rousseau et al. [76]	0.75
Seo and Milanfar [78]	0.88
Proposed method	**0.95**

was applied. To evaluate the results the traditional similarity index, SI $= 2 * \frac{GT \cap DT}{GT \cup DT}$ was used, with GT and DT denoting ground truth and detected result. Table 3.3 shows quantitative results, by comparing the achieved results with the ones reported in [76] and [78], the proposed method performs better.

High SI value was achieved due to the absence of registration errors and anatomical variance, which meant a challenge in case of real images. Two types of real images were tested, the first consisted of registered image pairs from [79], the second included non-registered pairs provided by Semmelweis University MR Research Center. The first dataset included two 'tours', where the patient was scanned more than twenty times in one year to follow the evolution of different lesions (see Figure 3.13). The 'time-lapse' images were registered with a software called Superpose. (No information could be found about this software.) The second dataset is from real, unregistered image pairs, with appeared lesion(s) in the target image. As no ground truth data was given, results of detection of appearing lesions (Fig. 3.16) can be adjudged only qualitatively.

3.5.3 Flying target detection

Visual detection, recognition, classification and tracking of stationary of moving targets are among the most active research areas in computer vision and image processing fields. Applications built on the results of these research areas are constantly sought to be deployed for both defensive and offensive scenarios, including civilian and military use. For civilian applications, wide area surveillance, crowd and traffic monitoring and target tracking are the most important fields, while for military applications troops and asset protection, region of interest surveillance, target detection and tracking are probably the most important scenarios. Aiding such tasks by intelligent and automatic visual processing is important since such methods can aid the detection, recognition and alerting tasks of security personnel. Also, visual processing sensors/nodes can provide a means for passive detection (without requiring active signals), thus making them harder to detect and disarm in case of sensitive scenarios.

This section presents a preprocessing step of a proposed solution for one aspect of the above described wide range of possibilities, focusing on automatic airborne

target detection. The presented approach can be used in defense and surveillance scenarios, where passive detection capabilities are preferred (or required) over a secured area or protected zone. The goals are to automatically detect a class of observed flying targets from varying angles, views, size and environmental conditions.

Wenga et al. present a flying target detection and tracking method [83] in infrared. Here the goal is detection and tracking, without recognition/classification. Image complexity, number of objects and number of other large areas (e.g. cloud objects) is taken into consideration, and detection is performed depending on the weather condition (clouds or clear skies). Also, clouds are separated based on histogram analysis, assuming white cloud color. On the other hand, the approach introduced in this section does not use or depend on such information, any background clutter (clouds, vapor trails, smoke, independent of their color) get automatically discarded based on their non-relevance as target candidates (based on their features), and the method is independent on the presence of such clutter.

Noor et al. [84] present a model generation approach for object recognition, using multiple views of objects to build a model database. They start with SIFT descriptors for relevant corner point extraction, used to build a region-neighborhood graph that is used for object matching. In this case a more robust interest point extraction is needed because of the variances in backgrounds and viewing angles, thus the formerly presented more robust Harris based point extraction approach is extended to handle multiple objects and background clutter simultaneously. The presented extension is robust against noise and clutter in the scene (e.g. clouds, vapor trails, other interference like illumination changes) and contains novelties with respect to classical multi-layer background/foreground modeling methods, shifting the complexity from the background modeling to a faster process of robust boundary point and contour segment extraction.

Figure 3.17 shows the main algorithmic steps of the whole recognition method, the red framed rectangle is the step that will be detailed in the next section.

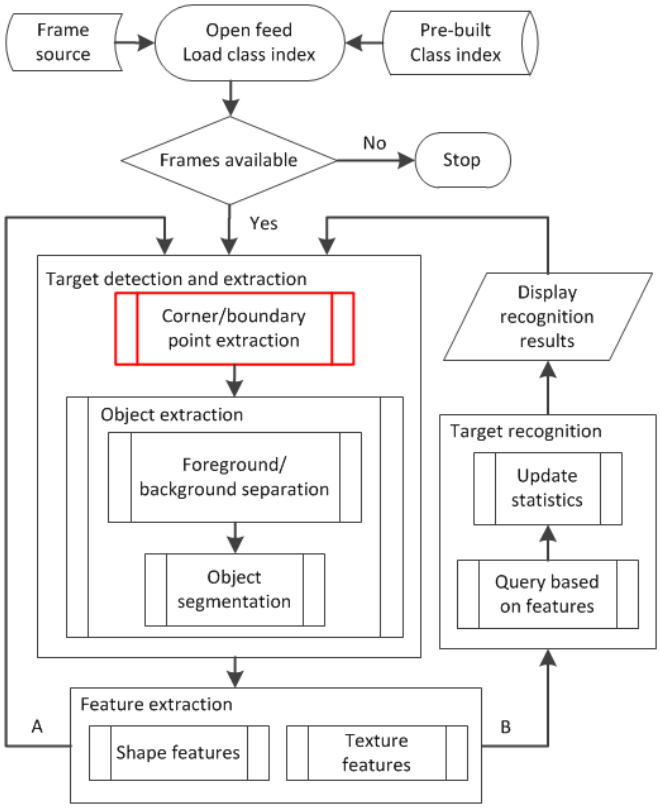


Figure 3.17: Sequence diagram of the whole approach. Branches A and B run in parallel.

3.5.3.1 Feature point extraction and target detection for multiple objects

For the extraction of the object silhouettes, Harris based feature point detector was applied, which is extended to handle multiple objects with different properties (e.g. size, color, shape, texture), on a changing cloudy sky background. The detector was previously tested for singular object recognition, therefore the novelty of this section relies in the issue of handling multiple objects simultaneously.

Before introducing the proposed method, some assumptions were formulated:

1. the camera is not required to be static, but the consecutive video frames should contain overlapping parts;

2. the moving objects are smaller than the background part of the frames;

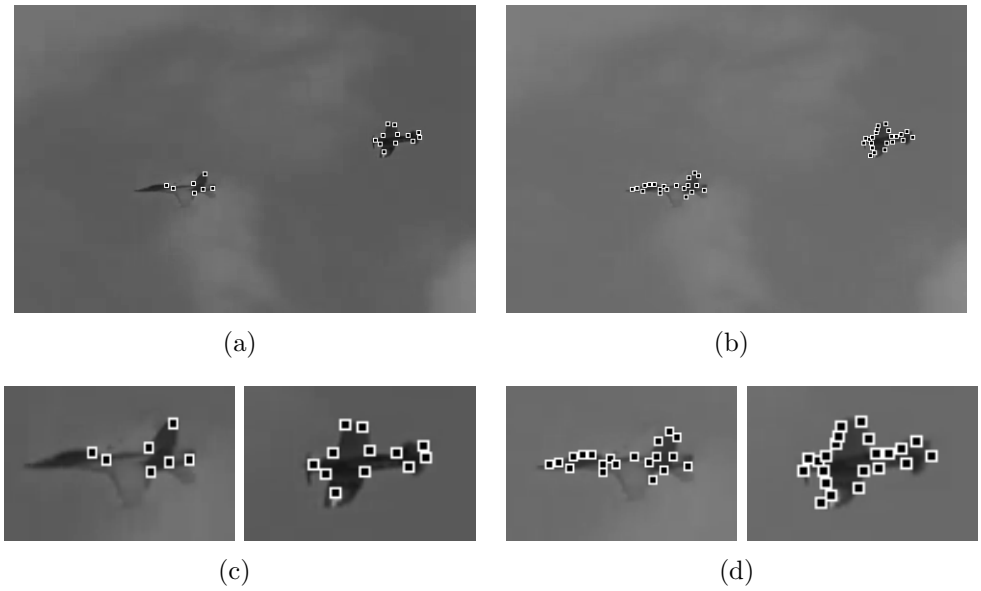


Figure 3.18: Contour point detection. (a): Original Harris corner detector [24]; (b): Proposed MHEC point detector; (c)-(d) show the respective objects zoomed.

3. the background is not required to be completely homogeneous (e.g. clear skies), but should contain large homogeneous areas (which can be sky, clouds, vapor trails, etc.).

The first step of localizing foreground objects is to extract interest/feature points in the image I_t at time t. The challenge in detecting such points is that the contours of the moving foreground objects are usually of low contrast, and contain high curvature parts. Therefore, traditional point detectors, like the Harris corner detector [24], cannot represent them accurately (see Figure 3.18(a)). For this reason, the formerly introduced R_{logmax} function (Equation 3.7) was calculated and feature points were extracted as local maxima of the function. The $\mathbf{C}$ set of contour points is shown in Figure 3.18(b). It is important to note, that the $\mathbf{C}$ point set also emphasizes parts that were dismissed by the original Harris implementation, like the frontal part of the left plane.

Now the $\mathbf{C}$ point set is defining contour points in the image belonging to different flying objects or background. The next step is to separate point subsets of various objects, while eliminating the points of the background.

The separation process of contour point subsets is based on the points connectivity in the Canny edge map [22]. If two contour points are connected by an edge in the edge map, then they are supposed to belong to the same object. The same graph representation is applied as in Section 2.3.2.5: a $\mathbf{G} = (\mathbf{C}, \mathbf{N})$ graph is described with the $\mathbf{C}$ vertex set and the $\mathbf{N}$ edge set, where $\mathbf{C}$ is the feature point set and $\mathbf{N}$ is built according to the connectivity of the vertices (points) in the edge map.

After constructing $\mathbf{N}$, the $\mathbf{G}$ graph will contain K disjoint subgraphs (denoting the k^{th} with $\mathbf{G}^k$) with contour point sets $\mathbf{C}^k$ representing separate objects:

$$\mathbf{C}^k = \left\{c_1^k, \ldots, c_{N_k}^k\right\} \ , \tag{3.17}$$

where N_k is the number of contour points in $\mathbf{G}^k$. Then the following conditions are satisfied by $\mathbf{C}^k$ point subsets:

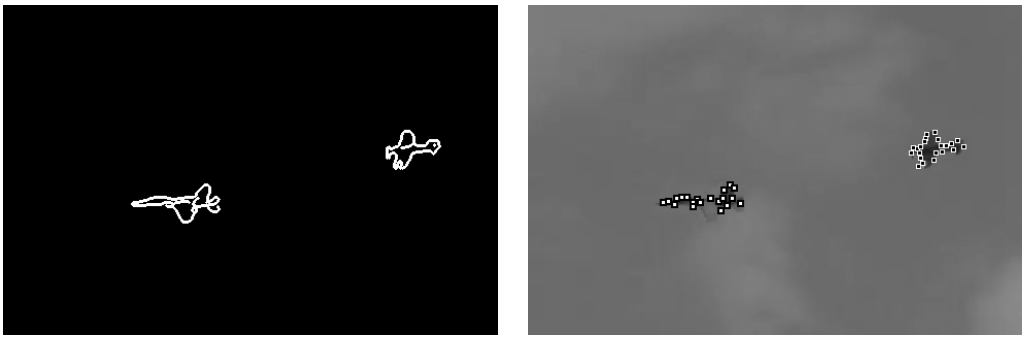
$$\mathbf{C} = \bigcup_{k=1}^{K} \mathbf{C}^k; \quad \mathbf{C}^i \cap \mathbf{C}^j = \emptyset \quad \forall i, j \ . \tag{3.18}$$

Subgraphs containing only a few points are supposed to indicate noise or background, therefore $\mathbf{G}^k$ subgraph was filtered out, if the number of its points is smaller than an n threshold ($N_k < n$). In this work, $n = 4$ was applied. After this filtering, the remaining set of contour points representing K' flying target is given as:

$$\mathbf{C}' = \bigcup_{k=1}^{K'} \mathbf{C}^k \ . \tag{3.19}$$

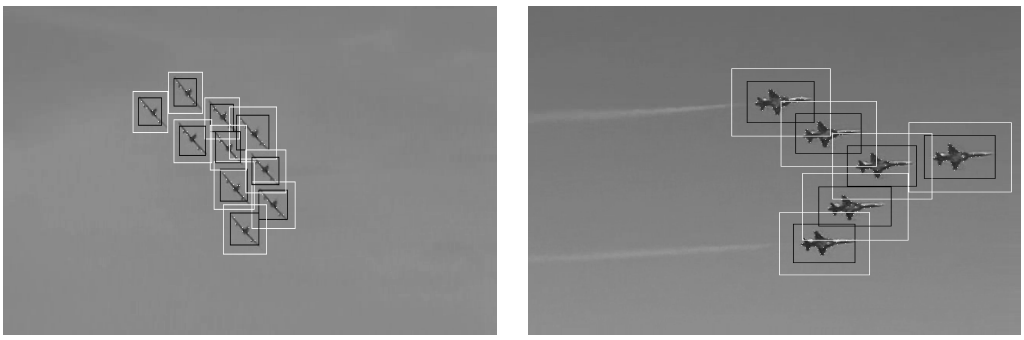
Figure 3.19(b) shows the two separated contour point sets representing the two objects. Multiple flying objects are localized in Figure 3.20.

In situations where objects are so close that they visually occlude one another, the objects might not get separated and the result would be a blob containing both objects. In such situations the latter recognition phase of the complete process (see Figure 3.17) will give a false classification of the blob. However, the recognition step builds a continuous statistics of the detected classes of objects in time, and if the objects will visually separate later, then their classes will be updated. Also, a tracker using these outputs could help in separation of such objects.



(a) (b)

Figure 3.19: Object separation. (a): Canny edge map; (b): Separated object contour points marked differently.



(a) (b)

Figure 3.20: Separation of multiple objects. Flying objects (marked by rectangles) are localized based on the separated feature point subsets. [9]

3.6 Conclusion

In this chapter a novel feature map was introduced, which is able to improve the efficiency of parametric active contour methods (GVF and VFC) when detecting complex boundaries with weak contrast and high curvatures. This feature map is based on a modified Harris characteristic function and describes better the principal curvatures and emphasizes both corners and edges of the object equally. By calculating local maxima of the feature map, a point set is given, which can be also applied successfully in other applications for complex object detection,

like searching for MS lesions in brain MRI image pairs or localization of small flying targets.

Although the method has been extended to multiple object detection, detailed evaluation is required to analyze the efficiency with more complex images, like in case of background clutter. It would be also interesting to extend the method to 3-D. However, the three dimensional representation of Harris corner detector is quite novel [85] and needs more extensive research and evaluation.

Chapter 4

Improved Harris Feature Point Set for Orientation Sensitive Detection in Aerial Images

This chapter addresses automatic detection tasks (urban area and building detection) in remotely sensed images. As manual administration is time consuming and unfeasible, researchers have to focus on automated processing techniques, which can handle various image characteristics and huge amount of data. The introduced method extracts feature points in the first step, which is followed by different detection techniques concentrating on urban areas or buildings. Orientation information in the close proximity of the feature points is exploited for constructing a more accurate voting map to represent urban areas [3] and to extract an orientation sensitive edge/connectivity map, emphasizing edges only in the main directions to detect buildings [11]. This chapter presents methodological contributions in two key issues of the detection process: (1) An automatically extracted, Harris based feature point set is introduced for the first step, which is able to represent urban areas more precisely. (2) Orientation information in the local neighborhood of points is exploited and applied as a new feature for more accurate urban area and building detection. Evaluation results show that the proposed contributions increase the detection accuracy of both urban areas and buildings.

4.1 Introduction

Automatic detection of urban areas and/or buildings in optical aerial images means a great support in urban development analysis, map updating, disaster management and also helps municipalities in long-term residential area planning. Unfortunately, the images may cover large areas and they can be taken in altering weather and illumination conditions, causing very different image features. Moreover, as urban areas are usually changing dynamically, continuous administration is required to have up-to-date information. Since manual administration is time consuming and unfeasible, researchers have to focus on automated processing techniques, which can handle various image characteristics and huge amount of data.

A wide range of automatic urban (a.k.a. built-in) area detection techniques have been introduced in recent years (overview of literature about automatic building detection can be read in Section 3.5.1.1). The first group of such methods needs specific training data to detect urban areas, like [91] using a differential morphological profile to record structural image information, then applied feature extraction and neural network for classifying the features; and [39] presenting a multiple conditional random fields ensemble model to incorporate multiple features and learn their contextual information.

This work follows another methodology, to construct a direct method which does not need any training data for urban area extraction. [92] showed that corner detectors (Harris [24] and SUSAN [100]) are efficient tools for distinguishing different types of structures (man-made versus natural structures) present in the image. Since then, other works also applied interest point detectors for urban area detection: [67] used SIFT (Section 2.2.1) integrated with graph theory. The results were promising, but the computational complexity and time was quite significant. To reduce computational requirements, the same authors introduced a novel technique using Gabor feature points and spatial voting [88].

In this chapter, first an automatically extracted feature point set, called Modified Harris for Edges and Corners (**MHEC**) is introduced for urban area detection, which was introduced for effective object contour detection in Section 3.3.2 and used in different applications (see Section 3.5). As the density of extracted

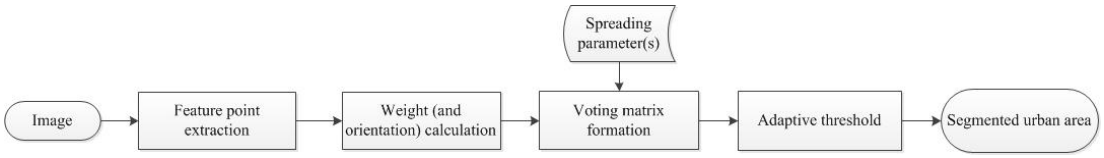


Figure 4.1: Simplified diagram of the workflow of urban area detection.

feature points is higher in the residential areas [92], building a probability map based on this feature point set can help in identifying urban areas and buildings. After having a local feature point set, the voting matrix strategy of [88] is applied to get a probability map of urban area. To improve the accuracy of this step, a novel orientation-sensitive technique is also proposed for constructing the voting matrix. Finally, urban areas are calculated by a decision-making step. Figure 4.1 shows the main algorithmic steps of the method. In the experimental part, it is demonstrated that the introduced method is able to outperform other interest point detectors, moreover the proposed orientation-sensitive voting matrix is also able to improve the performance of the previously introduced Gabor-based algorithm [88] as well.

Orientation sensitivity also exists in case of building detection. According to our assumption, a small urban area has buildings with connected orientation. In most cases, houses are oriented according to some bigger structure (e. g. the road network), therefore main orientation of the area can be defined. As the proposed MHEC point set represents the area efficiently, the idea was to calculate the main direction of the buildings of the urban area based on this point set. Orientation information of the points is calculated based on the edges of the local neighborhood, and main directions in an urban region are determined with statistical Gaussian function fitting.

After defining the main orientation(s), edges in the given direction have to be enhanced and for this issue, shearlet approach [93] is used. Shearlet is a multidimensional version of the traditional wavelet transform. Shearlets, unlike wavelets, are theoretically optimal in representing images with edges and have the ability to fully capture directional features. This technique was used to generate a connectivity map where main directional edges are emphasized.

Based on this connectivity map, the feature point set is divided into subsets, representing building candidates. Finally, Chan-Vese active contour method [47] estimates boundary of the building.

4.2 Modified Harris for edges and corners

The novel MHEC feature point set is introduced here for urban area detection, which adapts the R_{logmax} characteristic function (see Section 3.3.1, Equation 3.7). The advantage of the improved detector is that it is automatic and it is able to recognize not just corners, but edges as well. Therefore, it gives an efficient tool for characterizing contour-rich regions, such as urban areas in aerial images. The improved characteristic function was also successfully applied earlier for object contour recognition with parametric active contour algorithms [1].

Feature points are calculated as local maxima of R_{logmax}. A pixel $p_i = (x_i, y_i)$ is the element of the P feature point set, if it has the largest $R_{\text{logmax}}(p_i)$ 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 $R_{\text{logmax}}(p_i)$ value exceeds a given $T_{\max}$ threshold:

$$P = \left\{ p_i : R_{\text{logmax}}(p_i) > T_{\max} \text{ AND } p_i = \underset{r \in b_i}{\operatorname{argmax}} R_{\text{logmax}}(r) \right\}. \tag{4.1}$$

Here, the $T_{\max}$ threshold is calculated by Otsu's method [82] for each image adaptively, therefore it can handle different image characteristics.

By exploiting the properties of the λs (eigenvalues of the Harris matrix, see Section 3.3.1) and investigating the smaller eigenvalue $\min(\lambda_1, \lambda_2)$, edge and corner points can be separated: if $\min(\lambda_1, \lambda_2)$ is over an adaptive $T_{\min}$ threshold (also calculated by Otsu's method), then the point is a corner point, otherwise it is an edge point. Separation of edge and corner points will be used in Section 4.3.2 for the orientation sensitive urban area detection method.

The proposed feature point detector is completely automatic, does not need any parameter tuning, unlike the referred Gabor mechanism [88]. The method is also not sensitive to the g scale parameter of the Gaussian smoothing used in the Harris corner detection, $g = 0.5$ value is picked for every test case, regardless of resolution or any other image feature. Therefore, it is not presented in Figure 4.1

as input parameter. Additionally, the above $T_{\min}$ and $T_{\max}$ thresholds are defined with the adaptive Otsu's method.

In the next step (Figure 4.1), a weight value (presented in [88]) is calculated for each feature point, that represents the strength of the point's neighborhood. Following [88], the R_{logmax} map is binarized with the adaptive $T_{\max}$ threshold to get a B binary map. In B, pixels of R_{logmax} over $T_{\max}$ get 1, others 0 value. By definition, every member of the P feature point set will have a value of 1 (see Eq. 4.1). Let w_i denote the weight of the p_i point, which is calculated as the natural logarithm of the number of pixels connected to p_i in B binary map. Two pixels are connected in B if both pixels are 1, and a path of pixels with value 1 exists between them. If the number of connected pixels is less than 10 (defined based on the recommendation of [88]), the point is classified as non-significant and eliminated from the feature point set. Points in larger sections, like roads, may have higher weights than others, resulting in an unbalanced weighting function, therefore the reason for applying the natural logarithm function on the calculated weights is to get steady w_i values.

The extended MHEC point set for the $Szada_1$ image can be seen in Figure 4.2(a). The extended set has members in edge regions (like roads) as well and points are densely located in the built-in area. Only a few points are situated in the non-urban area.

4.3 Orientation sensitive urban area extraction

The properties of the extracted point set is applied for urban area detection. In the first step, a novel orientation sensitive voting matrix is constructed, which is able to exploit orientation information of the feature points. After this, the oriented matrix technique is adapted for the novel MHEC feature point set.

4.3.1 Orientation sensitive voting matrix formation

[88] assumes that around Gabor feature points [86] an urban area can be found with a high probability. Thus, it defines a Gaussian domain around each p_i point with σ_i standard deviation, where the highest vote is the point, and according to the spatial distance, the vote is decreasing. σ_i controls the spatial effect of

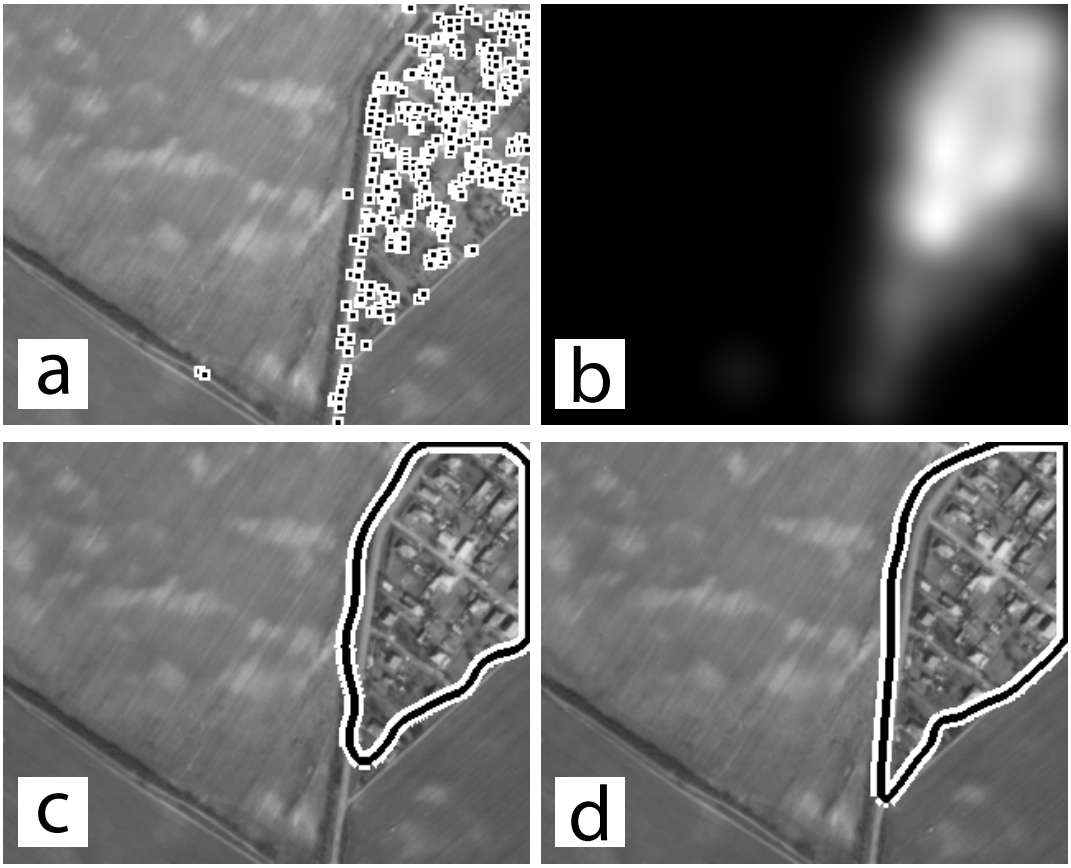


Figure 4.2: Steps of the urban area extraction for the $Szada_1$ image with the proposed MHEC feature point set. (a) Extracted feature point set. (b) Voting matrix of the referred, non-oriented process [88]. (c) Detected urban area applying the non-oriented process. (d) Detected urban area applying the improved, orientation sensitive process.

a point: if it is higher, a point has wider impact. The significance of a point is proportional to its weight, thus σ_i is estimated by w_i. V voting matrix for (x, y) pixel is defined as:

$$V(x,y) = \sum_{i=1}^{K} \frac{1}{2\pi\sigma_i^2} e^{-\frac{(x-x_i)^2+(y-y_i)^2}{2\sigma_i^2}}, \tag{4.2}$$

where K is the total number of feature points and $\sigma_i = 5 \times w_i$. Figure 4.2(b) shows the generated V map for the $Szada_1$ image. Finally, the urban area is defined by an adaptive decision making step. Otsu's thresholding technique [82]

(a) (b) (c)

Figure 4.3: Detection results for the $Szada_5$ image with different voting matrix formation techniques. (a): Original image. (b): Result of the original, non-oriented method [88] (F-measure value: 0.548). (c): Result of the proposed orientation sensitive method (F-measure value: 0.717).

is applied on V to distinguish the urban area from background. The detection result is shown in Figure 4.2(c).

The disadvantage of the method is that it is not able to detect straight lines accurately because of the circular Gaussian representation. If a straight urban area border (like a road segment or a row of houses) is represented by sparse feature points, then the voting matrix will have high values only in the neighborhood of the points, therefore the detection will result in a less accurate wavy urban area border (see Figure 4.2(c)). Moreover, high curvature boundaries are also detected inaccurately.

To compensate these drawbacks, the orientation sensitive extension of the voting matrix is introduced. The idea is to use the direction information of Gabor filters [86], when constructing the voting matrix. Therefore, the two-dimensional elliptical Gaussian function is proposed to be applied at a feature point (x_i, y_i) which can model orientation:

$$G_i(x, y) = A_i \cdot E_i(x, y), \tag{4.3}$$

$$E_i(x, y) = e^{-\left(a_i(x-x_i)^2+2b_i(x-x_i)(y-y_i)+c_i(y-y_i)^2\right)} \tag{4.4}$$

where the calculation of A_i is given in Eq. 4.6, a_i, b_i and c_i are calculated in the

following way:

$$\begin{aligned} a_i &= \frac{\cos^2\theta}{2\sigma_{i,x}^2} + \frac{\sin^2\theta}{2\sigma_{i,y}^2}, \\ b_i &= -\frac{\sin 2\theta}{4\sigma_{i,x}^2} + \frac{\sin 2\theta}{4\sigma_{i,y}^2}, \\ c_i &= \frac{\sin^2\theta}{2\sigma_{i,x}^2} + \frac{\cos^2\theta}{2\sigma_{i,y}^2}, \end{aligned} \tag{4.5}$$

where $\sigma_{i,x}$ and $\sigma_{i,y}$ are the spreads of the blob in x and y direction and θ is the orientation for the (x_i, y_i) feature point.

This interpretation represents a two-dimensional elliptical Gaussian function which is rotated by a clockwise angle θ.

In case of Gabor feature points, θ is defined as the orientation of the Gabor filter, which extracted the feature point. According to [88], ten different directions are chosen for Gabor filtering ($\theta = \{0, \pi/10, 2\pi/10, \ldots, 9\pi/10\}$ rad), therefore if a point was extracted with for example the $\theta = \pi/10$ filter, the orientation for that point is $\theta = \pi/10$. Then the additional direction information can be exploited in the voting matrix formulation by expanding a point's impact in the main θ direction.

Regarding this consideration, the following oriented voting matrix is proposed:

$$V_\theta(x,y) = \sum_{i=1}^{K} G_i(x,y) = \sum_{i=1}^{K} \frac{1}{2\pi(\sigma_{i,x}^2 + \sigma_{i,y}^2)} \cdot E_i(x,y). \tag{4.6}$$

The $\frac{\sigma_{i,y}}{\sigma_{i,x}} = 2$ ratio is set to avoid too moderate voting in the minor axis direction, resulting in false negative hits in the detection step.

The effect of the orientation sensitivity can be seen in Figure 4.3, where the image (b) shows the result of the original, non-oriented method for Gabor points [88], while image 4.3(c) shows the result of the improved, orientation sensitive technique ($\sigma_{i,x} = 3 \times w_i$ and $\sigma_{i,y} = 6 \times w_i$). To quantitatively compare the results, the gained F-measure scores are also presented. (For further evaluation details, see Section 4.4.)

4.3.2 Orientation sensitive voting matrix formation for the novel feature point set

In case of a Gabor feature point set, defining a point's orientation does not need any extra computation, therefore it is favorable to exploit the additional information. But there may be other interest point detectors where orientation has to be calculated in a distinct step. Moreover, in some cases, no such direction can be interpreted for points without further consideration. For example, when using the Harris corner detector, more than one main direction can be assigned to a point, which may pose many interpretation questions. (How to decide if a direction is significant? How many significant directions have to be calculated?)

In the present case, the MHEC feature point set proposed in Section 4.2 will consist of both edge and corner points. Thus, calculating multiple directions featuring corner points is problematic. Meanwhile, edge points represent a main orientation, which can be defined for example by analyzing the small neighborhood, as in [87]. Therefore, the proposed solution is to apply a mixture of voting matrices (V non-directional and V_θ directional components). After dividing the feature point set into two subsets (corner and edge points) based on λ eigenvalues (as described in Section 4.2), the original circular Gaussian field (V in Eq. 4.2) is calculated for corner points and the improved, orientation sensitive, elliptical field (V_θ in Eq. 4.6) for edge points. The final voting matrix is then given as the pixel-wise sum of the two pre-calculated matrices. Figure 4.2(d) shows the result of the joint detection technique for the proposed MHEC point set.

4.4 Experiments

For testing the performance of the proposed feature point set and the improved orientation sensitive voting matrix, the *Szada* dataset was used, which was provided by the Hungarian Institute of Geodesy Cartography and Remote Sensing (FÖMI) (earlier used in [40]). The *Szada* test set contains *eleven* aerial images taken by FÖMI in 1984, 2000 and 2005, showing Szada village and its close proximity. These aerial images have different spatial resolution, 1.5 m/pixel or 0.5 m/pixel; images are either grayscale or colored. The size of the images is varying

from 320×256 to 996×558 pixels. Therefore, the dataset contains diverse images to represent different urban region characteristics and to show the robustness of the proposed techniques.

In the first part of the evaluation process, the efficiency of different interest point detectors is investigated for urban area detection with the original non-oriented voting matrix technique. In the second part, the performance of the improved, orientation sensitive voting matrix is tested for selected feature point detectors.

For comparing the results, the P precision, R recall and F-measure values were used:

$$P = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}, \quad R = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}, \quad F = 2 \cdot \frac{P \cdot R}{P + R}, \tag{4.7}$$

where TP, FP and FN denote the number of true positive, false positive and false negative detected pixels respectively.

When defining an urban area, various interpretations might be considered:

- Are roads and gardens parts of urban area?
- Should buildings be marked separately or collectively?

Feature point detector	Average F-measure Score
MHEC (proposed)	$\mathbf{0.801 \pm 0.044}$
FAST [99]	0.792 ± 0.072
SUSAN [100]	0.778 ± 0.084
LoG [101]	0.750 ± 0.061
SIFT [19]	0.670 ± 0.093
Harris [24]	0.667 ± 0.154
Gabor [88]	0.646 ± 0.124

Table 4.1: Average F-measure Score (mean ± standard deviation) for the evaluated feature point detector methods for *Szada* dataset.

(a) (b)

(c) (d)

Figure 4.4: Ground truth results for $Szada_5$ image: (a)–(c) images were generated by three different individuals; (d) is the ground truth used for evaluation based on majority voting of previous images.

To handle these subjective cases, three individuals were asked to label the urban area manually. If a pixel was labeled as urban by at least two persons, then it was treated as a part of a built-in area in the ground truth based evaluation process (Figure 4.4).

4.4.1 Tests on different interest point detectors

To evaluate the performance of the proposed feature point detector, the original voting matrix technique was tested for different point detectors. Since there exists a large number of point detection methods [90], different detectors were tested to cover a wide range of detection techniques. Harris [24] and SUSAN [100] were proposed in [92] for extraction of man-made structures; previous works of Sirmaçek and Ünsalan used Lowe's SIFT [19] method in [67], Gabor filtering [86] in [88], FAST [99] and Lindeberg's blob detector (LoG) [101] in [89], thus these detectors were chosen for evaluation and comparison.

Table 4.1 shows the calculated average F-measure values (Eq. 4.7) for different interest point detection methods. The proposed MHEC method outperforms the other detectors and achieves more than 0.8 for mean F-measure value.

Computation time was calculated for the MHEC feature point extraction step on the $Szada_1$ test image (with 320×256 size) and altogether 0.6 s has been achieved on a PC with an Intel(R) Core™ i7 2.67GHz CPU with 4 GB RAM and MATLAB R2011b. This is a fairly good time, compared to [88], where Gabor filtering and local feature point extraction steps needed approximately 1.5 s for an image with similar size (235×265).

4.4.2 Tests on orientation sensitivity

To test the orientation sensitivity, such interest point detectors were chosen, where the orientation is reasonable. As mentioned in Section 4.3.2, this additional information cannot be interpreted for every detector (like Harris detector or blob detectors) and sometimes needs more consideration (like MHEC).

For the originally used Gabor points, orientation information for a point is given when determining the direction of the Gabor filter. Therefore, exploiting orientation information does not need any additional computation. Furthermore, detectors based on edge enhancing techniques extract edge points as well, thus can be used for orientation-sensitive urban area detection with further considerations.

In our evaluation step, the proposed MHEC and the SUSAN detectors were picked, to show the performance of the orientation-sensitive voting matrix. In both cases, the mixture of non-oriented and oriented voting matrix techniques

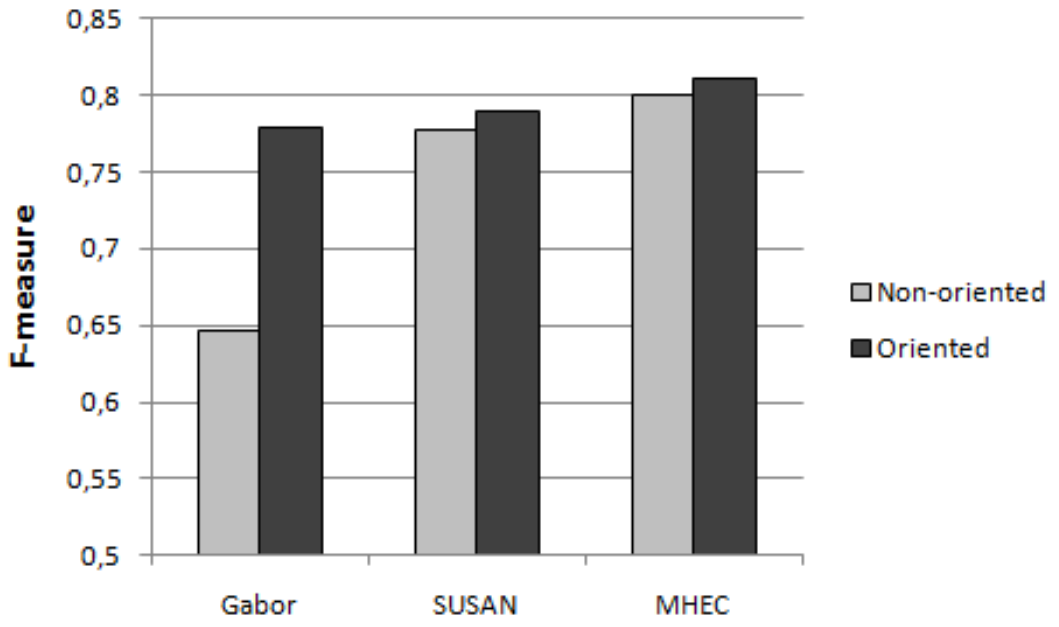


Figure 4.5: Detection results based for different voting matrix techniques. Left: Original, non-oriented [88]. Right: Proposed orientation sensitive.

(see Section 4.3.2) were used, as edge and corner points can be separated. The separation procedure for the MHEC method was described in Section 4.2. For the SUSAN method, the main principle can be used for both edge and corner detection purposes. Depending on the value of the geometric threshold, the algorithm is able to find edge or corner points (see [100] for further details), therefore it can be used to distinguish edges and corners.

Figure 4.5 shows the comparison of the performance for the three selected algorithms. The light bar is the original F-measure performance value, achieved by the referred, non-oriented method, while the dark one is the result achieved by the improved, orientation sensitive voting technique.

Spreading parameter values of the orientation sensitive voting matrix, $\sigma_{i,x}$ and $\sigma_{i,y}$ were selected based on the number of detected points and on the distribution of weight values. While the $\frac{\sigma_{i,y}}{\sigma_{i,x}}$ ratio handles the orientation sensitivity, the shape of the effect (depending on the number of points), $\sigma_{i,x}$ and $\sigma_{i,y}$ values are responsible for the coverage of a point (depending on the variance of weight values).

In case of Gabor points, $\sigma_{i,x} = 3 \times w_i$ and $\sigma_{i,y} = 6 \times w_i$ values were applied. As this detector gives the highest number of points and the weight values have large variance, orientation and saliency is represented effectively by the points.

In case of SUSAN, the oriented $\sigma_{i,x}$, $\sigma_{i,y}$ (for edge points) and the non-oriented σ_i (for corner points) values were tuned manually with the following restrictions: $\frac{\sigma_{i,y}}{\sigma_{i,x}} = 2$ ratio and $\sigma_{i,x} < \sigma_i < \sigma_{i,y}$ (a typical spreading parameter setting was $\sigma_{i,x} = 8 \times w_i$, $\sigma_{i,y} = 16 \times w_i$ and $\sigma_i = 14 \times w_i$ used for $Szada_1$ in Fig. 4.2). These considerations were defined empirically to balance the effect of corner and edge points. The variance of the weights was much lower than in case of Gabor detector, therefore larger multipliers were selected in case of $\sigma_{i,x}$ and $\sigma_{i,y}$ to exaggerate the saliency effect. The exact parameter values also depended on the resolution of the image, the higher the resolution, the wider spatial effect for a point has to be selected. Therefore, in case of images with 0.5 m/pixel resolution, a typical parameter setting $\sigma_{i,x} = 12 \times w_i$, $\sigma_{i,y} = 24 \times w_i$ and $\sigma_i = 20 \times w_i$ was applied.

In case of MHEC, the behavior of the detector was similar to the Gabor case, but the number of points was lower. Therefore, the $\frac{\sigma_{i,y}}{\sigma_{i,x}}$ ratio was set higher, to represent orientation sensitivity more efficiently, $\sigma_{i,x} = 2 \times w_i$, $\sigma_{i,y} = 6 \times w_i$ and $\sigma_i = 3 \times w_i$ was applied for lower resolution and $\sigma_{i,x} = 2 \times w_i$, $\sigma_{i,y} = 8 \times w_i$ and $\sigma_i = 6 \times w_i$ for higher resolution. As the variance of the weight values was high, the multipliers of w_i weight values are smaller than in case of SUSAN.

According to Figure 4.5, the orientation sensitive representation was able to improve the performance of urban area detection. The improvement was the most significant for Gabor points, where no additional calculation was needed. Moreover, it caused a slight increase in the performances of SUSAN and MHEC as well, enhancing accuracy altogether with 17% compared to previous method.

4.5 Orientation based building outline extraction

In the previous evaluation part, orientation was proved to be an efficient feature for detection tasks. In the present section, this novel information will be applied for the more complex building detection task [11], [13].

4.5.1 Orientation estimation

A small urban area has buildings with connected orientation. In most cases, houses are oriented according to some bigger structure (e. g. the road network), therefore the main orientation(s) of the area can be defined. As the proposed modified Harris point set represents the area, the idea was to calculate the main direction(s) of the buildings of the area based on this point set.

4.5.1.1 Unidirectional urban area

To confirm the assumption about connected orientation feature of closely located buildings, specific images were used in the first step of the work, presenting only small urban areas and having only one main direction. Such an image can be seen in Figure 4.7(a).

[74] used a low level feature, called local gradient orientation density, where the surroundings of a pixel was investigated whether it has perpendicular edges or not. Now, this method is adapted to calculate the main direction of a feature point's neighborhood. 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), the weighted density of φ_i^{∇} is as follows:

$$\lambda_i(\varphi) = \frac{1}{N_i} \sum_{r \in W_n(i)} \frac{1}{h} \cdot \|\nabla g_r\| \cdot k\left(\frac{\varphi - \varphi_r^{\nabla}}{h}\right), \tag{4.8}$$

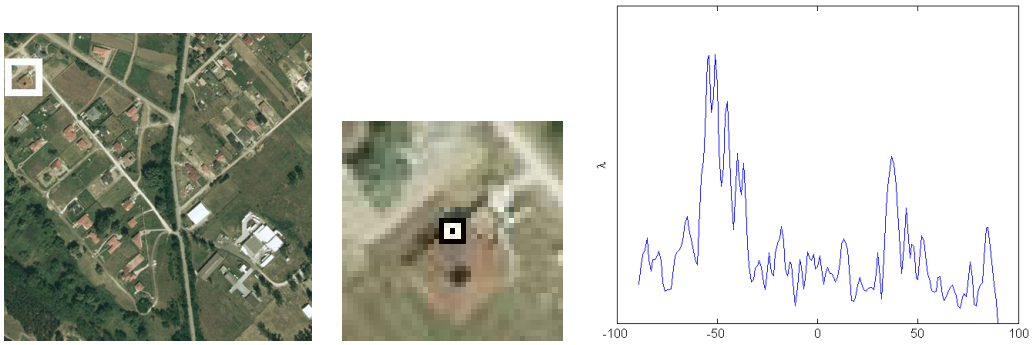
with $N_i = \sum_{r \in W_n(i)} \|\nabla g_r\|$ and $k(.)$ kernel function with h bandwidth parameter (See Figure 4.6).

Now, the main orientation for (i^{th}) feature point is defined as:

$$\varphi_i = \underset{\varphi \in [-90,+90]}{\operatorname{argmax}} \{\lambda_i\}. \tag{4.9}$$

After calculating the direction for all the K feature points, the density function ϑ of their orientation is defined:

$$\vartheta(\varphi) = \frac{1}{K} \sum_{i=1}^{K} H_i(\varphi), \tag{4.10}$$



(a) (b) (c)

Figure 4.6: Local gradient orientation density ($\lambda_i(\varphi)$ function) for the ith feature point : (a) is the original image denoting the neighborhood of the feature point by a white rectangle; (b) is the cropped image showing the neighborhood of the point; (c) shows the $\lambda_i(\varphi)$ function for the feature point, with $\varphi_i = -51$.

where $H_i(\varphi)$ is a logical function:

$$H_i(\varphi) = \begin{cases} 1, & \text{if } \varphi_i = \varphi \\ 0, & \text{otherwise} \end{cases} \tag{4.11}$$

The density function ϑ is expected to have two main peaks (because of the perpendicular edges of buildings), see Figure 4.7. This can be measured by correlating ϑ to a bimodal density function:

$$\alpha(m) = \int \vartheta(\varphi)\eta_2(\varphi, m, d_\vartheta)\, d\varphi, \tag{4.12}$$

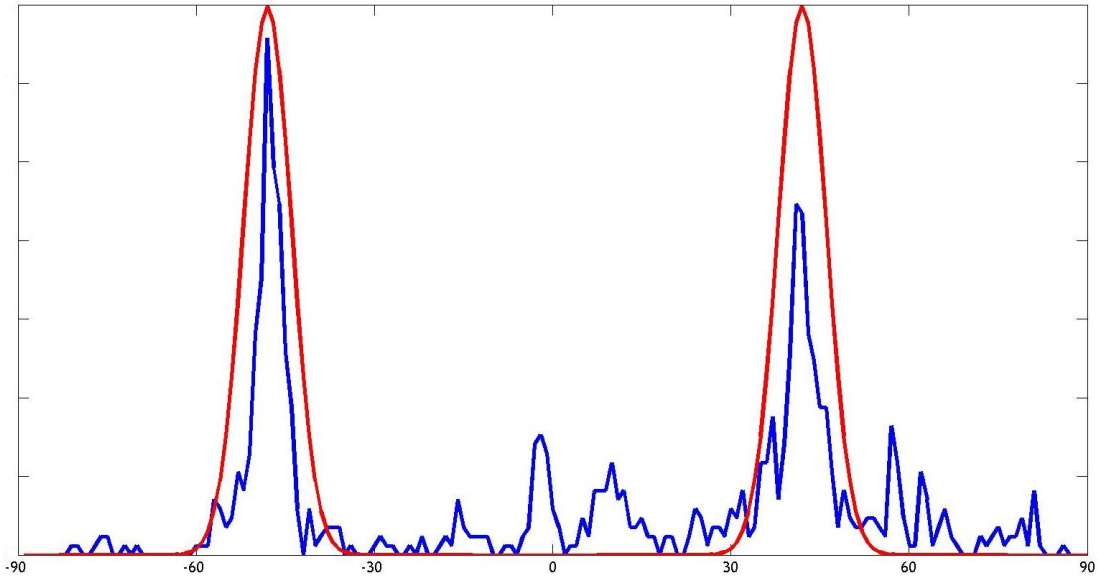
where $\eta_2(.)$ is a two-component Mixture of Gaussian (MG), with m and $m + 90$ mean values and d_ϑ is standard deviation for both components. The value θ of the maximal correlation can be obtained as:

$$\theta = \underset{m\in[-90,+90]}{\operatorname{argmax}} \{\alpha(m)\}. \tag{4.13}$$

And the corresponding orthogonal direction (the other peak):

$$\theta_{\text{ortho}} = \begin{cases} \theta - 90, & \text{if } \theta \geq 0 \\ \theta + 90, & \text{otherwise} \end{cases} \tag{4.14}$$

(a) (b)



(c)

Figure 4.7: Orientation estimation for a unidirectional image: (a) is the original image; (b) shows the feature points in yellow; (c) shows the $\vartheta(\varphi)$ orientation density function of the points in blue, calculated for 15×15 neighborhood. $\varphi \in [-90, +90]$ is the horizontal axis, the number of points is the vertical axis. The $\eta_2(.)$ two-component Mixture of Gaussian is in red, detected peaks are $\theta = -47$ and $\theta_{\text{ortho}} = +53$.

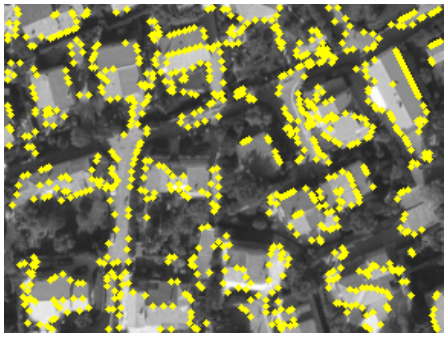
Thus, building edges are expected to be in the calculated main orientation and edges in the given directions have to be emphasized. The orientation estimation step has been tested for different datasets and different window sizes ($n \times n$ neighborhood). The results show that with larger size $n = 15$, the density function is smoother (see Figure 4.7(c)), while with smaller size ($n = 5$) rougher and blurred, but the main characteristics and the main peaks were obvious in both cases, therefore only the result of 15×15 neighborhood is shown in Figure 4.7.

4.5.1.2 Orientation based classification

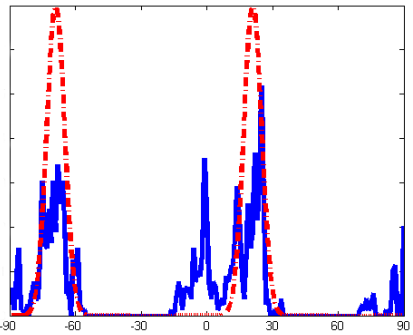
If the urban area is larger, there might be building groups with multiple orientations. However, the buildings are still oriented according to some bigger structure (road network) and cannot be located arbitrarily, orientation of the closely located buildings is coherent. In this case the ϑ density function of the φ_i values is expected to have more peak pairs: $2q$ peaks ($[\theta_1, \theta_{\text{ortho},1}], \ldots, [\theta_q, \theta_{\text{ortho},q}]$) for q main directions. As the value of q is unknown, it has to be estimated by correlating multiple bimodal Gaussian functions to the ϑ density function. The correlation is measured by $\alpha(m)$ (see Eq. 4.12), therefore the behavior of α values has been investigated for increasing number of $\eta_2(.)$ two-component MG functions. When the number of the correlating bimodal MGs is increasing, the α value should also be increasing or remaining nearly constant (a slight decreasing is acceptable), until a correct estimation number is reached, or the correlating data involves enough points (the number of correlated points has reached a given ratio), the ratio in this case has been set to 95%. Based on these criteria the value of the α_q parameter and the total number of the Correlated Points (CP_q) are investigated when correlating the data to q bimodal MGs.

Figure 4.8 shows the steps of defining the number of main directions (q). The calculated MHEC points for the image is in Figure 4.8(b), including altogether 790 points. The correlating bimodal MGs and the belonging parameters are in Fig. 4.8(c)-4.8(e). As one can see, the α_q parameter is increasing continuously and the CP_q parameter has reached the defined ratio (95%) in the second step (representing $768/790 \approx 97\%$ of the point set). The third MG (Fig. 4.8(e)) is just added for illustrating the behaviour of the correlation step: although α_q is still increasing, the newly correlated point set is small, containing only $CP_3 - CP_2 = 18$

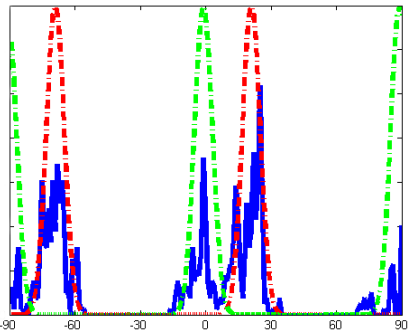
(a) Original image



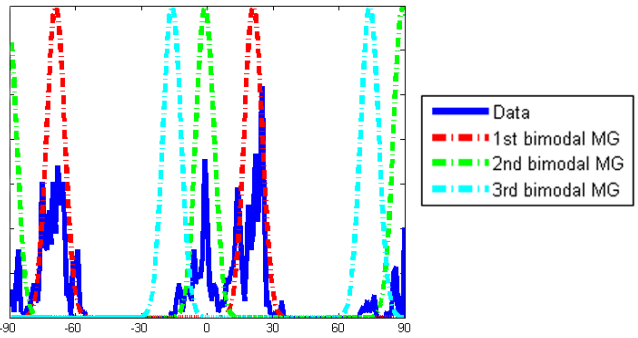
(b) MHEC point set ($\sum$ 790 points)



(c) 1 correlating bimodal MG:
$\alpha_1 = 0.042$; $CP_1 = 558$

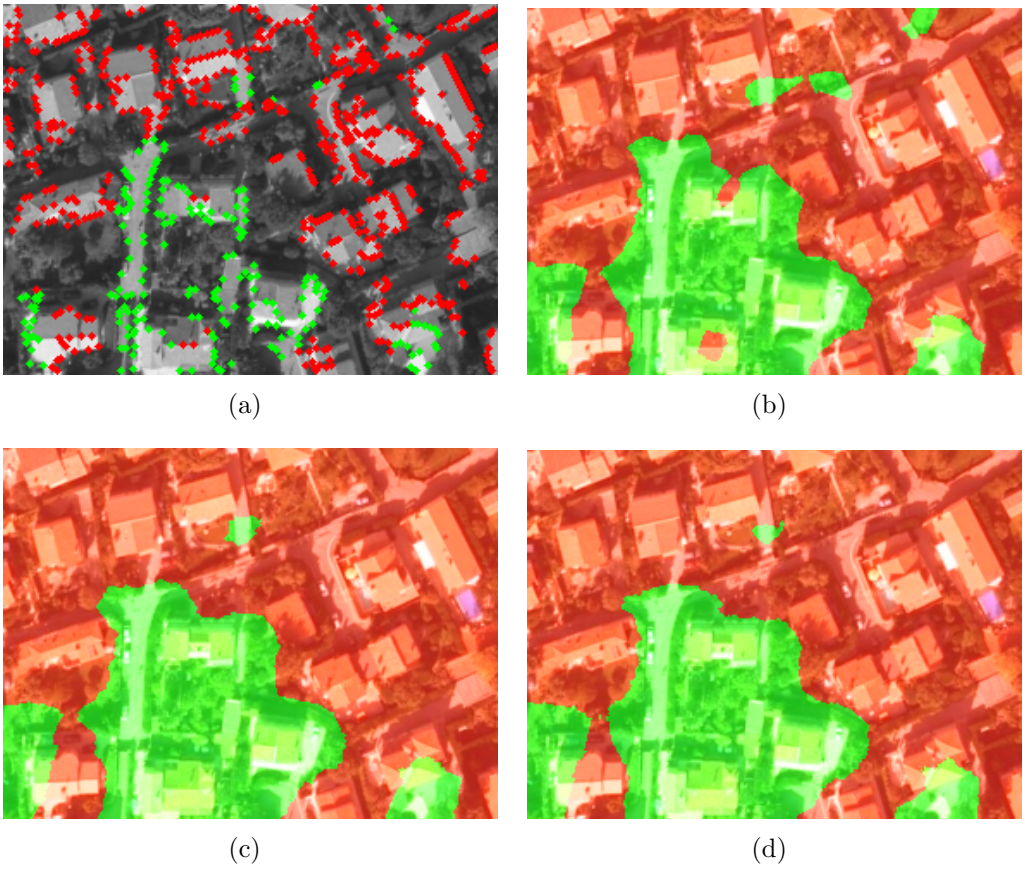


(d) 2 correlating bimodal MGs:
$\alpha_2 = 0.060$; $CP_2 = 768$



(e) 3 correlating bimodal MGs:
$\alpha_3 = 0.073$; $CP_3 = 786$

Figure 4.8: Correlating increasing number of bimodal mixture of Gaussians (MGs) with the ϑ orientation density function (marked in blue). The measured α_q and CP_q parameters are represented for each step. The third component is determined insignificant, as it covers only 18 MHEC points. Therefore the estimated number of main orientations is $q = 2$.



(a) (b)

(c) (d)

Figure 4.9: Orientation based classification for $q = 2$ main orientations with k-NN algorithm for image 4.8(a): (a) shows the classified MHEC point set, (b)–(d) is the classified image with $k = 3$, $k = 7$ and $k = 11$ parameter values. Different colors show the clusters belonging to the bimodal GMs in figure 4.8(d).

points. Therefore, the estimated number of main orientation is $q = 2$, with peaks $\theta_1 = 22$ ($\theta_{1,\text{ortho}} = -68$) and $\theta_2 = 0$ ($\theta_{2,\text{ortho}} = 90$).

The point set is then classified by K-means algorithm, where K is the number of main orientation peaks ($2q$) and the distance measure is the difference between the orientation values. After the classification, the 'orthogonal' clusters (2 peaks belonging to the same bimodal MG component) are merged, resulting in q clusters. The result of this step is in Figure 4.9(a).

The orientation based classification is then extended to the whole image, k-NN clustering is performed to classify the image pixelwisely. Classification has been

tested with different k values (3, 7 and 11), Figure 4.9(b)–(d) show the results respectively, different colors mark the clusters with different orientations. The same color is picked for the correlating bimodal MG-s in Figure 4.8(d) and for the area belonging to the corresponding cluster in Figure 4.9. The tests have proved that the classification results are not sensitive to the k parameter, therefore in the further evaluation, a medium value, $k = 7$ was chosen.

The classification map defines the main orientation for each pixel of the image, therefore in the edge detection part, connectivity information in the given direction has to be extracted.

4.5.2 Edge detection with shearlet transform

Now, the main orientations have been defined. The next step is to enhance edges in these directions to extract useful connectivity information from the image that can be later combined efficiently with the feature point set. There are different approaches which uses directional information like Canny edge detection [22] using the gradient orientation; or [94] which is based on anisotropic diffusion, but cannot handle the situation of multiple orientations (like corners). Other single orientation methods exist, like [95] and [96], but the main problem with these methods is that they calculate orientation in pixel-level and lose the scaling nature of orientation, therefore they cannot be used for edge detection. In the present case, edges constructed by joint pixels has to be enhanced, thus the applied edge detection method has to be able to handle orientation. Moreover, as searching for building contours, the algorithm must handle corner points as well. Shearlet transform [93] has been lately introduced for efficient edge detection, as unlike wavelets, shearlets are theoretically optimal in representing images with edges and, in particular, have the ability to fully capture directional and other geometrical features.

For an image u, the shearlet transform is a mapping:

$$u \rightarrow SH_{\psi}u(a, s, x), \tag{4.15}$$

providing a directional scale-space decomposition of u with $a > 0$ is the scale, s

is the orientation and x is the location:

$$SH_\psi u(a,s,x) = \int u(y)\psi_{as}(x-y)dy = u * \psi_{as}(x), \tag{4.16}$$

where ψ_{as} are well localized waveforms at various scales and orientations. When working with a discrete transform, a discrete set of possible orientations is used, for example $s = 1,\ldots,16$. In the present case, the main orientation(s) of the image θ are calculated (see Section 4.5.1), therefore the aim is to strengthen the components in the given directions on different scales as only edges in the main orientations have to be detected. The first step is to define the s subband for image pixel (x_i, y_i) which includes θ_i and $\theta_{i,\text{ortho}}$:

$$\begin{aligned}\tilde{s}_{1,\ldots,q} &= \left\{ s_i : (i-1)\frac{2\pi}{s} < \theta_{1,\ldots,q} \leq i\frac{2\pi}{s} \right\}, \\ \tilde{s}_{1,\ldots,q,\text{ortho}} &= \left\{ s_j : (j-1)\frac{2\pi}{s} < \theta_{1,\ldots,q,\text{ortho}} \leq j\frac{2\pi}{s} \right\}.\end{aligned} \tag{4.17}$$

After this, the $SH_\psi u(a, \tilde{s}_{1,\ldots,q}, x)$ and $SH_\psi u(a, \tilde{s}_{1,\ldots,q,\text{ortho}}, x)$ subbands have to be strengthened at (x_i, y_i). For this reason, the weak edges (values) have been eliminated with a hard threshold and only the strong coefficients are amplified.

Finally, the shearlet transform is applied backward (see Eq.4.16) to get the reconstructed image, which will have strengthened edges in the main directions. The strengthened edges can be easily detected by Otsu thresholding, results can be seen in Figure 4.10 for an urban area with one main direction. While the pure Canny method detects the edges sometimes with discontinuities, the shearlet based edge strengthening helps to eliminate these problems and the given result represents connectivity relations efficiently.

As building colours may vary largely and the shadow effect have to be reduced to eliminate false contours when detecting buildings, two color channels were used in the edge strengthening step: for red buildings the u* component of CIE L*u*v advised in [69]; for grey buildings the C_b component of YC_bC_r space was applied, which was found to separate grey coloured objects and their shadows the most effectively in [98]. First, the red building outlines have been extracted, based on the u* edge map, after that the remaining buildings were detected in the C_b-based map.

(a) (b) (c)

Figure 4.10: Comparing the edge maps for u* channel: (b) shows the result of the pure Canny edge detection; (c) is the result of the shearlet based edge strengthening.

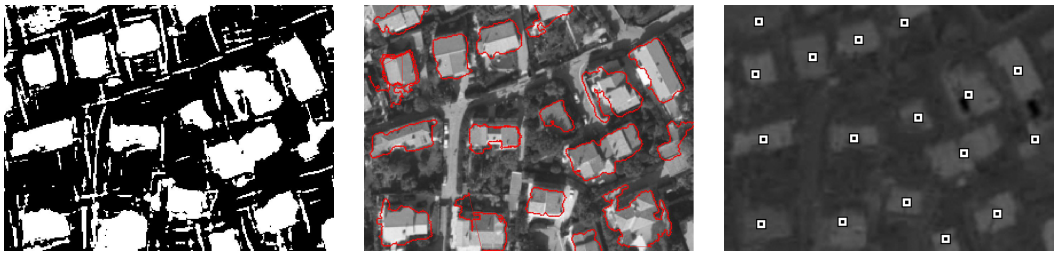
As the u* channel emphasizes the red roofs as well, the Otsu adaptive thresholding may also detect these pixels with high intensity values in the edge strengthened map (see Figure 4.11(a)), therefore the resulted edge map is better to be called as a connectivity map.

4.5.3 Building contour detection

After defining the main orientation for the extended Harris point set for the urban region, the shearlet based edge strengthening approach enhanced the edges in the main directions, resulting in a connectivity map S. In the next step, the feature point set and connectivity map will be fused with a graph based representation, which was introduced in previous sections; based on the generated edge map, connected feature point subgraphs are determined, indicating building candidates. The E edge network of $G = (V, E)$ graph is constructed by connecting $v_i = (x_i, y_i)$ and $v_j = (x_j, y_j)$, the i^{th} and j^{th} vertices of the V feature point set, if they satisfy the following conditions:

1. $S_{(x_i,y_i)} = 1$,
2. $S_{(x_j,y_j)} = 1$,
3. a finite path exists between v_i and v_j in S .

The result of this a graph composed of many separate subgraphs, where each subgraph indicates a building candidate. However, there might be some singular



(a) (b) (c)

Figure 4.11: Steps of multidirectional building detection: (a) is the connectivity map; (b) shows the detected building contours in red; (c) marks the estimated location (center of the outlined area) of the detected buildings.

points and some smaller subgraphs (points and edges connecting them) indicating noise. To discard them, only subgraphs having points over a given threshold are selected.

To detect the accurate contour of the buildings, Chan-Vese active contour algorithm [47] is applied and the contour for a building candidate is initialized as the convex hull of the vertices of the subgraph.

Main directional edge emphasis may also enhance road and vegetation contours, moreover some feature points can also be located on these edges. Therefore after the contour extraction step the results have to be supervised to filter out misdetections. When detecting false objects, like road parts or land section borders, the edges in the detected area are unidirectional, unlike buildings, which have either orthogonal or multidirectional contours. Thus, the directional distribution of edges is evaluated in the extracted area (see the technique in Section 4.5.1) and unidirectional hits are eliminated. Here, the correlation to a bimodal density function (Eq. 4.12) is measured again and the α value is thresholded to select multidirectional hits.

Figure 4.11 and 4.12 show the result of the building detection with the detected and filtered contours. Based on the contours, the location of the buildings can be estimated which will be useful in the further work for evaluation and comparison.

(a) (b)

Figure 4.12: Result of the building detection with one main direction: (a) shows the detected contours; (b) is the estimated locations of the detected buildings.

4.6 Experiments

The proposed method has been evaluated for the Szada dataset provided by the Hungarian Institute of Geodesy, Cartography and Remote Sensing. This dataset was also used in [74] for evaluation and for comparison with different methods ([67], [89], [97] and [66]), therefore quantitative test results are available.

There is a wide range of publications in remote sensing topic for urban area and building detection. However, only novel approaches which will be used for comparison. State-of-the-art building detection approaches can be divided into two main groups as it was already written in Section 3.5.1.1. The first group contains methods which only localize buildings without any shape information. From this group two methods have been selected for comparison.

In [67] a SIFT salient point based approach is introduced for urban area and building detection (denoted by SIFT-graph). This method uses two templates (a light and dark one) for detecting buildings. After extracting feature points representing buildings, graph based techniques are used to detect urban area. The given templates help to divide the point set into separate building subsets, then the location is defined. However, in many cases, the buildings cannot be represented by such templates, moreover sometimes it is hard to distinguish them from the background based on the given features.

[89] proposes a method to detect building positions in aerial and satellite images based on Gabor filters (marked as Gabor filters), where different local feature

Szada dataset (57 buildings)	Missing objects	False objects
SIFT-graph [67]	17	26
Gabor filter[89]	17	23
Features-Canny [66]	10	18
Segment-Merge [97]	11	5
bMBD [74]	4	1
Proposed	**2**	**0**

Table 4.2: Quantitative results for Szada dataset

vectors are used to localize buildings with data and decision fusion techniques.

The other group contains approaches which use some shape templates (e.g. rectangles) for detecting the buildings. In this case, beside the location, additional information is given about the size, orientation and shape. The following three polygon-fitting approaches were evaluated and compared to the proposed method.

In [97] a segment-merge technique is introduced (Segment-Merge), which represents a distinct trend. This method considers building detection task as a region level problem and assumes that buildings are homogeneous areas (either regarding color or texture information), and based on this fact, they can be distinguished from the background. In the first step, the background is subtracted, then some shape and size constraints are created to define building objects. However, the basic assumptions influences the success of the approach: sometimes buildings cannot be distinguished from the background effectively by using color and texture features, therefore the further steps will also fail.

[66] (named as Features-Canny) combines roof color, shadow and edge information in a two-step process. First, a built-in candidate is defined based on color and shadow feature, then a rectangle template is fitted using a Canny edge map. This sequential method is very sensitive to the deficiencies of both steps: the inappropriate shadow and color information results in false candidates, and accurate detection is not possible with a malfunctioning edge map.

A novel building detection approach is introduced in [74], using a global optimization process, considering observed data, prior knowledge and interactions

(a) (b) (c)

Figure 4.13: Qualitative comparison of MPP-based and proposed method: (a) is the original image part; (b) shows the result of MPP-based method; (c) is the result of the proposed approach.

between the neighboring building parts (marked later as bMBD). The method uses low-level (like gradient orientation, roof color, shadow, roof homogeneity) features which are then integrated to have object-level features. After having object (building part) candidates, a configuration energy is defined based on a data term (integrating the object-level features) and a prior term, handling the interactions of neighboring objects and penalizing the overlap between them. The optimization process is then performed by a bi-layer multiple birth and death optimization.

In Table 4.2 the quantitative results for Szada dataset is shown. The complete dataset contains 57 buildings out of which our method is able to detect 55 buildings (meaning 2 misdetections) with 0 false positive object. In this case only the location of the buildings (see Figure 4.12(b)) was used for evaluation. By comparing this with the other approaches, one can see that the proposed method is able to outperform the others.

For qualitative evaluation the detected outlines (see Figure 4.12(a)) were compared. A part of the image was cut and enlarged to compare the proposed method qualitatively with [74]. Figure 4.13 shows the detailed differences. However, rectangular templates provide a very close estimation for the shape of the buildings, the fine details are lost. Unlike shape templates, active contour based techniques do not apply any restrictions for the shape and able to detect the varying contour parts more accurately.

Although active contours are able to cope with the altering shapes, sometimes they suffer from the lack of contrast difference between the building and the background and have difficulties when detecting contours (like missing a part of the building outline, see the building in the bottom-middle of Figure 4.12(a)).

4.7 Conclusion

In this chapter a new feature point set has been introduced for the urban area detection. Moreover, orientation was proved to be an efficient tool for detection tasks in aerial images: either for urban area extraction or building outline detection without using any shape templates.

Chapter 5

Conclusions

This thesis have presented contributions in three main tasks of automatic detection. Novel features have been introduced which were proved to enhance object detection accuracy. The given solutions can all be labeled as techniques for object featuring, but the distinct aims and applications (like extraction, tracking, change detection) needed different tools and developments, tests in artificial and real images showed the advantages of the given contributions.

Active contour theory was the central element of the work, as it is a widely applied technique for detection purposes. We have confirmed with quantitative experiments that by introducing a novel feature map, the accuracy of detecting complex contours is enhanced compared to traditional algorithms. The proposed feature map has other advantages, it can be adapted for feature point selection to achieve automatic active contour initialization and efficient object representation as well. Local characteristics of the feature points was proposed to be applied for creating low dimensional descriptors for different computer vision tasks. We have also introduced a novel feature for aerial image analysis, which can be adapted in other frameworks as well for better object representation.

Our models have been tested on a wide range of images and we have compared the proposed algorithms to other state-of-the-art approaches of the selected fields. We have also given many real life applications, where the introduced contributions are beneficial.

Summary

5.1 Methods used in the experiments

The majority of my work is connected to shape analysis with Active Contour method. This approach is energy minimization, driven by different forces, representing the characteristics of the deformable curve and the image as well:

$$E = \int_0^1 \frac{1}{2}(\alpha \left|\mathbf{x'}(s)\right|^2 + \beta \left|\mathbf{x''}(s)\right|^2) + E_{\text{ext}}(\mathbf{x}(s))\mathrm{d}s, \qquad (5.1)$$

where α and β are weighting parameters for the elasticity and rigidity components of the internal energy; $\mathbf{x'}(s)$ and $\mathbf{x''}(s)$ are the first and second order derivatives with respect to s. In my work, I concentrated on the improvement of E_{ext}, the external energy, derived from the image.

Contributions of the thesis are presented in low-level shape description with Fourier methods, efficient feature point detection and feature extraction, edge detection by shearlets (wavelets) combined with mathematical toolkits: classification and graph theory.

Images used for evaluation in the thesis are partially coming from publicly available image datasets (Brodatz, Weizmann dataset). In *Task 1* I also used a video set taken by an outdoor surveillance camera of a city police central. Magnetic Resonance Imaging (MRI) scans used in *Task 2* were provided by Péter Barsi, MR Research Center, Semmelweis University. Airborne images used in the evaluations in the different tasks were bought from the Hungarian Institute of Geodesy, Cartography and Remote Sensing (FÖMI).

The software design and implementation was performed in Matlab environment. The thesis and my corresponding publications were written in LaTeX.

5.2 New scientific results

1. Thesis: *I have proposed a novel, active contour-based descriptor set for characterizing the neighborhood for scale-invariant feature points. I have experimentally shown that local contour generated around a feature point efficiently represents the main characteristics of the local neighborhood and the local contour descriptor, retrieved as a low dimensional interpretation of the local contour, is an efficient compressed descriptors which can be adapted for computer vision tasks (e.g. tracking, classification and change detection).*

Related publications: [4],[6], [7], [12], [14].

Related section(s) in the dissertation: 2.2.

Describing local patches to register image feature points (keypoints) is an important task for many applications in computer vision. When searching for an efficient descriptor, the task is twofold: features must describe the characteristic patches at a high efficiency, while the dimensionality should be kept at a manageable low value.

By investigating the potential applicability of methods in which some formal meaning of the local properties is represented at a reduced dimension, I have found that active contours generated around keypoints (called local contours) can be applied for designing an efficient descriptor set for comparing image regions. However, as the dimensionality of contour descriptor was high, I have introduced a feature set with reduced dimension, by characterizing the contour with Fourier descriptors and keeping the main coefficients. I have shown that the novel, low dimensional descriptor set can be efficiently adapted for different computer vision tasks. Point matching is performed by using a distance metric for the comparison of introduced descriptors.

For texture classification, I have introduced dynamic radius by investigating the variance of Fourier coefficients, to find the optimal size of point surroundings where the local contour has to be calculated. For change detection, I have presented an application to detect structural changes between registered aerial image pairs based on local contour descriptors. Test results confirmed that local contour descriptors can be comparable features against compressed descriptors, while the meaningful interpretation can help to design better keypoint descriptors.

2. Thesis: *I have introduced a novel feature map and experimentally confirmed that it can be applied efficiently in the energy term of parametric active contour methods for detecting noisy and high curvature object contours. The introduced method is based on my proposed modification of the traditional Harris detector's characteristic function.*

Related publications: [1], [2], [3], [5], [8], [9], [10], [12], [15].

Related section(s) in the dissertation: 3.3.

Deformable active contour (snake) models are efficient tools for object boundary detection. Existing alterations of parametric models have reduced sensitivity to noise, parameters and initial location, but high curvatures and noisy, weakly contrasted boundaries cause difficulties for them.

To address the limitation of initialization and curvature sensitivity, I have investigated the energy minimization process of the active contour theory and introduced a novel feature map for Gradient Vector Flow (GVF) and Vector Field Convolution (VFC) methods applied in the external energy part of the energy function (Eq. 5.1). The proposed modification is based on the Harris detector's characteristic function and it is able to emphasize high and low curvatures steadily. The experimental results confirmed that the proposed methods outperforms previous active contour models and detects high curvatures more accurately.

2.1. I have shown that the feature points of the intro-

duced map can be applied for accurate object localization and initialization of iterative contour detection. I have improved the method given in the thesis, to handle multiple objects simultaneously by separating point sets representing different objects, adapting the method for multi-target tracking.

Related section(s) in the dissertation: 3.3, 3.5.3.

Curve initialization is a challenging task, existing representations either take shape information into account or extract the focus area to define the region of interest, but in case of the detection of randomly shaped objects, the initial outline is usually defined with human interaction.

I have used local maxima of the introduced feature map as feature points and generated the convex hull of the point set to initialize a starting curve around the object. I have extended the introduced method to handle multiple objects simultaneously, by separating the feature point using graph methodology.

Tests, aiming to localize small objects in noisy background, showed that the given technique can successfully reach the required goals.

2.2. By combining the feature map with the local contour descriptors, introduced in Thesis 1, I have given a model for detecting structural changes in image pairs scanned with long time difference. I have tested the introduced method on single channel brain MRI image pairs for detecting appearing malignant lesions.

Related section(s) in the dissertation: 3.5.2.

Change detection is a crucial step for monitoring applications, where the changes may refer to important actions. The challenge resides in the altering image characteristics, which makes registration and detection more difficult.

I have given an automatic structural change detection method for long time-span image pairs, using the introduced feature map for ro-

bust difference image calculation. The local maxima of the difference image are change keypoint candidates and local contour descriptors are generated in their surroundings to measure the change rate and separate misregistration errors and real changes.

The method has been tested on single channel brain MRI image pairs to focus the radiologist's attention to appearing malignant lesions. Comparison of the introduced method with previous lesion detections on artificial and real images, confirmed the advantages of the proposed model.

2.3. Based on the analysis of many airborne images, I have revealed that the proposed feature point set represents built-in areas more precisely, than other point sets extracted by existing feature and corner point detector methods.

Related section(s) in the dissertation: 4.2.

Automatic detection of urban areas in optical aerial images means a great support in a wide range of applications, like urban development analysis, map updating, disaster management. I have applied the introduced feature point set for representing urban areas. I have built a probability map based on this point set and performed a decision-making step to identify urban areas. In the experimental part, I have demonstrated that the introduced feature point set enhances the detection accuracy versus other interest point detectors.

3. Thesis: *I have shown that orientation is an efficient feature for urban area characterization in airborne images. I have developed novel, orientation sensitive models for enhancing the localization of built-in areas and detection of building contours without shape templates, by estimating the main directions of the area surrounding feature points.*

Related publications: [3], [11], [13].

3.1. I have developed an orientation sensitive model, by improving the method used in Thesis 2.3. with inserting the orientation of feature point surroundings. By applying the

improved model for different feature point extraction methods, I have revealed that using orientation as a feature enhance the accuracy of built-in area detection. I have experimentally shown that the improved model applying the feature point set introduced in Thesis 2.1. combined with the novel feature outperforms previous techniques.

Related section(s) in the dissertation: 4.3.

Orientation has an important role when detecting residential areas. Orientation information of feature points is calculated based on the edges of the local neighborhood. I have shown that, inserting this information into the previously used probability model and introducing a novel orientation-sensitive voting map system, increases the accuracy of urban area detection.

Experiments showed that orientation-sensitivity is an efficiently adaptable feature and improves the performance of multiple feature point detector. Tests have also confirmed that the feature point set introduced in Thesis 2.1. fused with the orientation feature obtains the highest detection accuracy compared to previously used point detectors both using and missing orientation information.

3.2. I have given a novel, orientation sensitive model for detecting object contours without shape templates in airborne images. I have shown that by building a statistical model using the orientation information extracted from the surrounding of feature points, the main directions, representing the objects in the image, can be defined and more specific local features can be gained. I have experimentally confirmed the benefits of the proposed approach over previously used building detection methods, either detecting purely location or using shape templates.

Related section(s) in the dissertation: 4.5.

A small urban area has buildings with connected orientation, therefore this feature can also be efficient for building detection. In most cases, houses are oriented according to some bigger structure (e. g. the road

network), therefore orientation of such structure should be analyzed and used for complex tasks.

I have given an orientation sensitive building detection model. First, I have extracted the orientation information based on the introduced feature point set. Then, main directions representing an urban region are determined with bimodal Gaussian function fitting based on the orientation distribution. Edges in the defined directions are enhanced and for this issue, shearlet based edge detection is used. By fusing point and edge information, building candidates are initialized and active contour is applied to detect accurate boundaries.

The proposed model have been compared to previously used algorithm, giving either purely object location or adapting shape templates (like rectangles). Experiments showed that besides it's technical advantages (no templates and accurate detecting), the performance of the method is better than previous approaches.

5.3 Examples for application

The introduced algorithms and methods can all be integrated in complex systems for different computer vision tasks, like surveillance and change detection, medical and airborne image analysis, force protection and defense applications.

The selected tasks were corresponding to either ongoing research projects or cooperation with other institutes.

The aim of project No. 76159 of the Hungarian Scientific Research Fund (OTKA) is the analysis of structural information in the space of sensor networks, to measure and extract valuable information which can be used as a feature set for definite problems, like detecting the important changes in a dynamic scene.

OTKA project No. 80352 is about coherent attributes for interpreting the visual world and its perception, by addressing the fundamental

problem of automatic extraction of visual information from raw sensory data, giving coherent models for sensory understanding the visual world by investigating the human vision for solving special tasks like the analysis of medical or aerial images.

The aim of the cooperation with MR Research Center, Semmelweis University was to give an automatic support for Multiple Sclerosis lesion detection by focusing the radiologist's attention with providing a list of hypothetical lesions, corresponding to significant, but not anatomical changes.

Flying target detection and recognition is an important task for defense applications. The goal of the Multi Sensor Data Fusion Grid for Urban Situational Awareness (MEDUSA) project of the European Defence Agency was to realize a multi-sensor data fusion grid to improve situational awareness and Command & Control in the context of force protection in the urban environment. MEDUSA has also analyzed, developed and applied algorithms that facilitate usage across a range of different types of sensors, fusing the information obtained from them.

References

The author's journal publications

[1] **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. 3, 4.2, 5.2

[2] L. Kovacs, **A. Kovacs**, A. Utasi, and T. Sziranyi, "Flying target detection and recognition by feature fusion," *Optical Engineering*, vol. 51, no. 11, pp. 117002–1–13, 2012. 3, 3.5, 5.2

[3] **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. 4, 5.2, 5.2

The author's international conference publications

[4] **A. Kovacs** and T. Sziranyi, "Local contour descriptors around scale-invariant keypoints," in *Proceedings of IEEE International Conference on Image Processing*, (Cairo, Egypt), pp. 1105–1108, 2009. 2.1, 3.5.2, 5.2

[5] **A. Kovacs** and T. Sziranyi, "High definition feature map for GVF snake by using Harris function," in *Advanced Concepts for Intelligent Vision Systems, Lecture Notes in Computer Science 6474*, (Sydney, Australia), pp. 163–172, 2010. 3, 3.1, 3.3.1, 3.5.1.2, 5.2

[6] **A. Kovacs** and T. Sziranyi, "New saliency point detection and evaluation methods for finding structural differences in remote sensing images of long

time-span samples," in *Advanced Concepts for Intelligent Vision Systems, Lecture Notes in Computer Science 6475*, (Sydney, Australia), pp. 272–283, 2010. 2.3, 5.2

[7] **A. Kovacs** and T. Sziranyi, "Shape detection of structural changes in long time-span aerial image samples by new saliency methods," in *ISPRS Workshop on Modeling of Optical Airborne and Space Borne Sensors*, vol. XXXVIII-1/W17, (Istanbul, Turkey), 2010. 5.2

[8] **A. Kovacs**, C. Benedek, and T. Sziranyi, "A joint approach of building localization and outline extraction," in *IASTED International Conference on Signal Processing and Pattern Recognition*, (Innsbruck, Austria), pp. 721–113, 2011. (document), 3, 3.5, 3.8, 5.2

[9] **A. Kovacs**, A. Utasi, L. Kovacs, and T. Sziranyi, "Shape and texture fused recognition of flying targets," in *Proceedings of Signal Processing, Sensor Fusion, and Target Recognition XX, at SPIE Defense, Security and Sensing*, vol. 8050, (Orlando, Florida, USA), pp. 80501E–1–12, 2011. (document), 3, 3.5, 3.20, 5.2

[10] **A. Kovacs** and T. Sziranyi, "Improved force field for vector field convolution method," in *Proceedings of IEEE International Conference on Image Processing*, (Brussels, Belgium), pp. 2853–2856, 2011. 3, 3.1, 5.2

[11] **A. Kovacs** and T. Sziranyi, "Orientation based building outline extraction in aerial images," in *ISPRS Annals of Photogrammetry, Remote Sensing and the Spatial Information Sciences (Proc. ISPRS Congress)*, vol. I-7, (Melbourne, Australia), pp. 141–146, 2012. 4, 4.5, 5.2

[12] **A. Kovacs** and T. Sziranyi, "Automatic detection of structural changes in single channel long time-span brain MRI images using saliency map and active contour methods," in *Proceedings of IEEE International Conference on Image Processing*, (Orlando, Florida, USA), pp. 1265–1268, 2012. 3, 3.5, 5.2, 5.2

[13] **A. Kovacs** and T. Sziranyi, "Multidirectional building detection in aerial images without shape templates," in *ISPRS Workshop on High-Resolution Earth Imaging for Geospatial Information*, (Hannover, Germany), 2013. accepted. 4.5, 5.2

The author's other publications

[14] **A. Kovacs** and T. Sziranyi, "Detecting boundaries of structural differences in long time-span image samples for remote sensing images and medical applications," in *8th Conference of the Hungarian Association for Image Processing and Pattern Recognition*, (Szeged, Hungary), 2011. 5.2

[15] **A. Kovacs** and T. Sziranyi, "Új típusú, Harris függvény alapú tulajdonságtérkép és ponthalmaz objektumok körvonalának megkeresésére," in *9th Conference of the Hungarian Association for Image Processing and Pattern Recognition*, (Bakonybél, Hungary), 2013. 5.2

[16] A. Utasi and **A. Kovacs**, "Recognizing human actions by using spatio-temporal motion descriptors," in *Advanced Concepts for Intelligent Vision Systems, Lecture Notes in Computer Science 6475*, (Sydney, Australia), pp. 366–375, 2010.

Publications related to the dissertation

[17] M. Kass, A. P. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *International Journal of Computer Vision*, vol. 1, no. 4, pp. 321–331, 1988. 1, 2.2.2, 3.1, 3.2

[18] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proceedings of the International Conference on Computer Vision*, (Corfu, Greece), pp. 1150–1157, 1999. 2.1, 2.2.1

[19] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004. 2.1, 1, 2.2.1, 2.1, 2.2, 2.2.1.4, 3.5.2.1, 4.4, 4.4.1

[20] C. Xu and J. L. Prince, "Gradient vector flow: A new external force for snakes," in *Proceedings of Conference On Computer Vision and Pattern Recognition*, (San Juan, Puerto Rico), pp. 66–71, 1997. (document), 2, 2.2.2, 2.2.2, 3.1, 3.2.1, 3.3.1, 3.3.2, 3.4, 3.3, 3.1, 3.4, 3.5, 3.4.2, 3.2, 3.5.1.1, 3.5.1.8

[21] H. Seo and P. Milanfar, "Using local regression kernels for statistical object detection," in *Proceedings of International Conference on Image Processing*, (San Diego, California, USA), pp. 2380–2383, 2008. 2.1

[22] J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 8, no. 6, pp. 679–698, 1986. 2.1, 2.3.2.5, 3.5.3.1, 4.5.2

[23] S. Belongie, J. Malik, and J. Puzicha, "Shape matching and object recognition using shape contexts," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 24, no. 4, pp. 509–522, 2002. 2.1

[24] C. Harris and M. Stephens, "A combined corner and edge detector," in *Proceedings of the 4th Alvey Vision Conference*, (Manchester, UK), pp. 147–151, 1988. (document), 2.1, 2.2.1.2, 2.3.2, 2.3.2.1, 3.1, 3.3.1, 3.3.1, 3.5.1.1, 3.5.2, 3.18, 3.5.3.1, 4.1, 4.4, 4.4.1

[25] Y. Ke and R. Sukthankar, "PCA-SIFT: A more distinctive representation for local image descriptors," in *Proceedings of Conference on Computer Vision and Pattern Recognition*, vol. 2, (Washington, DC, USA), pp. 506–513, 2004. 2.1, 2.2.1.4

[26] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 27, no. 10, pp. 1615–1630, 2005. 2.1, 2.2.1.4

[27] A. Licsar and T. Sziranyi, "User-adaptive hand gesture recognition system with interactive training," *Image and Vision Computing*, vol. 23, no. 12, pp. 1102–1114, 2005. 4, 2.2.3

[28] C. Zahn and R. Roskies, "Fourier descriptors for plane closed curves," *IEEE Trans. Computers*, vol. 21, no. 3, pp. 269–281, 1972. 2.2.3

[29] Y. Rui, A. She, and T. Huang, "A modified Fourier descriptor for shape matching in MARS," in *Image Databases and Multimedia Search*, pp. 165–180, 1998. 3, 2.2.3

[30] B. Jähne, *Digital image processing (5th revised and extended edition)*. Springer-Verlag, 2002. 2.6

[31] T. Peng, I. H. Jermyn, V. Prinet, and J. Zerubia, "Incorporating generic and specific prior knowledge in a multi-scale phase field model for road extraction from VHR images," *IEEE Trans. Geoscience and Remote Sensing*, vol. 1, no. 2, pp. 139–146, 2008. 2.3.1

[32] F. Lafarge, X. Descombes, J. Zerubia, and M. Pierrot Deseilligny, "Automatic building extraction from DEMs using an object approach and application to the 3D-city modeling," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 63, no. 3, pp. 365–381, 2008. 2.3.1

[33] S. Ghosh, L. Bruzzone, S. Patra, F. Bovolo, and A. Ghosh, "A context-sensitive technique for unsupervised change detection based on hopfield-type neural networks," *IEEE Trans. Geoscience and Remote Sensing*, vol. 45, no. 3, pp. 778–789, 2007. 2.3.1

[34] G. Perrin, X. Descombes, and J. Zerubia, "2D and 3D vegetation resource parameters assessment using marked point processes," in *Proceedings of International Conference on Pattern Recognition*, pp. 1–4, 2006. 2.3.1

[35] R. Wiemker, "An iterative spectral-spatial bayesian labeling approach for unsupervised robust change detection on remotely sensed multispectral imagery," in *Proceedings of Conference on Computer Analysis of Images and Patterns*, pp. 263–270, 1997. 2.3.1

[36] L. Bruzzone and D. F. Prieto, "An adaptive semiparametric and context-based approach to unsupervised change detection in multitemporal remote-sensing images," *IEEE Trans. Image Processing*, vol. 11, no. 4, pp. 452–466, 2002. 2.3.1

[37] Y. Bazi, L. Bruzzone, and F. Melgani, "An unsupervised approach based on the generalized Gaussian model to automatic change detection in multitemporal SAR images," *IEEE Trans. Geoscience and Remote Sensing*, vol. 43, no. 4, pp. 874–887, 2005. 2.3.1

[38] P. Gamba, F. Dell'Acqua, and G. Lisini, "Change detection of multitemporal SAR data in urban areas combining feature-based and pixel-based techniques," *IEEE Trans. Geoscience and Remote Sensing*, vol. 44, no. 10, pp. 2820–2827, 2006. 2.3.1

[39] P. Zhong and R. Wang, "A multiple conditional random fields ensemble model for urban area detection in remote sensing optical images," *IEEE Trans. Geoscience and Remote Sensing*, vol. 45, no. 12, pp. 3978–3988, 2007. 2.3.1, 4.1

[40] C. Benedek and T. Szirányi, "Change detection in optical aerial images by a multi-layer conditional mixed Markov model," *IEEE Trans. Geoscience and Remote Sensing*, vol. 47, no. 10, pp. 3416–3430, 2009. 2.3.1, 4.4

[41] C. Benedek, T. Szirányi, Z. Kato, and J. Zerubia, "Detection of object motion regions in aerial image pairs with a multilayer Markovian model," *IEEE Trans. Image Processing*, vol. 18, no. 10, pp. 2303–2315, 2009. 2.3.1

[42] C. Benedek and T. Szirányi, "Bayesian foreground and shadow detection in uncertain frame rate surveillance videos," *IEEE Trans. Image Processing*, vol. 17, no. 4, pp. 608–621, 2008. 1, 2.3.1

[43] L. Castellana, A. d'Addabbo, and G. Pasquariello, "A composed supervised / unsupervised approach to improve change detection from remote sensing," *Pattern Recognition Letters*, vol. 28, no. 4, pp. 405–413, 2007. 1, 2.3.1

[44] C. Schmid, R. Mohr, and C. Bauckhage, "Evaluation of interest point detectors," *International Journal of Computer Vision*, vol. 37, no. 2, pp. 151–172, 2000. 2.3.2, 3.2.2

[45] V. Caselles, R. Kimmel, and G. Sapiro, "Geodesic active contours," *International Journal of Computer Vision*, vol. 22, no. 1, pp. 61–79, 1997. 3.1

[46] P. Brigger, J. Hoeg, and M. Unser, "B-spline snakes: A flexible tool for parametric contour detection," *International Journal of Computer Vision*, vol. 22, no. 1, pp. 61–79, 1997. 3.1

[47] T. F. Chan and L. A. Vese, "Active contours without edges," *IEEE Trans. Image Processing*, vol. 10, no. 2, pp. 266–277, 2001. (document), 3.1, 3.4, 3.4, 3.4.2, 3.2, 3.5.2.4, 4.1, 4.5.3

[48] A. Vasilevskiy and K. Siddiqi, "Flux maximizing geometric flows," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 24, no. 12, pp. 1565–1578, 2002. 3.1

[49] R. Kimmel and A. M. Bruckstein, "Regularized Laplacian zero crossings as optimal edge integrators," *International Journal of Computer Vision*, vol. 53, no. 3, pp. 225–243, 2003. 3.1

[50] X. Bresson, S. Esedoglu, P. Vandergheynst, J.-P. Thiran, and S. Osher, "Fast global minimization of the active contour/snake model," *Journal of Mathematical Imaging and Vision*, vol. 28, no. 2, pp. 151–167, 2007. 3.1, 3.5.1.1

[51] B. Li and T. Acton, "Active contour external force using vector field convolution for image segmentation," *IEEE Trans. Image Processing*, vol. 16, no. 8, pp. 2096–2106, 2007. (document), 3.1, 3.2.2, 3.3.2, 3.4, 3.3, 3.1, 3.4, 3.5, 3.2, 3.4.2, 3.5.1.1

[52] A. K. Mishra, P. W. Fieguth, and D. A. Clausi, "Decoupled active contour (DAC) for boundary detection," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 33, no. 2, 2011. (document), 3.1, 3.4, 3.4.1, 3.6(b), 3.6, 3.4.2

[53] M. G. Uzunbas, O. Soldea, D. Unay, M. Cetin, G. B. Ünal, A. Erçil, and A. Ekin, "Coupled nonparametric shape and moment-based intershape pose priors for multiple basal ganglia structure segmentation," *IEEE Trans. Medical Imaging*, vol. 29, no. 12, pp. 1959–1978, 2010. 3.1

[54] G. Zhu, S. Zhang, Q. Zeng, and C. Wang, "Gradient vector flow active contours with prior directional information," *Pattern Recognition Letters*, vol. 31, pp. 845–856, 2010. 3.1

[55] G. Sundaramoorthi and A. Yezzi, "Global regularizing flows with topology preservation for active contours and polygons," *IEEE Trans. Image Processing*, vol. 16, no. 3, pp. 803–812, 2007. 3.1

[56] L. Kovacs and T. Sziranyi, "Focus area extraction by blind deconvolution for defining regions of interest," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 29, no. 6, pp. 1080–1085, 2007. 3.1

[57] Y. Wang, L. Liu, H. Zhang, Z. Cao, and S. Lu, "Image segmentation using active contours with normally biased GVF external force," *Signal Processing Letters*, vol. 17, no. 10, pp. 875–878, 2010. 3.1

[58] 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. 3.1

[59] J. Cheng and S. Foo, "Dynamic directional gradient vector flow for snakes," *IEEE Trans. Image Processing*, vol. 15, no. 6, pp. 1563–1571, 2006. 3.1

[60] C. Chuang and W. Lie, "A downstream algorithm based on extended gradient vector flow field for object segmentation," *IEEE Trans. Image Processing*, vol. 13, no. 10, pp. 1379–1392, 2004. 3.1

[61] C. Tauber, H. Batatia, and A. Ayache, "Quasi-automatic initialization for parametric active contours," *Pattern Recognition Letters*, vol. 31, pp. 83–90, 2010. 3.1

[62] S. Alpert, M. Galun, R. Basri, and A. Brandt, "Image segmentation by probabilistic bottom-up aggregation and cue integration," in *Proceedings of Conference on Computer Vision and Pattern Recognition*, pp. 1–8, 2007. (document), 3.1, 3.4, 3.3, 3.1, 3.4.1

[63] N. M. Sirakov, "A new active convex hull model for image regions," *Journal of Mathematical Imaging and Vision*, vol. 26, pp. 309–325, December 2006. 3.3.2

[64] F. Zamani and R. Safabakhsh, "An unsupervised GVF snake approach for white blood cell segmentation based on nucleus," in *Proceedings of the 8th International Conference on Signal Processing*, vol. 2, 2006. 3.3.2

[65] C. B. Barber, D. P. Dobkin, and H. Huhdanpaa, "The quickhull algorithm for convex hulls," *ACM Trans. Mathematical Software*, vol. 22, no. 4, pp. 469–483, 1996. 3.3.2

[66] B. Sirmacek and C. Unsalan, "Building detection from aerial imagery using invariant color features and shadow information," in *International Symposium Computer and Information Sciences*, (Istanbul, Turkey), pp. 1–5, 2008. 3.5.1.1, 3.5.1.5, 4.6

[67] B. Sirmacek and C. Unsalan, "Urban-area and building detection using SIFT keypoints and graph theory," *IEEE Trans. Geoscience and Remote Sensing*, vol. 47, no. 4, pp. 1156–1167, 2009. 1, 2.3.2.5, 3.5.1.1, 4.1, 4.4.1, 4.6

[68] T. Szirányi and Z. Tóth, "Optimization of paintbrush rendering of images by dynamic mcmc methods," in *Proceedings of the Third International Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition*, (London, UK), pp. 201–215, 2001. 3.5.1.2

[69] S. Muller and D. Zaum, "Robust building detection in aerial images," in *ISPRS Workshop on Object Extraction for 3D City Models, Road Databases and Traffic Monitoring*, (Vienna, Austria), pp. 143–148, 2005. 3.5.1.1, 3.5.1.5, 4.5.2

[70] K. Karantzalos and N. Paragios, "Recognition-driven two-dimensional competing priors toward automatic and accurate building detection," *IEEE Trans. Geoscience and Remote Sensing*, vol. 47, no. 1, pp. 133–144, 2009. 3.5.1.1

[71] A. K. Mishra and A. Wong, "KPAC: A kernel-based parametric active contour method for fast image segmentation," *IEEE Signal Processing Letters*, vol. 17, no. 3, pp. 312–315, 2010. 3.5.1.1

[72] C. Benedek, X. Descombes, and J. Zerubia, "Building detection in a single remotely sensed image with a point process of rectangles," in *Proceedings of International Conference on Pattern Recognition*, (Istanbul, Turkey), 2010. (document), 3.5.1.2, 3.5.1.3, 3.7, 3.5.1.4, 3.5.1.5, 3.5.1.5, 3.5.1.6, 3.9

[73] X. Descombes, R. Minlos, and E. Zhizhina, "Object extraction using a stochastic birth-and-death dynamics in continuum," *Journal of Mathematical Imaging and Vision*, vol. 33, pp. 347–359, 2009. 3.5.1.3, 3.5.1.6

[74] C. Benedek, X. Descombes, and J. Zerubia, "Building development monitoring in multitemporal remotely sensed image pairs with stochastic birth-death dynamics," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 1, pp. 33–50, 2012. 1, 2.3.2.5, 3.5.1.2, 3.5.1.3, 3.5.1.4, 3.5.1.5, 3.5.1.6, 4.5.1.1, 4.6, 4.6

[75] R. J. Radke, S. Andra, O. Al-Kofahi, and B. Roysam, "Image change detection algorithms:A systematic survey," *IEEE Trans. on Image Processing*, vol. 14, pp. 294–307, 2005. 3.5.2

[76] F. Rousseau, F. Blanc, J. de Seze, L. Rumbach, and J.-P. Armspach, "An a contrario approach for outliers segmentation: Application to multiple sclerosis in MRI," in *IEEE Int. Symp. on Biomedical Imaging (ISBI)*, (Paris, France), pp. 9–12, 2008. 1, 3.5.2, 3.5.2.5, 3.3

[77] S. Shen, A. Szameitat, and A. Sterr, "Detection of infarct lesions from single MRI modality using inconsistency between voxel intensity and spatial location - A 3D automatic approach," *IEEE Trans. on Information Technology in Biomedicine*, vol. 12, pp. 532–540, 2008. 1, 3.5.2

[78] H. J. Seo and P. Milanfar, "A non-parametric approach to automatic change detection in MRI images of the brain," in *IEEE Int. Symp. on Biomedical Imaging (ISBI)*, (Boston, MA, USA), pp. 245–248, 2009. 3.5.2, 3.5.2.5, 3.3

[79] K. Johnson and J. A. Becker, "The whole brain atlas," 1995–1999. http://www.med.harvard.edu/aanlib/home.html. 3.5.2, 3.5.2.5, (document)

[80] P. J. Kostelec and S. Periaswamy, "Image registration for MRI," *Modern Signal Processing MSRI Publications*, vol. 46, 2003. 3.5.2.1

[81] X. Li and C. Wyatt, "Modeling topological changes in deformable registration," in *IEEE Int. Symp. on Biomedical imaging (ISBI)*, (Rotterdam, The Netherlands), pp. 360–363, 2010. 1, 3.5.2.1

[82] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. Systems, Man and Cybernetics*, vol. 9, no. 1, pp. 62–66, 1979. 3.5.2.3, 4.2, 4.3.1

[83] T. L. Wenga, Y. Y. Wang, Z. Y. Ho, and Y. N. Sun, "Weather-adaptive flying target detection and tracking from infrared video sequences," *Expert Systems with Applications*, vol. 37, no. 2, pp. 1666–1675, 2010. 1, 3.5.3

[84] H. Noor, S. H. Mirza, Y. Sheikh, A. Jain, and M. Shah, "Model generation for video-based object recognition," in *Proceedings of ACM International Conference on Multimedia*, pp. 715–719, 2006. 3.5.3

[85] I. Sipiran and B. Bustos, "Harris 3d: A robust extension of the harris operator for interest point detection on 3d meshes," *The Visual Computer*, vol. 27, no. 11, pp. 963–976, 2011. 3.6

[86] V. Kyrki and J. K. Kamarainen, "Simple Gabor feature space for invariant object recognition," *Pattern Recogn. Lett.*, vol. 25, no. 3, pp. 311–318, 2004. 4.3.1, 4.3.1, 4.4.1

[87] S. Kumar and M. Hebert, "Man-made structure detection in natural images using a causal multiscale random field," in *Proc. IEEE Conf. Comput. Vision Pattern Recogn.*, pp. 119–126, 2003. 1, 4.3.2

[88] B. Sirmaçek and C. Ünsalan, "Urban area detection using local feature points and spatial voting," *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 1, pp. 146–150, 2010. (document), 4.1, 4.1, 4.2, 4.3.1, 4.2, 4.3, 4.3.1, 4.3.1, 4.4, 4.4.1, 4.5

[89] B. Sirmaçek and C. Ünsalan, "A probabilistic framework to detect buildings in aerial and satellite images," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 1, pp. 211–221, 2011. 1, 4.4.1, 4.6

[90] K. Mikolajczyk and C. Schmid, "Scale and affine invariant interest point detectors," *Int. J. Comput. Vision*, vol. 60, no. 1, pp. 63–86, 2004. 4.4.1

[91] J. A. Benediktsson, M. Pesaresi, and K. Arnason, "Classification and feature extraction for remote sensing images from urban areas based on morphological transformations," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 9, pp. 1940–1949, 2003. 4.1

[92] L. Martinez-Fonte, S. Gautama, W. Philips, and W. Goeman, "Evaluating corner detectors for the extraction of man-made structures in urban areas," in *IEEE Int. Geosci. Remote Sens. Symp.*, pp. 237–240, 2005. 4.1, 4.1, 4.4.1

[93] S. Yi, D. Labate, G. R. Easley, and H. Krim, "A shearlet approach to edge analysis and detection," *IEEE Trans. Image Processing*, vol. 18, no. 5, pp. 929–941, 2009. 4.1, 4.5.2

[94] P. Perona, "Orientation diffusions," *IEEE Trans. Image Processing*, vol. 7, no. 3, pp. 457–467, 1998. 4.5.2

[95] R. Mester, "Orientation estimation: Conventional techniques and a new non-differential approach," in *Proc. 10th European Signal Processing Conference*, 2000. 4.5.2

[96] J. Bigun, G. H. Granlund, and J. Wiklund, "Multidimensional orientation estimation with applications to texture analysis and optical flow," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 13, no. 8, pp. 775–790, 1991. 4.5.2

[97] Z. Song, C. Pan, and Q. Yang, "A region-based approach to building detection in densely build-up high resolution satellite image," in *IEEE International Conference on Image Processing*, pp. 3225–3228, 2006. 4.6

[98] V. Tsai, "A comparative study on shadow compensation of color aerial images in invariant color models," *IEEE Trans. Geoscience and Remote Sensing*, vol. 44, pp. 1661–1671, June 2006. 4.5.2

[99] E. Rosten, R. Porter, and T. Drummond, "FASTER and better: A machine learning approach to corner detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 1, pp. 105–119, 2010. 4.4, 4.4.1

[100] S. M. Smith and J. M. Brady, "SUSAN - A new approach to low level image processing," *Int. J. Comput. Vision*, vol. 23, no. 1, pp. 45–78, 1997. 4.1, 4.4, 4.4.1, 4.4.2

[101] T. Lindeberg, "Feature detection with automatic scale selection," *Int. J. Comput. Vision*, vol. 30, no. 2, pp. 77–116, 1998. 4.4, 4.4.1