Detection of Fragmented Rectangular Enclosures in Very-High-Resolution Remote Sensing Images

Zingman Igor, Dietmar Saupe, Otávio A. B. Penatti, and Karsten Lambers

***Abstract*—We develop an approach for detection of ruins of livestock enclosures in alpine areas captured by high-resolution remotely sensed images. These structures are usually of approximately rectangular shape and appear in images as faint fragmented contours in complex background. We address this problem by introducing a rectangularity feature that quantifies the degree of alignment of an optimal subset of extracted linear segments with a contour of rectangular shape. The rectangularity feature has high values not only for perfectly regular enclosures, but also for ruined ones with distorted angles, fragmented walls, or even a completely missing wall. Furthermore, it has zero value for spurious structures with less than three sides of a perceivable rectangle. We show how the detection performance can be improved by learning a linear combination of the rectangularity and size features from just a few available representative examples and a large number of negatives. Our approach allowed detection of enclosures in the Silvretta Alps that were previously unknown. A comparative performance analysis is provided. Among other features, our comparison includes the state-of-the-art features that were generated by pre-trained deep convolutional neural networks (CNN). The deep CNN features, though learned from a very different type of images, provided the basic ability to capture the visual concept of the livestock enclosures. However, our handcrafted rectangularity-size features showed considerably higher performance.**

***Index Terms*—Object detection, incomplete rectangles, man-made structures, maximal cliques, rectangularity feature, deep features.**

I. Introduction

WE ADDRESS the problem of detecting remains of manmade enclosures used to hold livestock in grassland of mountainous regions. The livestock enclosures (LE) are of special archaeological interest because they offer important insights into historical development of alpine pastoralism [1]. Their automated spotting is the goal of a recent archaeological project [2]. Examples of such enclosures are shown in Fig. 1. These structures are usually composed of linear walls that may be heavily ruined. The most common shape of LE resembles a rectangular contour with greatly varying size and aspect ratio. Rectangle angles may deviate from right angles, and rectangle sides may be fragmented. The angle between adjacent fragments of the same side may deviate from 180 degrees. Moreover, the rectangular contours are sometimes incomplete such that even an entire side may be missing.

I. Zingman and D. Saupe are with the Department of Computer and Information Science, University of Konstanz, 78464 Konstanz, Germany (e-mail: igor.zingman@uni-konstanz.de).

O. A. B. Penatti is with the Advanced Technologies Group SAMSUNG Research Institute, Campinas, SP, 13097-160, Brazil.

K. Lambers is with the Faculty of Archaeology, Leiden University, 2333 CC Leiden, Netherlands.

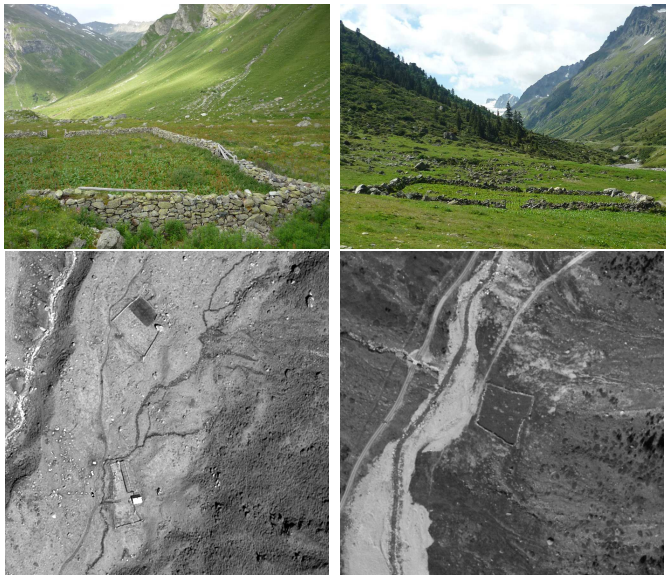

Fig. 1. Above: Livestock enclosures (LE) in alpine environment. Below: 600×600 satellite (©GeoEye 2011) and aerial (SWISSTOPO) images of 0.5m resolution with structures corresponding to livestock enclosures above.

We use satellite and aerial images of 0.5m resolution where the width of linear walls does not exceed two pixels. The ruined walls are of low height, which results in low contrast linear features in the images. The spectral properties of LEs are similar to the spectral properties of the surrounding terrain, rocks, and other irrelevant objects. The second row of Fig. 1 shows a satellite and an aerial image with structures corresponding to the LEs shown above. Nearby irrelevant structures, such as rivers, trails, or rocks, are often of similar or higher contrast either due to larger size (e.g. big rocks) or distinctive spectral properties (e.g. rivers). Detection of such faint enclosures in a complex terrain is a challenging task. Even the detection of easily modeled circular soil structures [3] had very limited success due to their low contrast and complex terrain. Only few examples of LE are available in our case, which presents another difficulty making most approaches that learn from the data inappropriate. Because of these difficulties, commonly used methods for rectangle detection are hardly applicable.

In contrast to spectral properties, the geometrical properties of LEs appear to be more distinctive and do not depend on image modality and conditions under which an image was captured. We therefore develop a measure that quantifies the distinctive geometry of approximately rectangular enclosures. Our approach relies on a new rectangularity feature that

discriminates rectangular patterns from other structures in complex cluttered background. The feature is based on a prior model of a fragmented rectangle, which is a convex polygon with constrained angles.

A. Related work

Detection of rectangular structures has previously been addressed in different contexts. Examples are detection of buildings in remotely sensed images [4]–[14], traffic signs [15]–[17], and particles of a rectangular shape in cryo-electron microscopy images [18], [19]. The methods used were based on Markov Random Fields [9], [15], Marked Point Processes [10], [14], search on a graph [6], [20], Hough Transform and other voting schemes [7], [8], [16]–[18], template matching [21], aggregation of local features [10], [11], [13], and heuristic rules [5].

Most techniques for detection of rectangular structures dealt with buildings in remotely sensed images. For example, in the graph-based approach in [6], a search for cycles was used to generate building hypotheses. The search was accompanied by an extensive set of rules and thresholds, which limits the robustness of the approach. Markov Random Fields (MRF) were used in [9] to delineate buildings. More recently, a similar approach was used in [15] for detection of traffic signs in color images. The approach is sensitive to inaccuracy of extracted edges and cannot detect incomplete rectangles, as it requires the presence of all four sides of a rectangular structure. The marked point processes (MPP) [22] recently became popular for extraction of various structures in remotely sensed images, including buildings (e.g. in [10], [14]). The MPP proved to be very powerful when applied to real data. However, these stochastic methods are still computationally expensive. Similarly to the MRF, they may not converge to a globally optimal solution and usually need careful tuning of a large number of parameters. Attempts have recently been made to address some of these problems, which are crucial for the analysis of large images. In [23] substantial improvements in performance have been achieved for the extraction of line networks (roads and rivers). In this work also the potential of GPUs was efficiently exploited.

An approach for detection of rectangular contours based on the Hough transform was developed in [8]. The approach relies on certain strict geometrical rules making it not suitable for detection of fragmented or incomplete structures. It may also result in detection of rectilinear configurations that cannot form a rectangular contour. Detection of such configurations is prevented in our approach by adding a convexity constraint.

In [11] a set of local features that carried local corner information were used to produce a probability map of building rooftops. Unfortunately, in the case of fragmented enclosures corners are not reliable features. Moreover, local features in general do not suffice in the case of faint contours appearing in a cluttered background. A more global description that takes into account spatial relations between local features is necessary. For example, in [10], [13] the gradient orientation density function (GODF) was computed from image gradients. A correlation of this function with a mixture of two Gaussians having mean values separated by ninety degrees served as a GODF-based feature indicating the presence of buildings.

Although there is a variety of methods developed for building detection, they are not applicable to our task because buildings are much more salient structures. In contrast to building rooftops, walls of ruined livestock enclosures are narrow and are of low height (low contrast features), may be highly fragmented, or even completely missing. Higher contrast irrelevant structures may appear inside or outside of rectangular structures in the immediate neighborhood. Various cues (rooftop color, shadows, 3D cues etc.) usually employed in building detection algorithms are not available.

B. Overview of our approach

Our approach follows the diagram in Fig. 2, which is briefly described below. A binary map of bar edges accompanied by angle information is computed first. Junction points of the medial axis of an inverted binary edge map are detected as candidate points (Sec. II), and a region within an analysis window centered at each candidate point is inspected. A windowed Hough transform is used to find linear segments and model them with a few parameters (Sec. III-A). An undirected graph is then constructed, nodes of which correspond to linear segments and graph edges encode spatial relations between linear segments. Particularly, we use angle and convexity properties to encode spatial relations (Sec. III-B). Due to the construction of the graph, its maximal cliques correspond to valid configurations of linear segments. The valid configurations are then ranked by a new rectangularity measure that encodes the goodness of grouping the segments into a rectangular structure (Sec. III-C). In contrast to [20], the new rectangularity measure does not rely on a heuristic partitioning of the set of linear segments into four subsets. Hard decisions are softened. Configurations better matching a rectangular structure result in a higher rectangularity measure. The rectangularity feature is introduced in Sec. III-D. It is defined as the maximal rectangularity measure of all valid configurations. In practice, the low number of corresponding maximal cliques within the analysis window allows exact maximization, which can efficiently be computed. The resulting rectangularity feature captures the presence of Π-like structures and is robust to their fragmentation. In Sec. IV we show how to improve the detection performance based solely on the rectangularity feature by introducing an additional feature proportional to enclosure size and learning the optimal feature combination from a large number of negative examples and just few positives. In our application we complete the core algorithmic steps summarized in Fig. 2 with a preprocessing stage in the beginning that filters out irrelevant regions (Sec. V), e.g. texture regions, and with a final detector at the end, e.g. a simple thresholding (Sec. IV).

This paper follows our work presented in [24]. Here, we give a more detailed description of the methods, design a more efficient detector of initial candidate locations (Sec. II), report on the results of application of our approach to a large region in the Silvretta Alps (Sec. V), and extend our experimental part (Sec. VI) by comparing the discrimination

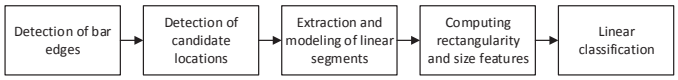

Fig. 2. Algorithmic steps for detection of approximately rectangular enclosures.

ability of the introduced and alternative features for our task. Particularly, we evaluate performance of the features generated by several pre-trained deep convolutional networks [25]–[29], the histogram of oriented gradients (HOG) [30], as well as 1D features, such as the GODF-based feature [10] designed for building detection, and the normalized maximal rectangularity (NMR) measure [20] that we developed earlier for detection of the livestock enclosures. We conclude in Sec. VIII with the discussion on shortcomings and advantages of the rectangularity and deep features for our problem.

II. Detection of candidate locations

As in [20], we detect candidate locations from a map of bar edges (ridges and valleys) that were extracted using the Morphological Feature Contrast (MFC) line detector [31], [32]. This technique extracts linear features, while suppressing texture elements of cluttered background. We also experimented with other approaches [33]–[35], but these are either not sensitive enough to extract faint edges of enclosures, or generate lots of clutter edges depending on the parameters used. The parameterless line segment detector of [36], which is known to provide robust results for a large range of images, misses faint edges of ruined enclosures.

In [20] the candidate points were obtained by sampling the medial axis of an inverted binary edge map. The medial axis was obtained by thinning the inverted edge map. The number of candidates was determined by the choice of the sampling rate. Decreasing the sampling rate reduces the number of candidates, which speeds up further validation. However, this may result in the loss of promising candidates. Instead, we suggest here to look for medial axis junction points only, which are yielded by at least three sides of an enclosure structure. This greatly reduces the number of candidates without the risk of loosing enclosures with at least three remaining sides.

In [37]–[39] the medial axis of a shape was extracted by thresholding the average flux of the gradient field of the Euclidean distance function D to the boundary of the shape. The average flux of the gradient field through the boundary $\partial\mathcal{N}$ of a region is defined as the corresponding flux normalized by the length of the boundary

$$F(\nabla D) = \frac{\oint_{\partial\mathcal{N}} \nabla D \cdot \mathbf{n} \, \mathrm{d}L}{\oint_{\partial\mathcal{N}} \mathrm{d}L}, \tag{1}$$

where $\mathbf{n}$ denotes the inward normal[1] to the boundary $\partial\mathcal{N}$ and $\mathrm{d}L$ is the boundary element. As the region $\mathcal{N}$ shrinks to a point, the average flux F approaches zero at non-medial points and non-zero values at the medial axis of the shape.

We detect candidate points by finding local maxima of a discrete approximation of the average flux $F(\nabla D)$ through the boundary of a small disk $\mathcal{N}$, where D is the distance function of the binary edge map. These local maximum points usually correspond to junction points of the medial axis of the inverted binary edge map. Only local maxima with the average flux greater than 0.5 were taken into account. Fig. 3 shows examples of detections (in red) overlayed on the average flux, which has positive extrema on the medial axis (white) and negative extrema on the bar edges (black)[2].

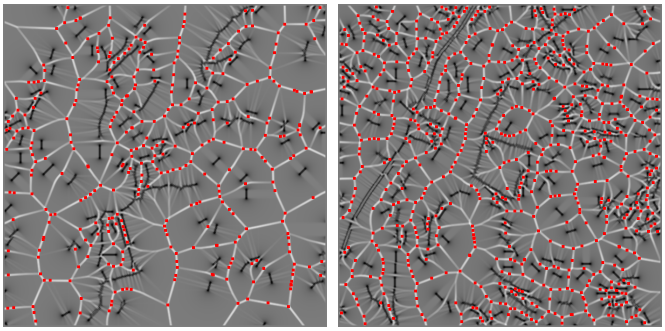

Fig. 3. Average flux of the gradient field of the distance function computed for edge maps of the images in Fig. 1. The medial axis coincides with positive singularities of the flux (white), while edges coincide with negative singularities (black). Local maxima (red points) are used as candidate locations. Best viewed in digital version.

In a related approach [40], non-maxima suppression was applied to the average flux of the normalized gradient vector flow (GVF) [41], [42] in order to detect medial feature points. Using the GVF instead of the gradient field of the distance transform of the edge map allowed detection of medial feature points directly from the gray-scale image without the need of edge extraction. However, GVF may ignore weak gradients of low contrast structures of our interest. In addition, computing GVF might be too slow on large images, depending on the number of predefined iterations.

We combine candidate points separately obtained from the binary maps of ridge and valley edges. The structures that are within a window around the candidate points p are further analyzed. The size of the analysis window can be determined adaptively, based on the value of the distance transform $D(p)$, i.e. the distance of p to a candidate structure. Similarly to [20], we use the circular analysis window of radius $D(p)\sqrt{a^2+1}$ centered at p that circumscribes a rectangle centered at p with (small) side length $2D(p)$ and aspect ratio a. In our experiments a was set to 1.4. We discarded all candidate points having a distance D less than 15 or greater than 90 pixels, which limits the distances between opposite walls of the structures to be in between 7.5m and 45m.

III. Measuring structure rectangularity

We introduce a rectangularity feature f_{R} computed from a set of linear segments $\mathbf{W} = \{\mathcal{S}_i, i = 1, ..., m\}$ that were extracted from a gray-scale image.

A. Grouping edge points into linear segments

Given a candidate location and edge points accompanied by estimated orientations, we extract and parameterize linear

[1] In [37]–[39] the outward normal was used.

[2] We used valley edges for the left figure and ridge edges for the right figure.

segments, each of which is a group of aligned edge points. Linear segments are represented by a triple of parameters (θ, r, l) found by the use of a local Hough transform centered at the candidate points. We use the Hough transform in the form introduced in [43], where a line is defined by the orientation θ of the normal and a distance r from the origin

$$r = x\cos\theta + y\sin\theta. \tag{2}$$

The spatial coordinates of an edge point are x, y. We use the parametrization $\theta \in [0, 360)$ and $r \in (0, \infty)$ of a Hough plane. A peak at (θ, r) in the Hough plane corresponds to a line. The peaks are detected as regional maxima in the Hough plane that was discretized with $\Delta\theta = 3°$ and $\Delta r = 1$ pixel. The detected line corresponds to either a single connected linear segment $\mathcal{S}$, or to several aligned connected components. In the latter case, the connected components with gaps smaller than a predefined threshold (3 pixels in our experiments) are considered a single linear segment (see the segment $\mathcal{S}_j$ in Fig. 4), otherwise they are considered separate linear segments. The parameter l in the triple (θ, r, l) is the number of points that belong to the linear segment. To better relate the parameter l to the length and avoid its dependence on the width of the extracted edges, we perform their thinning [44] prior to clustering in a Hough plane.

Since edges were extracted together with their orientations, r can be directly computed for each edge point (x, y) using Eq. (2). Thus, each edge point votes for a single point in the (θ, r) plane instead of voting for a curve as suggested in [43]. This idea, which was used already in [45] for clustering of short ridge features, considerably eases extraction of meaningful peaks in the Hough plane.

B. Valid configurations of linear segments

Below we define a valid configuration of linear segments $\mathbf{C} \subseteq \mathbf{W}$ that can be a part of a rectangular structure. We require angles $\beta_{k,j}$ between linear segments $\mathcal{S}_k, \mathcal{S}_j \in \mathbf{C}$ to be close to either zero, $180°$, or right angles. An angle tolerance α will be set to control the strictness of the angle constraint. We define $\beta_{k,j}$ as

$$\beta_{k,j} = \min(|\theta_{\mathcal{S}_k} - \theta_{\mathcal{S}_j}|, 360 - |\theta_{\mathcal{S}_k} - \theta_{\mathcal{S}_j}|). \tag{3}$$

Note that $\beta_{j,k} = \beta_{k,j}$ and $\beta \in [0, 180]$, since $\theta \in [0, 360)$.

The angle constraint alone does not suffice to restrict configurations to be perceptually close to rectangles or rectangle parts. We therefore define a second constraint that requires the valid configuration to be nearly convex in the sense that extension of all linear segments of the configuration can form a nearly convex contour. The convexity tolerance t will be defined to control the strictness of the convexity constraint. For a convex configuration of linear segments it is required that a half plane generated by each segment includes all other segments of the configuration. Additionally, we require all these half planes to contain the candidate point around which we search for a rectangular structure. Pair-wise convexity constraints suffice to verify the convexity of a configuration containing the given candidate point. We define the pair-wise convexity measure τ for a pair of linear segments $\mathcal{S}_k, \mathcal{S}_j$, each

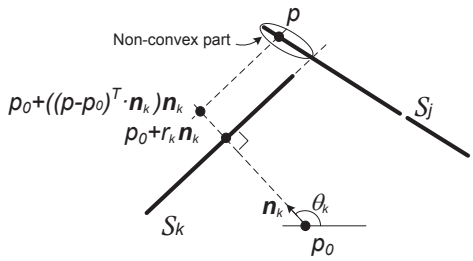

Fig. 4. The fraction of points p of $\mathcal{S}_j$ that violates the convexity constraint relative to $\mathcal{S}_k$ and p_0 is given by $\tilde{\tau}_{k,j}$. Note that linear segments can be fragmented having small gaps as in $\mathcal{S}_j$.

with corresponding attributes of size $l_{\mathcal{S}}$, orientation $\theta_{\mathcal{S}}$, and distance $r_{\mathcal{S}}$ to the candidate point p_0, as

$$\tau_{k,j} = \max(\tilde{\tau}_{k,j}, \tilde{\tau}_{j,k}), \tag{4}$$

$$\tilde{\tau}_{k,j} = \frac{1}{l_j}\sum_{p\in\mathcal{S}_j} H((p - p_0)^T \cdot n_k - r_k), \tag{5}$$

where $n_k = (\cos\theta_k, \sin\theta_k)^T$ is the unit normal of $\mathcal{S}_k$ and $H(u)$ is an indicator function equal one for $u > 0$ and zero otherwise. $\tilde{\tau}_{k,j}$ measures the relative number of points in the segment $\mathcal{S}_j$ that are behind the segment $\mathcal{S}_k$, relative to the given candidate point p_0 as illustrated in Fig. 4. Note that $\tau \in [0, 1]$, and $\tau_{k,j} = \tau_{j,k}$, while $\tilde{\tau}_{k,j} \neq \tilde{\tau}_{j,k}$.

Definition 1. Let $\alpha \in [0, 45], t \in [0, 1]$, a candidate point p_0, and a configuration $\mathbf{C}$ of linear segments be given. If for all pairs $\mathcal{S}_k, \mathcal{S}_j \in \mathbf{C}$, $j \neq k$, one of the inequalities of the angle constraint

$$\beta_{k,j} \leq \alpha \text{ or } |90 - \beta_{k,j}| \leq \alpha \text{ or } 180 - \beta_{k,j} \leq \alpha \tag{6}$$

and the convexity constraint

$$\tau_{k,j} \leq t \tag{7}$$

both hold, then $\mathbf{C}$ is called a (t,α)-valid configuration located around p_0, and denoted by $\mathbf{C}^{t,\alpha}_{p_0}$.

For the sake of brevity, we usually omit the indices t, α and the reference point p_0, mentioning that $\mathbf{C}$ is a valid configuration. Valid configurations include not only perfect rectangles, but also convex polygons or their parts with angles around either 90 or 180 degrees. This is important in practice since approximately rectangular structures are better modeled by such polygons rather than by perfect rectangles.

C. Rectangularity measure of a valid configuration

A couple of poorly aligned short segments can be a valid configuration as far as the tolerances t, α allow. There is a need to rank valid configurations according to their similarity to a canonical rectangle. To find and rank valid configurations we construct an undirected graph $\mathbf{G}^w$ from the given set $\mathbf{W}$ of linear segments in a window centered at a candidate point p_0. The graph $\mathbf{G}^w$ has nodes $j = 1, .., m$ corresponding to the segments $\mathcal{S}_1, .., \mathcal{S}_m \in \mathbf{W}$. Each node j is attributed by a triple of parameters (θ_j, r_j, l_j), i.e. orientation, distance to the reference point p_0, and size of the linear segment. An

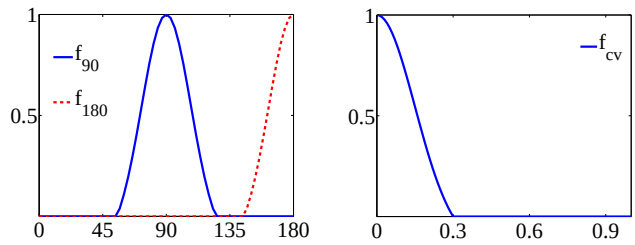

Fig. 5. Functions f_{90} (left figure, solid blue curve), f_{180} (left figure, dashed red curve), and f_{cv} (right figure) used in the rectangularity measure in Eq. (8).

edge $\{k,j\}$ is attributed with the angle $\beta_{k,j}$ and the pair-wise convexity $\tau_{k,j}$ of the corresponding pair of segments $\mathcal{S}_k, \mathcal{S}_j$. An edge $\{k,j\}$ is included in the graph $\mathbf{G}^w$ if $\beta_{k,j}$ and $\tau_{k,j}$ satisfy the constraints in Eqs. (6, 7). This attributed graph encodes properties of linear segments and their spatial relationships. Due to the graph construction and Definition 1, valid configurations $\mathbf{C}$ correspond to fully connected subgraphs $\mathbf{G}^c$, also called cliques, of the graph $\mathbf{G}^w$.

Below we introduce the new rectangularity measure $\rho(\mathbf{G}^c)$ that ranks a clique $\mathbf{G^c}$ corresponding to a valid configuration $\mathbf{C} \subseteq \mathbf{W}$. We define the measure with the following properties in mind. The rectangularity measure shall yield higher values for configurations with

1) higher degree of convexity given by lower values of the convexity measure τ
2) higher degree of angle alignments given by angles β
3) longer linear segments given by larger l.

In addition, the proposed rectangularity measure shall

4) have the increasing property $\rho(\mathbf{G}_1^c) \leq \rho(\mathbf{G}_2^c)$ for $\mathbf{G}_1^c \subseteq \mathbf{G}_2^c$. Thus, the rectangularity measure of a larger encompassing clique has a higher value
5) yield a zero value for configurations of linear segments with less than three sides of a rectangle.

We define the rectangularity measure of a graph clique $\mathbf{G}^c$ in terms of sums over its undirected edges $\{k,j\} \in E^c$

$$\rho(\mathbf{G}^c) = \left(\left(\sum_{\{k,j\} \in E^c} l_k l_j \mathrm{f}_{90}(\beta_{k,j})\mathrm{f}_{cv}(\tau_{k,j})\right) \times \left(\sum_{\{k,j\} \in E^c} l_k l_j \mathrm{f}_{180}(\beta_{k,j})\mathrm{f}_{cv}(\tau_{k,j})\right)\right)^{\frac{1}{4}}, \quad (8)$$

where $\mathrm{f}_{90}, \mathrm{f}_{180}$, and f_{cv} are mode functions depicted in Fig. 5. f_{90} and f_{180} equal zero for angles β that deviate from the mode center larger than the angle tolerance α. f_{cv} equals zero for the convexity measure τ larger than the convexity tolerance t. In our experiments we used $\alpha = 35°$ and $t = 0.3$. The exact definition of the mode function is not critical and is not given here due to space considerations.

The first factor of $\rho(\mathbf{G}^c)$ in Eq. (8) yields a non-zero value only if the valid configuration $\mathbf{C}$ contains at least one pair of approximately perpendicular linear segments that fulfill the convexity constraint in Eq. (7). The second factor is non-zero only if the valid configuration contains at least one pair of approximately parallel linear segments[3]. The product of these two factors is non-zero only if the valid configuration $\mathbf{C}$ contains at least one pair of parallel and one pair of perpendicular linear segments. The angles between linear segments of these parallel and perpendicular pairs are restricted to be approximately 0, 180, or 90 degrees since $\mathbf{C}$ is a valid configuration with linear segments constrained by Eq. (6). Thus, a non-zero rectangularity measure insures a valid configuration $\mathbf{C}$ containing at least one triple of segments arranged in a Π-like structure, as stated in property 5 above. This property allows suppression of a large number of configurations originating from clutter (e.g. lines, corners, junctions etc.). It is easy to verify that the other four properties above are also satisfied by the rectangularity measure in Eq. (8). Also note that the rectangularity measure scales linearly with the spatial size of rectangles.

D. Rectangularity feature

Given a set of linear segments $\mathbf{W}$ in an analysis window, we define below the rectangularity feature f_{R} of the corresponding graph $\mathbf{G}^w$. We denote the set of cliques of $\mathbf{G}^w$ as $\mathcal{K}(\mathbf{G}^w)$. The rectangularity feature of $\mathbf{G}^w$ is defined as

$$f_{\mathrm{R}}(\mathbf{G}^w) = \max_{\mathbf{G}^c \in \mathcal{K}(\mathbf{G}^w)} \rho(\mathbf{G}^c). \quad (9)$$

The corresponding optimal clique is

$$\mathbf{G}_{opt}^c = \operatorname*{argmax}_{\mathbf{G}^c \in \mathcal{K}(\mathbf{G}^w)} \rho(\mathbf{G}^c). \quad (10)$$

Due to the increasing property of ρ (the fourth property of the rectangularity measure stated in Sec. III-C), the maximum can be searched over the set of maximal cliques[4] only, denoted here by $\mathcal{M}(\mathbf{G}^w)$. That is

$$f_{\mathrm{R}}(\mathbf{G}^w) = \rho(\mathbf{G}_{opt}^c) = \max_{\mathbf{G}^c \in \mathcal{M}(\mathbf{G}^w)} \rho(\mathbf{G}^c). \quad (11)$$

Since the set of maximal cliques $\mathcal{M}(\mathbf{G}^w) \subseteq \mathcal{K}(\mathbf{G}^w)$ is much smaller than the set of graph cliques $\mathcal{K}(\mathbf{G}^w)$, the number of times the rectangularity measure ρ needs to be evaluated in Eq. (11) is considerably reduced in comparison to Eq. (9). Since, in addition, there are efficient algorithms for finding maximal cliques, e.g. [46], we compute the rectangularity feature by an exhaustive search for the maximum in Eq. (11).

Fig. 6 (left) shows an example of a given set $W = \{\mathcal{S}_1, \mathcal{S}_2, .., \mathcal{S}_6\}$ of linear segments and the optimal configuration $\mathbf{C}_{opt} = \{\mathcal{S}_1, \mathcal{S}_2, \mathcal{S}_3, \mathcal{S}_5\}$ in red, while Fig. 6 (right) shows the corresponding graph $\mathbf{G}^w$ and the optimal maximal clique $\mathbf{G}_{opt}^c$ in red. There are two additional maximal cliques $\mathbf{G}_1^c$ and $\mathbf{G}_2^c$ and corresponding valid configurations $\mathcal{C}_1 = \{\mathcal{S}_2, \mathcal{S}_3, \mathcal{S}_4, \mathcal{S}_6\}, \mathcal{C}_2 = \{\mathcal{S}_1, \mathcal{S}_2, \mathcal{S}_3, \mathcal{S}_4\}$. They, however, have lower rectangularity values $\rho(\mathbf{G}_1^c) < \rho(\mathbf{G}_{opt}^c), \rho(\mathbf{G}_2^c) < \rho(\mathbf{G}_{opt}^c)$.

Fig. 7 shows a couple of examples of the rectangularity feature computed for the real satellite and areal images. The

[3] f_{cv} in the second term has only a small impact on results. It reduces the rectangularity measure for configurations with badly aligned opposite sides with a non-zero convexity measure.

[4] Maximal cliques are cliques that are not contained in larger cliques.

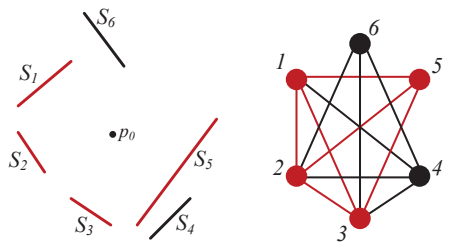

Fig. 6. Left: A set $W = \{\mathcal{S}_1, \mathcal{S}_2, .., \mathcal{S}_6\}$ of linear segments around a candidate point p_0. Right: A graph $\mathbf{G}^w$ for the set of linear segments. We assume an angle tolerance α such that all angle constraints are satisfied. Several node pairs of the graph are not connected by an edge due to the convexity constraint, which is not satisfied for an assumed convexity tolerance t. The red nodes of the graph are the nodes of the optimal maximal clique $\mathbf{G}^c_{opt}$. The corresponding valid configuration $\mathbf{C}_{opt}$ is marked in red on the left figure.

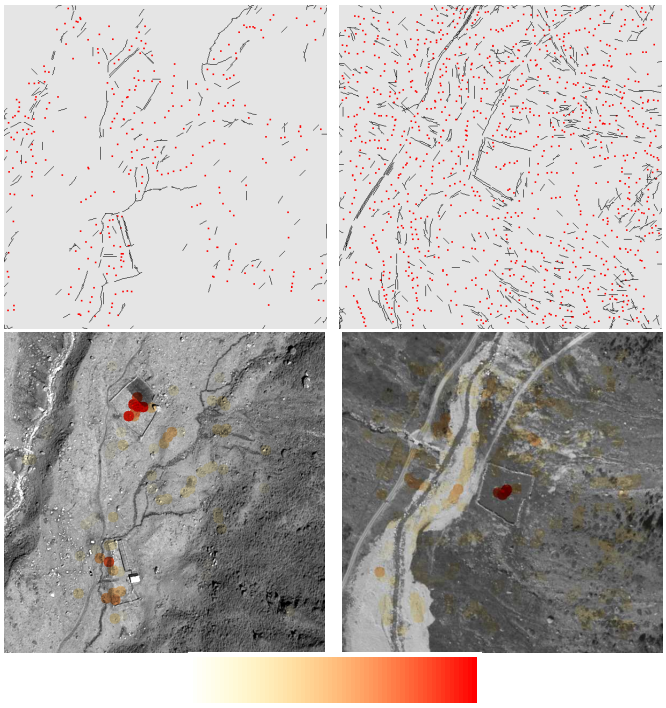

Fig. 7. First row: Bar edges (black) and candidate points (red) generated from the images in Fig. 1. Second row: The rectangularity feature computed at each candidate point and visualized by a colored disk. Color saturation increases and hue is changing from yellow to red for growing values of the features in accordance with the color bar in the bottom.

first row shows detected bar edges and candidate points[5] (Sec. II). The rectangularity feature f_{R} computed at the candidate points is visualized by colored disks in the second row. As expected, high values were obtained at positions of LE while zero or low values were obtained at most other candidate positions. One can see that the rectangularity feature map is quite sparse. This is partially because the rectangularity feature has zero value for spurious structures with less than three sides.

[5]Note that not all of the candidate points are the same as in Fig. 3. In contrast to Fig. 3, the map of candidate points in Fig. 7 resulted from the union of points coming from both valley and ridge edges. On the other hand, candidate points that are to too distant or too close to the edges were removed (see Sec. II) and do not appear in Fig. 7.

IV. Learning in the Rectangularity-Size Feature Space

The rectangularity feature scales with the structure size having lower values for small structures. A detector based on such a feature is prone to dismiss small rectangles. On the other hand, false structures of a small size are more frequent. We, therefore, introduce an additional feature f_{S} proportional to the structure size and learn a classifier from the available data in the two-dimensional rectangularity-size feature space. This may improve the trade-off between the sensitivity and the number of false detections in comparison to the one-dimensional case. We define the size of the structure, represented by the optimal clique $\mathbf{G}^c_{opt} \subseteq \mathbf{G}^w$, as

$$f_{\mathrm{S}}(\mathbf{G}^w) = \frac{\sum_j l_j r_j}{\sum_j l_j}, \tag{12}$$

where the sums are over all nodes of the optimal clique $\mathbf{G}^c_{opt}$. f_{S} is computed as the weighted distance of the linear segments of $\mathbf{C}_{opt}$ from the corresponding candidate point, where the weights are segment sizes.

Since only a few positive examples are available in our case, a classification approach should be carefully chosen. The linear classifiers are favorable when there is a danger of overfitting the data due to a limited number of available examples. They also are not computationally demanding. The normal w of the separating hyperplane of a linear classifier can be found by means of the Fisher Linear Discriminant analysis (FLD). In this approach, the optimal direction is determined such that the data from two classes projected on w is maximally separated. The separation is measured by the squared distance between class means normalized by the sum of their variances [45], [47]. This approach results in a simple solution represented in terms of class means and covariance matrices. In our case, however, the number of positive examples is very limited and the covariance matrix cannot reliably be estimated.

We optimize the normal direction w based on the large number of available samples from the dominant class of negatives and just a few examples from the class of positives. Let us define the expected signed distance between a deterministic point y (positive example) and the distribution X of negatives, both projected to the direction w and normalized by the standard deviation of the projected distribution

$$D_w(y, X) \equiv \frac{E_x\left[w^T y - w^T x\right]}{\sqrt{E_x\left[(w^T x - w^T \mu_x)^2\right]}} = \frac{w^T(y - \mu_x)}{\sqrt{w^T C_x w}}, \tag{13}$$

where μ_x and C_x are the mean and the covariance matrix of the distribution X, respectively. Next, we define the average signed distance between a set of deterministic points $\{y_i, i = 1, ..., n\}$ and the distribution X

$$\bar{D}_w(\{y_i\}, X) \equiv \frac{1}{n}\sum_{i=1}^{n} D_w(y_i, X) = \frac{w^T(\bar{y} - \mu_x)}{\sqrt{w^T C_x w}}, \tag{14}$$

where $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$. We now define the optimal direction w as the direction that maximizes the absolute value of the average signed distance between a set of points corresponding to positive examples and the distribution of the dominant class of negatives X, i.e.

$$w_{opt} \equiv \underset{w}{\operatorname{argmax}} |\bar{D}_w(\{y_i\}, X)|. \tag{15}$$

From Eqs. (14, 15) we obtain

$$w_{opt} = \underset{w}{\operatorname{argmax}} \frac{|w^T(\bar{y} - \mu_x)|}{\sqrt{w^T C_x w}}. \tag{16}$$

It can be shown that

$$w_{opt} = C_x^{-1}(\bar{y} - \mu_x) \tag{17}$$

is a solution of Eq. (16). The obtained direction w_{opt} is similar to the one in the FLD analysis [45]. In contrast to the FLD solution, Eq. (17) includes the covariance matrix of the class of negatives only, preferring the solution in the direction of the small variance of negatives. Negatives are well sampled in our problem and their covariance matrix can be robustly estimated. The positives are treated as deterministic points in the feature space and influence the solution only via their average. Literally, the average only weakly guides the solution pointing to the relevant location in the feature space. Note that the signed distance in Eq. (14) for w_{opt} given in Eq. (17) yields a positive value equal to the Mahalanobis distance $\bar{D}_{w_{opt}}(\{y_i\}, X) = \sqrt{(\bar{y} - \mu_x)^T C_x^{-1}(\bar{y} - \mu_x)}$ with the metric C_x.

The samples of X may include outliers. Therefore, in Eq. (17) we use the robust Multivariate Trimming (MVT) estimates of the mean and the covariance matrix [48]. The MVT is an iterative technique with mean and covariance matrices recomputed at each iteration. Given the current estimates of μ_x and C_x the Mahalanobis distance is computed for all the data points. A specified percentage of the observations with the largest Mahalanobis distance is discarded and the remaining data is used to recompute the estimates of μ_x and C_x. The technique is initialized with the sample mean and covariance matrix computed from the whole data. The samples with zero rectangularity f_{R}, which correspond to non-valid configurations, were excluded from such a training procedure. In our experiments we used three iterations and discarded 10% of observations in each iteration.

We will refer to

$$f_{\mathrm{RS}} = \begin{pmatrix} f_{\mathrm{S}} \\ f_{\mathrm{R}} \end{pmatrix}$$

as the rectangularity-size features. Given the optimal direction w_{opt}, the LE structures are detected by

$$f_{\mathrm{RS}}^T w_{opt} > b, \tag{18}$$

where b is a threshold to be set. It determines the trade-off between the sensitivity and the rate of false detections. The optimal linear feature combination $f_{RS}^T w_{opt}$, which is computed at candidate points, can be seen as a confidence measure of an enclosure structure being present in the area around the candidate point. Note that learning the optimal feature combination w_{opt} as shown above is not limited to two-dimensional feature spaces, but directly extends to higher dimensions. This will be used in our experiments in Sec. VI in order to compare the developed features with high-dimensional generic features.

Somewhat similar ideas of using Linear Discriminant Analysis (LDA) adapted to a small number of positives within the context of pedestrian detection appeared in [49]. Relying on the high-dimensional HOG features [30] and LDA, the authors modeled the background class with the mean and covariance matrix learned from unlabeled image patches. Their model trained on a few positives was highly competitive. In contrast to [49], our model was explicitly derived from optimization of Eq. (15) that was defined as a way to cope with the settings of the highly unbalanced problem at hand.

V. Detection of enclosures: the Silvretta Alps case study

We built a user interface that allows a user to explore large images, shows detections and their confidence, and allows to quickly examine and reject falsely detected sites. We determine a threshold parameter b for the LE detector in Eq. (18) based on the number of allowed false detections. The user interface allows choosing the number of false detections to be generated for an analyzed image.

We applied our detector to the region of the Silvretta Alps [50] of about 550 km^2 size. We used panchromatic images at 0.5m resolution captured by the GeoEye1 satellite. The data stems from a recent archaeological project in the Silvretta Alps [2]. We also repeated our experiments using the red channel of SWISSTOPO aerial images of 0.5m resolution that covered a slightly larger area of the same Silvretta region. However, the technical details below refer to the case of the satellite imagery.

The rectangularity feature may result in a large number of false detections in textured regions (e.g. forests). Since we are only interested in livestock enclosures that sparsely appear in grassland areas, high contrast texture regions were filtered out using the Morphological Texture Contrast (MTC) descriptor [31], [32], [51] thresholded with the Otsu method [52]. The size parameters r_1 and r_2 in the MTC were set to 30 and 60 pixels, respectively. This filters out urban areas, forests, rocky mountains, and other high contrast texture regions, while preserving individual structures.

For learning the optimal direction w_{opt} we used the 9 available examples of livestock enclosures and 49584 negative examples of structures around candidate points (see Sec. II) extracted from a 11000×17000 pixel satellite image. The detection threshold b in Eq. (18) was set such that the number of generated detections was 5000 in the analyzed 550km^2 area. After visual inspection of detected sites with our user interface, we found 13 structures resembling livestock enclosures. Some of these detections were found to be livestock enclosures that were hitherto unknown. An example of such an enclosure is shown in Fig. 8 on the left. Fig. 8 on the right shows a typical false detection. False detections are usually caused by streams and roads. The use of 3D data (e.g. LiDAR or based on stereo image pairs) would allow the discrimination of such false detections.

In our experiments we used a Matlab software. The satellite imagery that covers 550km^2 was divided into 17 partially overlapping images. Processing of all the images (including

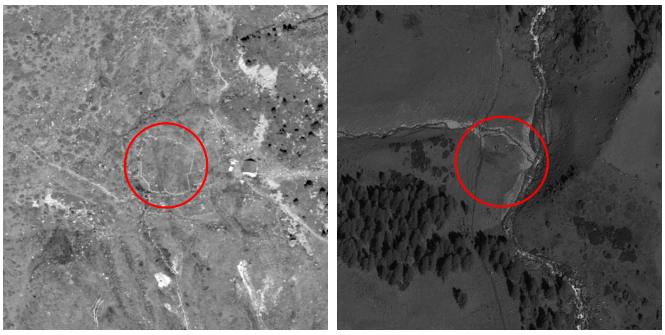

Fig. 8. Left: An example of a previously unknown enclosure that was detected using the proposed rectangularity-size feature. Right: A typical false detection.

all the stages), which is equivalent to processing of a single 53000×53000 pixel image, took three hours and forty minutes on a machine with an Intel Core i5 3.3 GHz Quad-Core processor and 32GB RAM.

VI. Comparative Experiments

A. Features for comparison

We evaluated the discrimination ability of the introduced rectangularity feature f_{R} and provided a comparison with the NMR measure f_{NMR} we developed earlier in [20] and the GODF-based feature f_{GODF} in [10] recently proposed for building detection. The GODF, denoted $\lambda(\theta)$, is a weighted gradient orientation histogram with gradient magnitudes as weights and discrete orientation $\theta \in [0, 180)$. The correlation of $\lambda(\theta)$ with a function having two modes separated by $90°$ served as a GODF-based feature f_{GODF} indicating the presence of rectilinear structures. The normalization constant was set such that $\lambda(\theta)$ is a unit vector, which gave us better results than for the normalization constant equal to the sum of the weights used in [10]. More implementation details can be found in [24]. Note that we did not compare the rectangularity feature with the whole approach developed in [10], because it is based on additional features not appropriate in the case of enclosures. We have also tested other methods for building detection (e.g. [8], [11]) applied to detection of livestock enclosures. Unfortunately, these methods completely failed to detect enclosures. Thus a corresponding quantitative comparison cannot be made.

We also used the learning framework described in Sec. IV in order to evaluate and compare the developed rectangularity-size features f_{RS}, high-dimensional HOG feature vectors f_{HOG} [30] and deep convolutional neural networks (CNN) [53], [54] generating so called deep features, denoted f_{CNN} with CNN substituted by the name of a particular CNN architecture. These deep features are neural activations generated by pre-trained CNNs at some intermediate layer of the deep network. Usually these features are extracted either from the last convolutional layer or from one of the following fully connected layers but before the final one. There is mounting evidence that such features generated by CNNs pre-trained on a very large dataset of labeled images have a sufficient representation power to perform recognition tasks on completely different types of target images. Several recent works successfully used deep features in conjunction with either a fully connected neural network [55] or even a simple linear classifier [56]–[58] trained on a relatively small target set of images. Moreover, deep features were also shown to be useful for classification of remotely sensed images [59]. Such an approach allows us using CNNs even though we have a very limited amount of positive examples to learn from.

We generated deep features using several CNN models pre-trained on the subsets of the ImageNet database [60]. Deep features $f_{\text{Vgg-f}}, f_{\text{Vgg-m}}, f_{\text{Vgg-m-2048}}, f_{\text{Vgg-s}}$[6] were extracted from the CNN architectures described in [27]. $f_{\text{Vgg-deep-16}}$, f_{OverFeat}, and $f_{\text{GoogLeNet}}$ were extracted from the networks described in [26], [28], and [29], respectively. f_{AlexNet} features were extracted from the network described in [25], while f_{CaffeNet} features were generated by an independently trained variation of that network as mentioned in [61], [62]. All deep features, except OverFeat, were computed using the Matlab toolbox MatConvNet [63] using the pre-trained models taken from the webpage [64] accompanying the toolbox. The OverFeat pre-trained model was taken from the webpage [65], which has the source code implementing the deep network presented in [28]. The "fast" network was used. Following recommendations of the authors of the CNN models, except for OverFeat, we subtracted the mean image of the training dataset from each image presented to the CNNs. Since we detect structures in gray-scale images, while the CNN models require RGB channels as an input, we set each channel equal to the given gray-scale image.

The HOG features were computed for 14×14 pixels cell size (with 7 pixels of overlap from one cell to the next) and 9 orientation bins, which gave us the best results among all configurations with which we experimented. We used the C source code of the HOG implementation available in the VLFeat package [66].

All the features were computed for image regions around the candidate points p. The size of these candidate regions was taken proportionally to the distance transform $D(p)$ (see Sec. II). To keep the size of HOG feature vectors constant we resized the candidate regions to 160×160 pixel patches. Deep features were computed for the candidate regions resized to the size required by a particular CNN architecture.

B. Measuring discrimination power of the features

Using a particular type of features f, livestock enclosures can be detected with $f > b$ for 1D features ($f_{\mathrm{R}}, f_{\mathrm{NMR}}, f_{\mathrm{GODF}}$) and with $f^T w_{opt} > b$ for multi-dimensional features ($f_{\mathrm{RS}}, f_{\mathrm{HOG}}, f_{\mathrm{CNN}}$), where b is an appropriate threshold to be set. Setting a particular threshold defines the true positive rate (TPR) and the false positive rate (FPR), or correspondingly the number of detected true and false positives (TP and FP). In our case, the effectiveness of the features is their ability to discriminate livestock enclosures from irrelevant structures and clutter. A possible measure of this ability is the

[6]We have also experimented with Vgg-m-1024 and Vgg-m-128 CNNs that have a smaller last hidden layer (1024 and 128 versus 4096 neurons in Vgg-m), but they gave worse results compared to Vgg-m. Therefore, we did not consider these features in our comparative experiments.

minimal number of FP detected with the threshold that insures $TPR \geq \xi$, where ξ is the predefined rate of true positives[7]. We computed FP for $\xi = 1$, denoted in the following by FP_{100}. This was done by setting the detection threshold b to the minimum value of f for 1D features and $f^T w_{opt}$ for multi-dimensional features computed for all positive examples. Obviously, the threshold used to obtain the detection rate $TPR = 1$ on a small number of available examples does not insure a detector with 100% detection rate. However, it allows us to measure and compare the discrimination ability of the features. FP_{100} is related to the spread of the class of positives toward the samples of the class of negatives, similarly to the Fisher criterion of discrimination ability [47]. However, FP_{100} also gives a rough estimate of the minimal number of false detections per area size that should be allowed in order to have a reasonable detection rate. Unfortunately, the actual detection rate cannot be reliably estimated due to a very small number of positive examples.

We also used an alternative measure of the discrimination ability that is the area under receiver operating characteristic (ROC) curve. It is especially useful in the presence of unbalanced classes [68], [69]. In contrast to FP_{100}, the area under receiver operating characteristic (AUC) does not rely on a particular threshold and a corresponding operating point on the ROC curve, but instead summarizes the detection performance for different values of the threshold. In fact, it is an average of true positive rates estimated for all false positive rates. The AUC has an important statistical property. It equals the probability that a randomly chosen sample y from the population of positives $\mathcal{P}$ has a higher score $f(y)$ (e.g. the rectangularity feature) than the score $f(x)$ for randomly chosen sample x from the population of negatives $\mathcal{N}$, i.e. $AUC(f) = P(f(y \in \mathcal{P}) > f(x \in \mathcal{N}))$. We estimated this probability of a correct ranking by means of the Wilcoxon-Mann-Whitney statistic [69], [70] as

$$\frac{1}{n_P n_N} \sum_{i=1}^{n_P} \sum_{j=1}^{n_N} I(f(y_i \in \mathcal{P}), f(x_j \in \mathcal{N})), \qquad (19)$$

where n_P and n_N denote the number positive and negative samples, respectively, and $I(u, v)$ is the indicator function defined as

$$I(u,v) = \begin{cases} 1, & u > v \\ 0, & u < v \\ 0.5, & u = v. \end{cases}$$

It should be noted that for the case of normally distributed features $f(x), f(y)$, the AUC has a simple relation to the Fisher criterion [43], [47], which is also frequently used as a separability measure between distributions. Namely, $AUC = \Phi\left(\frac{\mu_p - \mu_n}{\sqrt{\sigma_p^2 + \sigma_n^2}}\right)$, where Φ is the normal cumulative distribution function, evaluated for the Fischer criterion for positive and negative populations with distribution means μ_p, μ_n and standard deviations σ_p, σ_n, respectively.

[7]This corresponds to the so-called Neyman-Pearson task [67].

C. Evaluation procedure

In our experiments here we used panchromatic satellite images at 0.5m resolution that cover mountainous regions of the Silvretta Alps. Similarly to Sec. V, high contrast texture regions were filtered out using the Morphological Texture Contrast (MTC) descriptor [31], [32], [51]. We generated 49584 negative samples for training. The samples were taken around candidate points in a 11000×17000 pixel satellite image. For testing we used 57504 negative samples taken from a different satellite image of 10000×17000 pixel size that covers about 42.5 km^2. Overall only 9 examples of enclosures (positives) taken from aerial and satellite images were available to us. We augmented this data with additional 135 rotated versions of the same enclosures. 16 rotation angles were taken uniformly in the interval $[0, 360)$ degrees. This results in 144 positive examples, which is hardly enough for training and evaluation on separate train and test subsets as we have done with negative samples. In the case of high-dimensional feature vectors, the learned classifier parameters w_{opt} and the estimated performance may largely vary, depending on the selected subset of positives. In order to use most of the positives for training and also make reliable evaluation of the classier performance based on the data not used for training, we perform 9-fold cross validation. Note that we do not have hyper-parameters associated with the classifier that need to be set a priori or optimized on spare data. On each fold we use 16 examples of the same enclosure at different angles for testing and other 128 positives for training.

In the following we report the average value of the performance measures and the standard deviation over the nine folds of cross validation. In addition, we compare the sensitivity of FP_{100} to the reduction in number of positives used for training. To do so we compute the "inverted" 9-folds cross validation where on each fold we use only 16 examples of a single enclosure at different angles for training. For testing we use all 144 positives on each fold. Thus, the results may vary only due to the used training data, since the same data set is used for the performance evaluation. Note that for the case of FP_{100} measure, using all the positives for testing including a single training example can yield only higher (worse) FP_{100}, because the worst positive sample defines FP_{100}.

D. Results

The quantitative measures of the discrimination performance of the rectangularity f_{R} and the rectangularity-size features f_{RS} are summarized in Table I. The performance measures FP_{100} and AUC evaluate the discrimination ability of the features for our task. The FP_{100} (see Sec. VI-B) measures the number of false detections obtained in an area of approximately 42.5km^2, when all available positives are detected. This measure is particulary useful as it helps to decide how many false detections should be allowed in order to have a high detection rate, i.e. the rate that insures detection of at least all available positives examples. The AUC measure yields a performance ranking of different feature types similar to that of FP_{100}. On the other hand, unlike FP_{100}, the absolute values of AUC are quite close to each other giving

TABLE I
COMPARISON OF DISCRIMINATION MEASURES FOR THE FEATURES f_{NMR}, f_{GODF} AND THE PROPOSED FEATURES f_{R}, f_{RS}.

	FP_{100}	$AUC \times 10^2$	dim.
f_{GODF}	**6862**	99.262	1
f_{NMR}	**707**	99.925	1
f_{R}	**292**	99.967	1
f_{RS}	**201**	99.977	2

the wrong impression of similar performance. The high and close values of AUC are due to the ability of the detectors to reject most negatives while detecting a modest number of all available positives. The corresponding ROCs saturate at the maximum detection rate already for small values of false positive rate and differ only for lower false positive rates. Nevertheless, along with the FP_{100}, which is more intuitive and useful for our application, we also provide AUC because it is commonly used for evaluation of detector performance. The last column in Table I indicates the dimensionality of the features.

Table I shows that the discrimination ability of the rectangularity-size features f_{RS} is superior to the others. It allowed reduction of false positives by 31% relative to f_{R}, which is in turn considerably better than the NMR measure. Though effective for building detection, the GODF-based feature turned out to be far worse for detecting faint enclosures in cluttered background. This feature is not useful when computed over large windows, where the relative number of points belonging to an enclosure is small.

To compute the rectangularity-size features f_{RS} we learned w_{opt} from the separate training dataset of 49584 negative examples (see Sec. VI-C). The set of 144 augmented positives used for testing of all the feature types was also used for training the linear classifier. Learning the two-dimensional w_{opt} involves positives only via their average $\bar{y}$ (see Eq. (17)) and uses separate large datasets of negatives, therefore it is unlikely that the data is overfitted. Nevertheless, below we carried out another set of experiments, where we avoid the use of the same positives for training and testing by means of cross validation procedure (see Sec. VI-C for details). Using cross validation also allowed us comparison with high-dimensional HOG and deep CNN based features (deep features), which are much harder to keep from overfitting.

Table II shows the discrimination performance of the rectangularity-size features f_{RS}, high-dimensional histogram of oriented gradients f_{HOG}, and deep features f_{CNN} generated by several pre-trained CNNs. For all CNN architectures, the table gives the layer used to extract features that produced the best result. Given a particular set of features, we use the methodology described in Sec. IV based on training the linear classifier (learning the hyperplane w_{opt}). The table shows mean values, standard deviation, and worse values (max or min) for the discrimination measures FP_{100} and AUC over nine folds of cross validation. On each fold of cross validation 128 positives (augmented from 8 real examples) were used for training and the remaining 16 positives (augmented from the 9th remaining example) for testing (see Sec. VI-C for details). Mean values of FP_{100} show that the rectangularity-size features f_{RS} outperform all the other features by a large margin. Surprisingly however, two architectures of CNN, AlexNet [25] and Vgg-f [27], provided us with deep features that showed relatively high performance. This is a remarkable result, because the CNNs were trained on a completely different image dataset, while the linear classifier w_{opt} was trained on 128 examples augmented from just 8 real enclosures. Note that no fine-tuning of the pre-trained CNN was performed. These results also indicate that the simple methodology we developed in Sec. IV for learning from a few positive and a large number of negative examples is useful even for the case of high-dimensional features. Though, as we show below, performance of such features is much more sensitive to the number of positives used for training the linear classifier.

We experimented with CNNs with one or two final fully connected layers or the softmax function of the last layer removed. The table shows the layer that yields the best performing features. For AlexNet the best result was obtained when two final fully connected layers were kept. For all the other CNNs the best results were obtained when features were taken from the first layer that generates data reduced to 1×1 spatial dimension, which is an average pooling for the GoogLeNet and a fully connected layer for all the other CNNs. Our results support the hypothesis that convolutional layers of pre-trained CNNs generate generic intermediate features that might be useful for various tasks. In contrast, the final fully connected layers generate task specific features.

Table II also shows how the performance for all features dropped when only 16 augmented samples were used for training (see Sec. VI-C for details) within each fold of cross validation. However, in contrast to the other features, the rectangularity-size f_{RS} still yielded relatively high performance, while all the other high-dimensional features became not useful. This experiment showed high sensitivity of the performance to the number of training examples for the case of high-dimensional features. This also suggests that collecting additional examples might substantially improve their performance.

We also notice that the deeper architectures (Vgg-deep-16, Vgg-deep-19, and GoogLeNet) did not have superior performance for our task. The architecture of CNNs was more important than just their depth, which is in line with a recent observation in [71]. Note that the best performing CNNs AlexNet and Vgg-f have similar architecture [27]. The importance of the particular architecture is also evident from the large variability of the performance of the different CNNs in Table II. Moreover, though CaffeNet and AlexNet are supposed to perform similarly (the first network is a minor variation of the second [62]) they produce substantially different results. The differences in particular training procedures may be responsible for such a discrepancy. The choice of the CNN architecture and training procedure was crucial for our task and seems likely to be critical for other applications. However, it seems that currently there is no established alternative to the trial-and-error based choice of the most suitable architecture for the task at hand.

TABLE II
COMPARISON OF DISCRIMINATION MEASURES FOR MULTI-DIMENSIONAL CNN, HOG, AND THE PROPOSED RECTANGULARITY-SIZE FEATURES f_{RS}. 9-FOLD CROSS VALIDATION WAS PERFORMED ON 144 (AUGMENTED) POSITIVES WITH EITHER 128 OR 16 SAMPLES (N_{pos}) USED FOR TRAINING. LARGE SEPARATE DATASETS OF NEGATIVES WERE USED FOR TRAINING AND TESTING. "fc" ADJACENT TO THE LAYER NUMBER IN THE TABLE STANDS FOR "FULLY CONNECTED".

feature	$N_{pos} = 128$						$N_{pos} = 16$	dimensionality	output layer
	FP_{100}			$AUC \times 10^2$			FP_{100}		
	mean	std	max	mean	std	min	mean		
f_{RS}	**49.3**	64.3	203	99.976	0.036	99.884	206.3	2	
f_{AlexNet}	**178.6**	373.7	1145	99.943	0.124	99.615	35834.1	4096	fc_7
$f_{\mathrm{Vgg\text{-}f}}$	**195.4**	269.5	847	99.849	0.213	99.358	22699.8	4096	fc_6
$f_{\mathrm{Vgg\text{-}m\text{-}2048}}$	**365.8**	520.6	1688	99.814	0.287	99.091	25895.0	4096	fc_6
$f_{\mathrm{Vgg\text{-}deep\text{-}19}}$	**578.1**	1694.3	5096	99.718	0.826	97.515	32573.9	4096	fc_{17}
f_{CaffeNet}	**609.7**	1153.8	3498	99.771	0.384	99.000	46673.3	4096	fc_6
$f_{\mathrm{Vgg\text{-}m}}$	**631.3**	1039.9	2600	99.752	0.484	98.527	31120.2	4096	fc_6
$f_{\mathrm{GoogLeNet}}$	**1002.4**	2176.5	6549	99.727	0.608	98.161	18603.1	1024	avg pool_{21}
$f_{\mathrm{Vgg\text{-}s}}$	**2911.2**	7097.7	21720	99.353	1.348	95.835	38231.3	4096	fc_6
f_{OverFeat}	**2967.3**	8164.2	24715	99.416	1.554	95.281	37536.7	3072	fc_6
$f_{\mathrm{Vgg\text{-}deep\text{-}16}}$	**3473.0**	10294.5	30925	99.447	1.576	95.244	39626.8	4096	fc_{14}
f_{HOG}	**7472.8**	8352.8	27155	97.926	2.404	92.045	54154.9	4356	
$f_{\mathrm{Rand\text{-}4096}}$	**54039.1**	3124.5	57019.9	49.871	6.694	39.629	57054.1	4096	
$f_{\mathrm{Rand\text{-}2}}$	**54227.0**	2819.5	57087.5	50.736	6.164	41.098	57034.4	2	

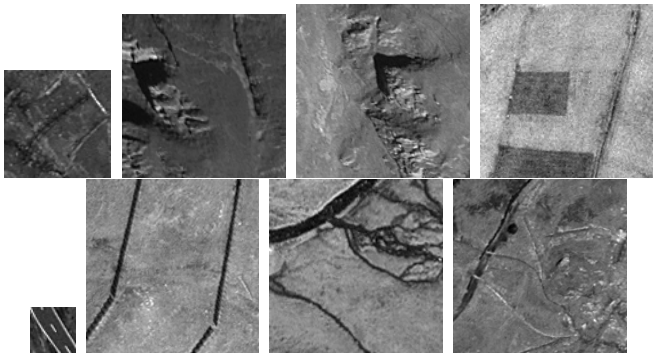

Fig. 9. Candidate patches generating highest responses by AlexNet (top row) and Vgg-f (bottom row) architectures of pre-trained CNNs followed by a linear classifier trained on 8 real examples (128 augmented examples) of rectangular enclosures and large number of negatives. Note that these patches contain structures that are conceptually close to rectangles.

In Fig. 9 we show candidate patches seen by AlexNet and Vgg-f CNNs that generate top responses of the linear classifier $f_{\mathrm{CNN}}^T w_{opt}$. The patches were taken out of 57504 negative samples used for testing. These top response patches resemble the structures of our interest, indicating that corresponding deep features might be powerful enough to capture the concept of the rectangular enclosures.

For reference purposes, in Table II we also evaluate performance of random feature vectors f_{Rand} using the same evaluation strategy. The random feature vectors with 2 and 4096 entries of independently identically distributed variables were drawn from the standard normal distribution. As expected, such features give average AUC values close to 0.5. The resulting mean values for false positives FP_{100} are not far from the overall number of samples used for testing (57504).

VII. APPLICATION TO DETECTION OF BUILDINGS

Though the rectangularity feature was developed for a particular task of detecting ruined livestock enclosures, it can also be used for other tasks. Here we illustrate its application to building detection. Since the MFC based line detector extracts bar edges only, we replaced it with the line segment detector of [36] that also extracts step edges that are more appropriate for detection of buildings. In the case of strong object contrasts (unlike the case of livestock enclosures) it reliably detects object borders and yields relatively small number of edges caused by clutter.

Once the edges were extracted we used the same algorithms as described in Sec. II and Sec. III in order to generate the rectangularity feature f_{R}. Since we do not have a labeled training dataset for buildings we did not experiment here with the rectangularity-size features f_{RS}. We also do not expect that it can essentially be better than f_{R}, because the variability of buildings sizes is much smaller than for the case of the enclosures. Nevertheless, we still took into account the dependence of the f_{R} on the size of the structure by normalizing the rectangularity feature as $f_{\mathrm{R}}/f_{\mathrm{S}}$. Fig. 10 illustrates $\ln(f_{\mathrm{R}}/f_{\mathrm{S}})$ computed for a SWISSTOPO 4000×4000 aerial image of 0.25m resolution taken over the Bernese Alps. The logarithm was taken in order to make weak detections better visible. We used the same parameters as before for 0.5m resolution images, except for the maximal size of building structures. The maximal size was reduced to 65 pixels, while the minimal size was kept to 15 pixels. One can see that most of the buildings were detected. On the other hand, there are false detections mostly caused by occasional configurations of forest and field edges or roads. In urban areas our detector may produce many false detections in between adjacent buildings or other manmade structures. Therefore, in the original form, the detector might be more appropriate for rural or mountainous areas when high sensitivity is needed for detection of possibly occluded individual rare structures. For better performance it can be adapted to detect buildings (instead of enclosures) by, for example, incorporating region and/or corner cues. Such adaptations, however, are out of the scope of this paper, as is the quantitative evaluation of performance for building

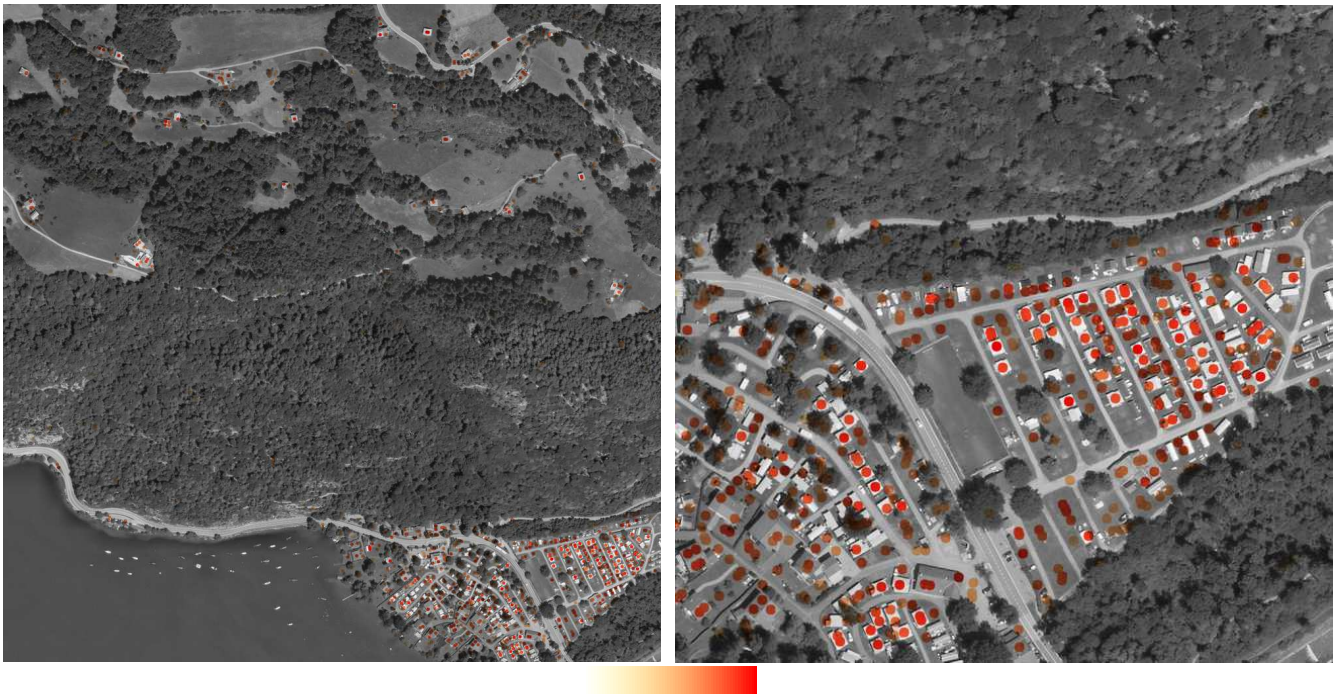

Fig. 10. Left: Building detections in 4000×4000 aerial (SWISSTOPO) image of 0.25m resolution visualized by colored disks. Right: Enlarged bottom right part of the image in the left. Color saturation increases and hue is changing from yellow to red for growing values of $\ln(f_{\mathrm{R}}/f_{\mathrm{S}})$ in accordance with the color bar in the bottom.

detection tasks.

VIII. Conclusion

We introduced the scalar rectangularity feature f_{R} for detection of approximately rectangular livestock enclosure structures in remotely sensed imagery. It has shown high performance in discriminating ruined enclosures from irrelevant structures and clutter. Due to the inherent difficulties of our problem, such a performance is hardly achievable with other approaches for detection of rectangular contours, nor with related approaches, e.g. for detection of buildings. Note that, while the building detection problem can be addressed using the enclosure detector (see Sec. VII for example), detection of enclosures cannot be approached using building detectors. In general, methods for building detection are not suitable for our case because of considerably lower heights (resulting in low feature contrasts) and feature sizes (ruined walls versus building rooftops), and due to the absence of various cues (roof colors, roof homogeneity, shadows, 3D cues, etc.). Some walls or parts of them may be missing or may also be missed in the edge extraction (the width of linear features does not exceed two pixels in images of 0.5m resolution). Various irrelevant structures (trails, streams, rocks etc.) with sizes or/and reflectance properties similar to those of enclosure walls may occasionally form rectilinear configurations. In contrast to enclosures, building rooftops are much more distinctive structures. As an example we have shown that the GODF-based feature used for detection of buildings reveals a poor discrimination ability for our task.

We also designed a size feature and introduced a methodology used for learning a linear classifier in the two-dimensional rectangularity-size feature space that improves over the detector based solely on the rectangularity feature. The same methodology in high-dimensional feature space was used to learn the classifier based on HOG feature vectors and feature vectors generated by pre-trained deep convolutional neural networks (deep features). Quantitative comparison has shown that the rectangularity-size features f_{RS} clearly outperform these state-of-the-art features for our task.

However, we have found that several pre-trained deep CNN architectures (that yield generic deep features) along with the linear classifier trained using our methodology from just a few positive samples (augmented using rotations) and a large number of negatives may still result in a well performing detector. Although these CNN-based features did not perform as well as the rectangularity-size features f_{RS}, they may still be useful for detection of the livestock enclosures of very low contrasts. Contrary to f_{RS}, they do not require a separate stage of extracting bar edges, which may fail in the cases of very low contrasts (e.g. due to low heights of ruined walls). Moreover, given examples of enclosures of non-rectangular shape, we could easily retrain our linear classifier using the same generic deep features. The resulting performance is likely to be improved by learning from more augmented examples using additional transformations, e.g. scaling, flipping, brightness transformations etc. Availability of additional (real) positive examples is certainly critical for improving performance of the deep CNN based detector and may also enable performing fine-tuning of the CNN itself for further gain in performance. The above issues are interesting for future research.

We reported that by using our algorithms (detection of bar edges, candidate generation, computation of the rectangularity-

size features, linear classification) we detected livestock enclosures in the Silvretta Alps that were hitherto unknown. We also discussed a different application of the rectangularity feature for detection of buildings in rural or mountainous areas. For better performance in such an application, the rectangularity feature should be accompanied by other features capturing additional properties of buildings.

ACKNOWLEDGMENT

Igor Zingman thanks the German Research Foundation (DFG) for financial support within project A05 of SFB/Transregio 161. Otávio A. B. Penatti thanks the RECOD lab at University of Campinas for the infrastructure used for computation of HOG and OverFeat features.

REFERENCES

[1] K. Kothieringer, C. Walser, B. Dietre, T. Reitmaier, J. N. Haas, and K. Lambers, "High impact: early pastoralism and environmental change during the Neolithic and Bronze Age in the Silvretta Alps (switzerland/austria) as evidenced by archaeological, palaeoecological and pedological proxies," *Zeitschrift für Geomorphologie, Supplementary Issues*, vol. 59, no. 2, pp. 177–198, 2015.

[2] K. Lambers and I. Zingman, "Towards detection of archaeological objects in high-resolution remotely sensed images: the Silvretta case study," in *Archaeology in the Digital Era, vol. II (e-papers) from the 40th Conf. on Computer Applications and Quantitative Methods in Archaeology, Southampton, March 2012*, G. Earl *et al.*, Eds. Amsterdam University Press, 2013, pp. 781–791.

[3] Ø. D. Trier, S. Ø. Larsen, and R. Solberg, "Automatic detection of circular structures in high-resolution satellite images of agricultural land," *Archaeological Prospection*, vol. 16, pp. 1–15, 2009.

[4] H. Mayer, "Automatic object extraction from aerial imagery - a survey focusing on buildings." *Computer Vision and Image Understanding*, vol. 74, no. 2, pp. 138–149, 1999.

[5] C. Lin and R. Nevatia, "Building detection and description from a single intensity image," *Computer Vision and Image Understanding*, vol. 72, no. 2, pp. 101–121, 1998.

[6] T. Kim and J.-P. Muller, "Development of a graph-based approach for building detection," *Image Vision Comput.*, vol. 17, no. 1, pp. 3–14, 1999.

[7] A. Croitoru and Y. Doytsher, "Right-angle rooftop polygon extraction in regularised urban areas: Cutting the corners," *The Photogrammetric Record*, vol. 19, no. 108, pp. 311–341, 2004.

[8] C. R. Jung and R. Schramm, "Rectangle detection based on a windowed Hough transform," in *Proceedings of the Computer Graphics and Image Processing (SIBGRAPI), XVII Brazilian Symposium*, 2004, pp. 113–120.

[9] S. Krishnamachari and R. Chellappa, "Delineating buildings by grouping lines with MRFs," *IEEE Transactions on Image Processing*, vol. 5, no. 1, pp. 164–168, 1996.

[10] 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.

[11] B. Sirmacek and C. Unsalan, "A probabilistic framework to detect buildings in aerial and satellite images." *IEEE Tran. Geoscience and Remote Sensing*, vol. 49, no. 1-1, pp. 211–221, 2011.

[12] ——, "Urban-area and building detection using SIFT keypoints and graph theory." *IEEE T. Geoscience and Remote Sensing*, vol. 47, no. 4, pp. 1156–1167, 2009.

[13] A. Manno-Kovacs and T. Sziranyi, "Multidirectional Building Detection in Aerial Images Without Shape Templates," *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, no. 1, pp. 010 000–232, May 2013.

[14] M. Ortner, X. Descombes, and J. Zerubia, "A marked point process of rectangles and segments for automatic analysis of digital elevation models," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 30, no. 1, pp. 105–119, 2008.

[15] Y. Liu, T. Ikenaga, and S. Goto, "An MRF model-based approach to the detection of rectangular shape objects in color images." *Signal Processing*, vol. 87, no. 11, pp. 2649–2658, 2007.

[16] C. G. Keller, C. Sprunk, C. Bahlmann, J. Giebel, and G. Baratoff, "Real-time recognition of US speed signs," in *Intelligent Vehicles Symposium*. IEEE, 2008, pp. 518–523.

[17] G. B. Loy and N. M. Barnes, "Fast shape-based road sign detection for a driver assistance system," in *2004 IEEE/RSJ International Conference on Intelligent Robots and Systems, Sendai, Japan*, 2004, pp. 70–75.

[18] Y. Zhu, B. Carragher, F. Mouche, and C. S. Potter, "Automatic particle detection through efficient hough transforms." *IEEE Trans. Med. Imaging*, vol. 22, no. 9, pp. 1053–1062, 2003.

[19] Z. Yu and C. Bajaj, "Detecting circular and rectangular particles based on geometric feature detection in electron micrographs," *Journal of Structural Biology*, vol. 145, no. 12, pp. 168 – 180, 2004.

[20] I. Zingman, D. Saupe, and K. Lambers, "Automated search for livestock enclosures of rectangular shape in remotely sensed imagery," in *Proc. SPIE, Image and Signal Processing for Remote Sensing XIX*, L. Bruzzone, Ed., vol. 8892, Dresden, Germany, 2013, pp. 88 920F–1 – 88 920F–11.

[21] H. Moon, R. Chellappa, and A. Rosenfeld, "Optimal edge-based shape detection." *IEEE Transactions on Image Processing*, vol. 11, no. 11, pp. 1209–1227, 2002.

[22] X. Descombes and J. Zerubia, "Marked point process in image analysis," *Signal Processing Magazine, IEEE*, vol. 19, no. 5, pp. 77–84, 2002.

[23] Y. Verdie and F. Lafarge, "Detecting parametric objects in large scenes by Monte Carlo sampling," *International Journal of Computer Vision*, vol. 106, no. 1, pp. 57–75, 2014.

[24] I. Zingman, D. Saupe, and K. Lambers, "Detection of incomplete enclosures of rectangular shape in remotely sensed images," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2015.

[25] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[26] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *CoRR*, vol. abs/1409.1556, 2014.

[27] K. Chatfield, K. Simonyan, A. Vedaldi, and A. Zisserman, "Return of the devil in the details: Delving deep into convolutional nets," in *British Machine Vision Conference*, 2014.

[28] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. LeCun, "OverFeat: Integrated recognition, localization and detection using convolutional networks," in *International Conference on Learning Representations (ICLR)*, April 2014.

[29] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015.

[30] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2005, pp. 886–893.

[31] I. Zingman, D. Saupe, and K. Lambers, "A morphological approach for distinguishing texture and individual features in images," *Pattern Recognition Letters*, vol. 47, pp. 129 – 138, 2014, advances in Mathematical Morphology.

[32] ——, "Detection of texture and isolated features using alternating morphological filters," in *Mathematical Morphology and Its Applications to Signal and Image Processing*, ser. Lecture Notes in Computer Science, C. Hendriks, G. Borgefors, and R. Strand, Eds. Springer, 2013, vol. 7883, pp. 440–451.

[33] T. Lindeberg, "Edge detection and ridge detection with automatic scale selection," *International Journal of Computer Vision*, vol. 30, no. 2, pp. 117–156, 1998.

[34] C. Grigorescu, N. Petkov, and M. Westenberg, "Contour and boundary detection improved by surround suppression of texture edges," *Image and Vision Computing*, vol. 22, pp. 609 – 622, 2004.

[35] G. Papari and N. Petkov, "An improved model for surround suppression by steerable filters and multilevel inhibition with application to contour detection," *Pattern Recognition*, vol. 44, pp. 1999 – 2007, 2011.

[36] R. Grompone von Gioi, J. Jakubowicz, J.-M. Morel, and G. Randall, "LSD: A fast line segment detector with a false detection control," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 4, pp. 722–732, Apr. 2010.

[37] K. Siddiqi, S. Bouix, A. Tannenbaum, and S. W. Zucker, "Hamilton-jacobi skeletons," *International Journal of Computer Vision*, vol. 48, no. 3, pp. 215–231, 2002.

[38] S. M. Pizer, K. Siddiqi, G. Székely, J. N. Damon, and S. W. Zucker, "Multiscale medial loci and their properties," *International Journal of Computer Vision*, vol. 55, no. 2-3, pp. 155–179, 2003.

[39] P. Dimitrov, J. N. Damon, and K. Siddiqi, "Flux invariants for shape," in *Proceedings of the the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 1. IEEE, 2003, pp. I–835.

[40] D. Engel and C. Curio, "Scale-invariant medial features based on gradient vector flow fields," in *International Conference on Pattern Recognition (ICPR)*, 2008, pp. 1–4.

[41] C. Xu and J. L. Prince, "Snakes, shapes, and gradient vector flow," *Image Processing, IEEE Transactions on*, vol. 7, no. 3, pp. 359–369, 1998.

[42] ——, "Generalized gradient vector flow external forces for active contours," *Signal processing*, vol. 71, no. 2, pp. 131–139, 1998.

[43] R. O. Duda and P. E. Hart, "Use of the Hough transformation to detect lines and curves in pictures," *Commun. ACM*, vol. 15, no. 1, pp. 11–15, Jan. 1972.

[44] L. Lam, S.-W. Lee, and C. Y. Suen, "Thinning methodologies-a comprehensive survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 9, pp. 869 – 885, 1992, p.879.

[45] R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis*. John Willey & Sons, 1973.

[46] C. Bron and J. Kerbosch, "Algorithm 457: finding all cliques of an undirected graph," *Commun. ACM*, vol. 16, no. 9, pp. 575–577, Sep. 1973.

[47] K. Fukunaga, *Introduction to Statistical Pattern Recognition*. Academic Press., 1990.

[48] S. J. Devlin, R. Gnanadesikan, and J. R. Kettenring, "Robust estimation of dispersion matrices and principal components," *Journal of the American Statistical Association*, vol. 76, no. 374, pp. 354–362, 1981.

[49] B. Hariharan, J. Malik, and D. Ramanan, "Discriminative decorrelation for clustering and classification," in *Proc. of the European Conference on Computer Vision*. Springer, 2012, pp. 459–472.

[50] K. Lambers and T. Reitmaier, "Silvretta Historica: satellite-assisted archaeological survey in an Alpine environment," in *CAA 2010 Fusion of cultures: Proceedings of the 38th annual conference on Computer Applications and Quantitative Methods in Archaeology, Granada, Spain, April 2010*, F. Contreras, M. Farjas, and F. J. Melero, Eds. Oxford: Archaeopress, 2013.

[51] I. Zingman, D. Saupe, and K. Lambers, "Morphological operators for segmentation of high contrast textured regions in remotely sensed imagery," in *Proc. of the IEEE Int. Geoscience and Remote Sensing Symposium*, Munich, Germany, July 2012, pp. 3451–3454.

[52] N. Otsu, "A Threshold Selection Method from Gray-level Histograms," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 9, pp. 62–66, Jan. 1979.

[53] S. Haykin, *Neural networks and learning machines*, 3rd ed., 2009, ch. 4.17.

[54] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[55] M. Oquab, L. Bottou, I. Laptev, and J. Sivic, "Learning and transferring mid-level image representations using convolutional neural networks," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2014.

[56] J. Donahue, Y. Jia, O. Vinyals, J. Hoffman, N. Zhang, E. Tzeng, and T. Darrell, "Decaf: A deep convolutional activation feature for generic visual recognition," *arXiv preprint arXiv:1310.1531*, 2013.

[57] A. S. Razavian, H. Azizpour, J. Sullivan, and S. Carlsson, "CNN features off-the-shelf: an astounding baseline for recognition," in *Computer Vision and Pattern Recognition Workshops (CVPRW)*. IEEE, 2014, pp. 512–519.

[58] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Region-based convolutional networks for accurate object detection and segmentation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 2015.

[59] O. A. B. Penatti, K. Nogueira, and J. A. dos Santos, "Do deep features generalize from everyday objects to remote sensing and aerial scenes domains?" in *Computer Vision and Pattern Recognition Workshops (CVPRW)*. IEEE, 2015.

[60] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet Large Scale Visual Recognition Challenge," *International Journal of Computer Vision (IJCV)*, pp. 1–42, April 2015.

[61] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. B. Girshick, S. Guadarrama, and T. Darrell, "Caffe: Convolutional architecture for fast feature embedding," *CoRR*, vol. abs/1408.5093, 2014.

[62] Caffe Model Zoo. [Online]. Available: http://caffe.berkeleyvision.org/model_zoo

[63] A. Vedaldi and K. Lenc, "Matconvnet – convolutional neural networks for matlab," in *Proceeding of the ACM Int. Conf. on Multimedia*, 2015.

[64] Pre-trained CNN models. Accessed: 18.11.2015. [Online]. Available: http://www.vlfeat.org/matconvnet/pretrained

[65] Pre-trained OverFeat. Version: v04-2. [Online]. Available: https://github.com/sermanet/OverFeat

[66] A. Vedaldi and B. Fulkerson. VLFeat: An open and portable library of computer vision algorithms. Version: 0.9.16. [Online]. Available: http://www.vlfeat.org/

[67] M. I. Schlesinger and V. Hlavac, *Ten lectures on statistical and structural pattern recognition*. Springer, 2002.

[68] T. Fawcett, "An introduction to ROC analysis," *Pattern Recogn. Lett.*, vol. 27, no. 8, pp. 861–874, Jun. 2006.

[69] W. J. Krzanowski and D. J. Hand, *ROC Curves for Continuous Data*. Chapman & Hall/CRC, 2009.

[70] J. A. Hanley and B. J. McNeil, "The meaning and use of the area under a receiver operating characteristic (ROC) curve," *Radiology*, vol. 143, no. 1, pp. 29–36, 1982.

[71] B. Pepik, R. Benenson, T. Ritschel, and B. Schiele, "What is holding back convnets for detection?" in *Pattern Recognition*. Springer, 2015, pp. 517–528.