

# USING REDUNDANT INFORMATION FROM MULTIPLE AERIAL IMAGES FOR THE DETECTION OF BOMB CRATERS BASED ON MARKED POINT PROCESSES

C. Kruse\*, F. Rottensteiner, C. Heipke

Institute of Photogrammetry and GeoInformation, Leibniz Universität Hannover, Germany  
(kruse, rottensteiner, heipke)@ipi.uni-hannover.de

Commission II, ICWG II/III

**KEY WORDS:** Marked Point Processes, RJMCMC, Multiple Aerial Wartime Images, Bomb Craters, Duds, Impact Maps

## ABSTRACT:

Many countries were the target of air strikes during World War II. Numerous unexploded bombs still exist in the ground. These duds can be tracked down with the help of bomb craters, indicating areas where unexploded bombs may be located. Such areas are documented in so-called impact maps based on detected bomb craters. In this paper, a stochastic approach based on marked point processes (MPPs) for the automatic detection of bomb craters in aerial images taken during World War II is presented. As most areas are covered by multiple images, the influence of redundant image information on the object detection result is investigated: we compare the results generated based on single images with those obtained by our new approach that combines the individual detection results of multiple images covering the same location. The object model for the bomb craters is represented by circles. Our MPP approach determines the most likely configuration of objects within the scene. The goal is reached by minimizing an energy function that describes the conformity with a predefined model by Reversible Jump Markov Chain Monte Carlo sampling in combination with simulated annealing. Afterwards, a probability map is generated from the automatic detections via kernel density estimation. By setting a threshold, areas around the detections are classified as contaminated or uncontaminated sites, respectively, which results in an impact map. Our results show a significant improvement with respect to its quality when redundant image information is used.

## 1. INTRODUCTION

### 1.1 Motivation

Many countries were bombed during World War II. Even though the last combat operations date back more than three quarters of a century, numerous unexploded bombs still remain in the ground. Experts of Lower Saxony's explosive ordnance disposal service estimate that 10 % - 15 % of all dropped bombs did not detonate. Surveillance flights were carried out before and after an air strike, but also on other occasions. Hence, there are usually numerous images covering the same area and the number of bomb craters for a given area may vary between images taken at different times. These images are used today by the explosive ordnance disposal service of Lower Saxony to find potentially dangerous sites. In this context, a central task is the identification of duds. Typically, such investigations are restricted to particularly endangered or otherwise relevant areas. However, the processing effort is still immense as the analysis is carried out manually. For many applications, it is sufficient to have comprehensive information on the basic occurrence of war impacts. Such knowledge can be represented in an "impact map" which indicates whether areas are likely to be contaminated or not. In this context, contaminated areas should be very likely to contain one or more duds, whereas uncontaminated areas should not contain any dud. The automatic creation of such an impact map could accelerate the manual evaluation process. To achieve this goal, an automatic detection of bomb craters in aerial wartime images is indispensable.

As duds are difficult to detect, we focus on estimating the probability for their occurrence using bomb craters in aerial wartime images as indicators for the areas where duds may be located. This probability can be used to identify contaminated

areas to be represented in an impact map. We are interested in detecting areas that have a very high likelihood of containing a dud so that it makes sense to send a team of experts to that area to probe it, i.e. to visit the site and take measurements using geophysical detectors. As this is expensive, false detections should be avoided, i.e. the correctness of the results is most important in the proposed scenario. The main benefit for the explosive ordnance disposal service is that the particular images in these areas would then no longer have to be checked manually.

Probabilistic models have been successfully developed to detect a configuration of predefined objects. Approaches integrating probabilistic models of context such as Conditional Random Fields (Kumar and Hebert, 2006) are restricted to local interactions, e.g. between neighbouring pixels; more global constraints about the objects, e.g. their shape, are difficult to integrate. This can be handled by marked point processes (MPPs; Descombes and Zerubia, 2002; Daley and Vere-Jones, 2003). A MPP is a random variable whose realizations are configurations of geometrical objects of a certain type. This approach uses a strong object model that offers much flexibility in integrating knowledge about the objects and their mutual relationships. The globally optimal configuration of an unknown number of objects is determined by sampling. Thus, knowledge about the objects is expressed in a more holistic way and characteristics of objects can be integrated beyond pixel-based relations. MPPs have shown to achieve good results in various object detection problems (e.g. Lafarge et al., 2010; Favreau et al., 2019).

The individual positions of bomb craters are not very descriptive to indicate areas that probably contain duds. Therefore, we are interested in deducing area-based information about the occurrence of warlike impacts. This is done by statistical modelling, e.g. of the probability density function (PDF) for a

\* Corresponding author

dud to occur. In general, parametric approaches may be used where an analytical model for the PDF is assumed, whose parameters are derived from the data. On the other hand, non-parametric approaches estimate the PDF directly from the data. This approach avoids having to select a PDF model and to estimate its parameters. We use kernel density estimation, a quite popular non-parametric technique (e.g. Scott, 2015).

This work is based on the method for bomb crater detection of Kruse et al. (2019), where MPPs are applied to single images. However, the appearance of aerial images can vary considerably. Consequently, in this paper we propose to combine the MPP results of multiple images covering the investigated area. To do so, detections from multiple roughly georeferenced images that refer to the same object are matched. Our approach allows to deal with inaccurate georeferencing and also takes into account that the detections should support each other. In the experiments, we investigate the benefit of using redundant information and assess sensitivity of the method with respect to the degree of bombing. Finally, similar to (Kruse et al., 2019), a probability map based on the automatically detected bomb craters is created by kernel density estimation.

## 1.2 Related Work

Here, we focus on applications of MPPs and on methods for detecting craters. In this context, we also include planetary craters, because they have a similar appearance as bomb craters.

Model knowledge can be integrated into MPPs in different ways. Typically, simple geometric primitives, which can be described by a small number of parameters, are employed to represent the objects to be detected. Rectangles are often used to extract buildings or other human-made objects in the scene (Chai et al., 2012; Brédif et al., 2013). In these papers, MPPs are applied to digital surface models. A rectangle is included in the object configuration during sampling if high gradient magnitudes of the heights at the rectangle border are present while rectangles overlapping with each other are penalized. Wenzel and Förstner (2016) use rectangles to interpret facades of buildings based on rectified images. In addition to rectangles, circles, e.g. to detect tree crowns in laser scanning data sets (Yu et al., 2012; Zhang et al., 2013) or ellipses, e.g. for the detection of flamingos (Descamps et al., 2008), seed products (Dubosclard et al., 2014) or boats in harbours (Craciun and Zerubia, 2013), are used. The latter work makes use of objects being locally aligned. In clusters of bomb craters, no specific patterns exist, which is why such prior information cannot be used in our case. Benedek (2017) proposes a method for extracting complex hierarchical object structures from digital images using different types of objects, namely ellipses, rectangles and isosceles triangles, thus considering multiple object models. Inside this MPP framework, object-subobject ensembles in parent-child relationship are admitted and corresponding objects may form coherent object groups. Bomb craters have no object-subobject relationships, so this type of model is not useful in the context of this work. Descombes (2017) uses circles and ellipses to detect cells in biological images. In order to improve the resulting poor approximation of the cell shapes, the object space may be defined as a dictionary of precomputed shapes. Such a dictionary can be obtained from previous segmentation methods (Poullain et al., 2015) or by constructing an exhaustive description of convex shapes inside a small region (e.g. bounded by 5 x 5 pixels; Cedilnik et al., 2018). However, we have another application domain and are not interested in the more precise shape of craters. To the best of our knowledge there are two contributions dealing with MPPs in the context of planetary crater detection (Troglio et

al., 2010; Solarna et al., 2017). To reduce the computational effort in the optimization process, Solarna et al. (2017) create a birth map from the available contour map via generalized Hough transform and Gaussian filtering. Other approaches are based on unsupervised (e.g. Meng et al., 2009) or supervised (e.g. Wetzler et al., 2005; Urbach and Stepinski, 2009) methods. The former apply common edge filters to highlight the edges in the image followed by a Hough transform to reconstruct the circular shape of the craters. However, an expert is needed to choose the best filters and the procedure is not very robust to noise. Supervised methods used for crater detection include boosting (Bandeira et al., 2012; Wang and Wu, 2019), support vector machines (Ding et al., 2013) and Convolutional Neural Networks (Cohen et al., 2016; Emami et al., 2019). Apart from (Kruse et al., 2018; 2019), we are aware of four contributions dealing with the detection of bomb craters in aerial wartime images. Jensen et al. (2010) use a two-step approach. First, candidates are searched via cross correlation with representative crater-templates. Afterwards, the candidates are classified by linear discriminant analysis. Merler et al. (2005) use different boosting approaches for the classification of image sections, which requires a high computational effort. Their result is a map of the spatial density of craters, an indicator for the risk of finding duds. Clermont et al. (2019) and Brenner et al. (2018) propose methods based on Convolutional Neural Networks. These algorithms need a large set of training data, which are currently not available (see Clermont et al., 2019). To the best of our knowledge, in connection with redundant image information, only (Brenner et al., 2018) make use of it to combine the individual detections (the outputs of the CNNs were converted to individual crater positions of each image). For that purpose, they apply a neighbourhood-based clustering, where neighbourhood is defined by a maximum, empirically determined, Euclidian distance. Finally, the clusters are replaced by their centroids. However, unlike in our case, double detections from different images are eliminated in order to not miss any crater in an investigated area. Thus, the idea of supporting detections is not pursued here, which is of importance for us, especially with respect to the proposed application scenario. Further details, e.g. on the clustering method or the georeferencing accuracy of the images, are not provided.

The listed articles show the potential of stochastic methods based on MPPs in different fields and they allow a flexible integration of knowledge about the objects, too. That is why we suppose the MPP-based procedure of Kruse et al. (2019) is suitable for us, but we assume the results of MPPs on single wartime images to be more error-prone due to their varying appearances, e.g. caused by differing contrast or sun positions. Hence, this paper proposes a new approach that combines the individual detection results of the MPP procedure stemming from multiple aerial wartime images covering the same location. As the investigated images are only coarsely georeferenced, matching of the individual detections referring to the same object (e.g. a bomb crater), becomes more challenging. We expect that by exploiting redundant image information, the above-mentioned aspects (contrast, position of the sun) and others can be partially counteracted and, thus, the impact map derived from the detected bomb craters will be improved.

## 2. MATHEMATICAL BASICS

### 2.1 Marked point processes

MPPs (Daley and Vere-Jones, 2003) are stochastic processes that describe random events depending on the position in space. They model a scene by an unordered set of objects of a certain type in

a bounded region  $F$  (a digital image). An object  $u_i = (p_i, m_i)$  is characterized by its position  $p_i = (x_i, y_i)$  and a parameter vector  $m_i$  (mark). A homogeneous Poisson point process assumes a purely random distribution of objects in space that are not related to each other and the probability  $P_\lambda(n)$  for the number of objects  $n$  is assumed to follow a discrete Poisson distribution. The *intensity parameter*  $\lambda$  describes the expected number of objects within  $F$  and the object positions are uniformly distributed. In practice, the assumption of complete randomness between the objects mostly does not apply, and more complex models are needed to measure the quality of the object configuration. To achieve this goal, a probability density  $h(\cdot)$  of the MPP can be formulated with respect to a reference point process, which is usually defined as the Poisson point process. We define  $h(\cdot)$  by a Gibbs energy  $U(\cdot)$  in the form of  $h \propto \exp - U(\cdot)$ . This energy consists of a data term  $U_D(\cdot)$  and a prior term  $U_P(\cdot)$ . The relative influence is modelled by a weight parameter  $\beta \in [0, 1]$ .

$$U(\cdot) = \beta \cdot U_D(\cdot) + (1 - \beta) \cdot U_P(\cdot). \quad (1)$$

The conformity of the object configuration with the input data is measured by  $U_D(\cdot)$ , while interactions between the objects are taken into account by  $U_P(\cdot)$ . The optimal object configuration  $u^* = \{u_1, \dots, u_n\}$  can be determined by maximizing the probability density  $h(\cdot)$ , i.e.  $u^* = \arg \max h(\cdot)$ . The probability density  $h(\cdot)$  is usually multi-modal and is defined in a configuration space with a variable dimension, because the number of objects can change. For this reason, a Reversible Jump Markov Chain Monte Carlo (RJMC) sampler in combination with simulated annealing is often used to estimate the global minimum to find an approximation of  $u^*$ .

## 2.2 Reversible Jump Markov Chain Monte Carlo sampling

In contrast to MCMC techniques (Metropolis et al., 1953; Hastings, 1970), RJMC methods (Green, 1995) can model scenes with an unknown number of objects. This is achieved by defining a set of changes (jumps) of the current configuration. These jumps are reversible, i.e. one can always return to a previous state. In each iteration  $t$ , the sampler proposes a change of the current object configuration from the predefined set of jumps. For each type of change, there is a density function  $Q_m$ , also called *kernel*.  $Q_m$  leads from an object configuration  $X_t$  to a new configuration  $X_{t+1}$  according to a probability  $Q_m(X_t \rightarrow X_{t+1})$ . The new configuration is accepted with a probability  $\alpha$  depending on the energy difference of states  $X_t$  and  $X_{t+1}$

$$\alpha = \min \left( 1, \frac{Q_m(X_{t+1} \rightarrow X_t)}{Q_m(X_t \rightarrow X_{t+1})} \cdot \exp - \left( \frac{U(X_{t+1}) - U(X_t)}{T_t} \right) \right). \quad (2)$$

In equ. 2, the kernel ratio  $Q_m(X_{t+1} \rightarrow X_t)/Q_m(X_t \rightarrow X_{t+1})$  describes the ratio of probabilities for changing the configuration from  $X_{t+1}$  to  $X_t$  and vice versa. The Gibbs energies (equ. 1) of the new and current object configuration are represented by  $U(X_{t+1})$  and  $U(X_t)$ , respectively.  $T_t$  is the temperature used in simulated annealing at iteration  $t$ . To minimize the energy, RJMC is combined with simulated annealing (Kirkpatrick et al., 1983). The temperatures  $T_t$  tends towards zero while  $t \rightarrow \infty$ . While a logarithmic cooling schedule guarantees convergence to the global optimum, usually a faster cooling scheme based on a geometric sequence is applied (Van Laarhoven and Aarts, 1987).

## 2.3 Kernel Density Estimation

Kernel density estimation allows estimating the PDF of a random variable directly from the data in a non-parametric way (Scott,

2015). Given a sample  $(x_1, x_2, \dots, x_n)$  drawn from a distribution with an unknown density  $p$ , an estimate  $\hat{p}$  of this density can be calculated via

$$\hat{p}(x) = \frac{1}{nh} \sum_{i=1}^n K \left( \frac{x - x_i}{h} \right). \quad (3)$$

Here,  $h$  is the bandwidth parameter and  $K$  is a kernel function (not to be confused with the kernels for RJMC sampling from Section 2.2). The kernel function  $K(k)$  has to be a non-negative function ( $K(k) \geq 0$ ) that integrates to one ( $\int K(k) dt = 1$ ). Equ. 3 can be thought of as an estimate of the PDF by averaging the effect of a set of kernel functions centred on each data point. For this purpose, often the Gaussian kernel is considered, but other functions (e.g. triangular) can also be used.

## 3. METHODOLOGY

In order to detect bomb craters in individual images, MPPs are combined with RJMC sampling and simulated annealing. This method, based on (Kruse et al., 2019), is outlined in Section 3.1. In Section 3.2, the combination of the individual detection results is described, which is the core contribution of this paper. Based on the final object configuration, a probability map differentiating between potentially contaminated and uncontaminated areas is created (Section 3.3).

### 3.1 MPP-based bomb crater detection

**3.1.1 Object model:** Like (Kruse et al., 2019), we model bomb craters as circles, each being described by its position  $(x, y)$  in the image and one mark, the radius  $r \in [r_m, r_M]$ , where  $r_m$  and  $r_M$  are the minimum and maximum value.

**3.1.2 Kernels:** In each iteration of sampling, the object configuration changes according to a kernel  $Q_m$  with proposition probability  $p_{Q_m}$ . There are four types of change and, thus, four kernels (see Section 2.2): the *birth* ( $Q_B$ ), *death* ( $Q_D$ ), *translation* ( $Q_T$ ) and *mark-variation* ( $Q_M$ ) kernels with the related proposition probabilities  $p_{Q_B}, p_{Q_D}, p_{Q_T}, p_{Q_M}$ . An object can be added to ( $Q_B$ ) or removed from ( $Q_D$ ) the current configuration. In the case of a death, a randomly selected object is removed whereas for a birth, the position of a new object is sampled from likely positions for bomb craters detected in the way described in Section 3.1.4. This procedure provides information about the size of the associated crater, which is used for the initialization of the circles radius. The kernel ratio in equ. 2 considers the probability of changing the configuration from  $X_t$  to  $X_{t+1}$  and vice versa; we model the kernel ratio of the birth event by

$$\frac{Q_B(X_{t+1} \rightarrow X_t)}{Q_D(X_t \rightarrow X_{t+1})} = \frac{p_{Q_D}}{p_{Q_B}} \cdot \frac{\lambda}{n}. \quad (4)$$

In equ. 4, the Poisson parameter  $\lambda$  describes the expected number of objects and  $n$  is the actual number of objects in the scene. In the case of the death event, the kernel ratio corresponds to the inverse birth rate. In addition, the position of an object can be changed by  $Q_T$ . For this purpose, a randomly chosen object is shifted from its current position by a random displacement vector in the local neighbourhood. The movement is realized in a given interval based on a uniform distribution and the kernel ratio is set to one. Finally,  $Q_M$  allows to change the marks of the object. A circle of the current configuration is randomly selected and its radius is changed to a new value drawn from a uniform

distribution within predefined intervals. The kernel ratio is also set to one.

**3.1.3 Energy function:** In each iteration step, we compare the current object configuration with the new one based on the Gibbs energy. Here, we describe the energy terms of equ. 1.

The **data energy**  $U_D(X_t)$  checks the consistency of the object configuration with the input data. Bomb craters are predominantly characterized by locally darker grey values in comparison with the surrounding area. This is mainly due to the shadow cast by the sun and its shape is often circular within the bomb craters. Consequently, bomb craters are assumed to have high gradient magnitudes in the transition region from dark to bright and more or less homogenous grey values inside. Following (Kruse et al., 2019), the data energy is modelled by

$$U_D(X_t) = U_G(X_t) + U_H(X_t). \quad (5)$$

The two terms are explained in detail below. Note that before evaluating the data energy for an object, the grey values in a window centred over the object centre with the size of the object's radius and additional  $n_{pix}$  pixels (set to  $n_{pix} = 20$ ) in each direction are normalized to the interval  $[0, 255]$ .

The first term in equ. 5 is related to **gradient magnitudes**. According to the previously made assumptions, a newly created or modified object leads to a reduction of the data energy if high gradient magnitudes occur along the edges of the object. We determine the gradients along the edge of the circle and model the corresponding data term by

$$U_G(X_t) = f_G \cdot \sum_{o_j \in X_t} \left( c - \frac{1}{n_b} \sum_{j=1}^{n_b} \nabla_{IMG_{p_j}}^b \right). \quad (6)$$

Here,  $\nabla_{IMG_{p_j}}^b$  is the component of the grey value gradient at the border pixel  $p_j$  in the direction of the normal vector of the object  $o_j$  pointing outside. To calculate the sum of the gradients along the border of the object,  $n_b$  pixels  $p_j$  are used. The edge of the object is approximated by a polygon with a constant number of  $n_p$  vertices (we use  $n_p = 32$  in our experiments). The term is weighted by a factor  $f_G$  and the constant  $c \geq 0$  ensures that the energy only decreases if the sum in equ. 6 is larger than  $c$ . Without considering  $c$ , objects with very small gradient magnitudes at the object border would already reduce the energy, so that the optimal configuration would consist of an extremely large number of (mostly false positive) objects.

The second term in equ. 5 requires the grey values inside the object to be **homogeneous**. Homogeneity is measured by the grey value standard deviation  $\sigma$  within the object, which is assumed to be higher for a false positive (FP) object than for a bomb crater. A new or modified object  $o_j$  increases the energy if  $\sigma_j$  is higher than a predefined threshold  $H_t$ , which results in

$$U_H(X_t) = f_H \cdot \sum_{o_j \in X_t} \max(0, \sigma_j - H_t). \quad (7)$$

In equ. 7, the energy term is weighted by  $f_H$ . The grey values of all pixels inside the object  $o_j$  are used to compute  $\sigma_j$ . In this context, a certain number of pixels around the border of the object can be excluded, because the shapes bomb craters may deviate slightly from the geometrical model. This possibility is controlled via a parameter  $H_e$  describing the width of the excluded area.

The **prior energy** term penalizes overlapping objects to avoid the accumulation of objects in regions with low data energy. Like (Kruse et al., 2019), we consider all possible combinations of overlapping object pairs  $o_i, o_j$ . The overlapping areas  $A_{ij}$  of the objects  $o_i$  and  $o_j$  as well as the respective relative overlapping areas  $A_{ij}/A_i(o_i)$  and  $A_{ij}/A_j(o_j)$  are computed. Here,  $A_i(o_i)$  and  $A_j(o_j)$  are the areas of the objects  $o_i$  and  $o_j$ , respectively. The prior energy with weight  $f_P$  becomes

$$U_P(X_t) = f_P \cdot \sum_{o_i, o_j} \max\left(\frac{A_{ij}}{A_i(o_i)}, \frac{A_{ij}}{A_j(o_j)}\right). \quad (8)$$

**3.1.4 Limiting the search space:** In order to reduce the computational effort for the MPPs in the sampling process, the search space in the image is restricted. For that purpose, we use the blob detector described in (Mallick, 2015) and implemented in OpenCV providing the coordinates centres of each valid blob as well as its size. Thus, during sampling the birth of an object is only possible at a blob location anymore. In order not to miss any craters, the parameters of the blob detector are selected appropriately (Section 4.1.2). In the first step of blob detection, the image is converted into several binary images by applying different thresholds. Starting at a minimum threshold  $B_{T,min}$ , this threshold is increased by step size  $B_{T,step}$  up to a maximum  $B_{T,max}$ . After extracting connected components from each binary image, they are grouped according to the distance of their centres to form blobs. Blobs located closer than  $B_D$  are merged. The algorithm also provides filter options (circularity, convexity, inertia ratio and size  $B_N$ ) and allows to detect only dark blobs, bright blobs or both types.

## 3.2 Fusing the results from multiple images

In the presence of multiple images, the MPP process is applied independently to all images, which results in a set of detected objects for each image. In a subsequent step, these detections have to be combined before generating an impact map. This means that detections from multiple images that refer to the same object (e.g. a bomb crater) have to be matched. A major problem is that the georeferencing information of the images may not be very accurate because precise orientation of a very large set of images is too expensive. It may in fact be impossible because camera information is usually not available for wartime images and it is often impossible to identify reliable ground control after such a long time. For instance, the images used in the experiments have a georeferencing accuracy of about 5 m - 40 m (cf. Section 4.1.1), which has to be taken into account when merging the detections. Furthermore, as already indicated in Section 1.2, it should be exploited that several detections of the same object in different images may be more likely to indicate a correct detection. Our method is presented in the following.

In order to merge the detections of every image that refer to the same object, we first select a master image (the detections within this image will be referred to as master detections MD) and detections from the remaining images are assigned to the respective MD. However, due to the coarse georeferencing, the positions of detections may differ from those of the MD by up to 40 m. For this reason, we define a search radius  $dist_{assign}$  in order to be able to cover all possible detections that belong to the same object. The assignment process starts with the detections from the master image. One MD is randomly selected and all neighbouring detections within the radius of  $dist_{assign}$  form a point set  $PS_i$  with this MD. Obviously, when searching for neighbours, as a rule only detections from other images, only one

detection per image as a neighbour (the closest) and each detection only once, can be considered. After all MD are taken into account, all the other detections from the remaining images are treated in the same way. It should be noted that due to the coarse georeferencing, problems with the assignment might occur. Since the distance between craters can be only a few meters, it is possible that e.g. nearby detections of a detection forming a  $PS$  belong de facto to another  $PS$ . How this can be addressed is discussed in the Outlook (Section 5). Finally, the gravity centre of the coordinates is calculated for each  $PS$ , whereas for  $PS$  containing a MD, the coordinate of the MD is used for the subsequent generation of the impact map. Note, that the centres of gravity from  $PS$  not including a MD are usually not located at the positions of the craters in the master image due to the inaccuracy of the georeferencing. Consequently, this will have an impact on the evaluation of the results and must be counteracted (see Section 4.1.3). Given the fact that  $PS$  with one (e.g. image errors) or a few (e.g. unfortunate shadows) detections may be uncertain, we define the following criterion: only  $PS$  consisting of at least  $MinNumDet_{PS}$  detections are preserved.

Figure 1 illustrates the principle of combining the individual detection results based on a subset of an aerial wartime image (master image) covered by three additional images. Here, the centres of the respective detection result from the MPP procedure are shown in red, orange, green and purple. First, one MD (red) is randomly selected (here the top left one) and all neighbouring detections within the radius of  $dist_{assign}$  form  $PS_1$  with the MD. According to the previously mentioned rules, the purple detection in the upper left corner is assigned to  $PS_1$  (next to the orange and green one). In analogy,  $PS_2$  is formed by the red, orange and purple detection. Now that all MD were taken into account, all the other detections from the remaining images are considered. Let us assume that we now assign the orange detections. The detections within  $PS_1$  and  $PS_2$  have already been used, so the remaining orange detection forms  $PS_3$ , to which the green detection is assigned. Only the purple detection in the upper right corner has not yet been considered, which then forms  $PS_4$ .

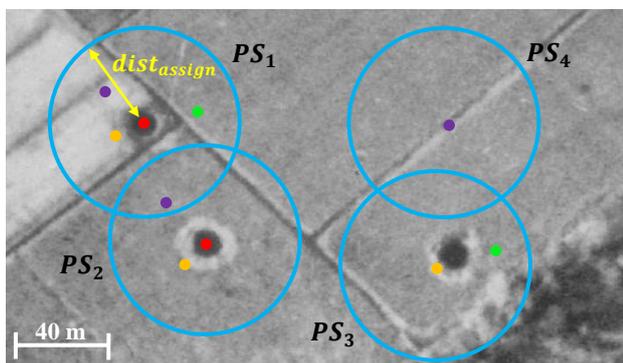


Figure 1: Subset of an aerial wartime image covered by a total of four different images with the centres of the respective detection results of the MPP procedure in red, orange, green and purple; these detections are assigned to point sets  $PS_i$  according to criteria described in the main text.

### 3.3 Impact map

We use the detections, obtained either from combined (Section 3.2) or non-combined detection results, to derive a probability for each location that there are duds nearby. The associated probability map is generated from the centres of the detections by means of kernel density estimation with the kernel function  $K(k) = (1 - |k|)$ . In this context, the bandwidth  $h$  in equ. 3

indicates how large the area of influence of a detection is. Using the probability map, the entire scene is classified into potentially contaminated and uncontaminated areas. For that purpose, a threshold  $p$  is applied to the probabilities resulting in an impact map. This threshold specifies from which probability within the probability map an area is classified as contaminated. We are interested in detecting areas that have a very high likelihood of containing a dud so that it makes sense to send a team of experts to that area to probe it. Thus, the focus of our work is on avoiding false detections, because these cause high unnecessary costs, i.e. areas classified as contaminated should in fact be contaminated.

## 4. EXPERIMENTS

We evaluate our method based on coarsely georeferenced panchromatic aerial wartime images. Section 4.1 describes the test data and the setup of the experiments. The main goal of our experiments is to investigate the influence of redundant image information on the results. First, we compare the combined with the detection results based on individual images. First, only images with an (almost) identical degree of bombing are considered (Section 4.2). Secondly, the extent to which additional images with fewer or no craters will affect the results is investigated (Section 4.3). While in both, Sections 4.2 and 4.3, the goal is to achieve results with a high F1-score, this is different in Section 4.4. Here, the procedure that combines the individual detection results is tuned to have a high correctness, which is important for the proposed application scenario. For this purpose, the parameter  $MinNumDet_{PS}$  (Section 3.2) is adapted.

### 4.1 Test data and test setup

**4.1.1 Data:** Our method is evaluated based on panchromatic aerial images acquired by the Allied forces during World War II over Lower Saxony, Northern Germany. They are provided by its explosive ordnance disposal service, who also generated the reference by manual annotation, i.e. for each reference crater we know its position and radius. These images are coarsely georeferenced (approx. 5 m - 40 m) and often distorted. They have ground sampling distances (GSDs) of approx. 0.17 m to 0.36 m and cover areas of about 3 km<sup>2</sup> to 12 km<sup>2</sup>. Each image consist of approx. 10.000<sup>2</sup> pixels and the number of bomb craters per image varies. A total of 186 images from three different regions are available. These regions include 95, 60 and 31 images, respectively, with the degree of bombing being (almost) the same for 52, 48 and 23 images of the three test regions, respectively, based on visual inspection. Hence, a region is much larger than an image within that region (in terms of the area they cover). The investigations are carried out as follows: From each of the three regions 6 (master) images are selected. Consequently, in the context of combined detection results (Section 3.2), all the images having an overlap with the respective master image are additionally considered. The amount of coverage varies locally because the footprints of the images inside a region are not aligned. Hence, each  $PS$  (Section 3.2) of a master image is covered by a different number of images (reported in the respective sections). The total of 18 images were selected in such a way that for each region there are two images with a rather difficult, moderate and easy image content (relative to the region), respectively. The focus of the investigations and developments is on rural sites. In densely built-up areas, it is not possible to clearly identify craters in the images because they are covered by the debris of destroyed buildings. The investigated images of the three regions are representative for certain cases (e.g. different lighting situations). Finally, it should be mentioned that the image content becomes more and more challenging for

the MPP procedure from region 1 to region 3, as more objects (e.g. forests, trees, houses and their shadows) that make correct detection more difficult are included in the images and the number of craters decreases. The quality of the images differs considerably; in particular, there are both underexposed and overexposed areas. To counteract this, Contrast Limited Adaptive Histogram Equalization (Pizer et al., 1987 – we use the OpenCV implementation) is applied. It requires two parameters (a block size  $C_b$  and a contrast limit  $C_l$ ), which are set to  $C_b = 60 \times 60 \text{ pixels}$  and  $C_l = 2$  based on a coarse visual inspection.

**4.1.2 Parameter settings:** Following (Kruse et al., 2019), we set the free parameters in our experiments to values that were determined empirically. If not specifically indicated, the parameters for all images and tests were set to identical values which makes the procedure more relevant for a potential use case.

We select the parameters of the blob detector (Section 3.1.4) as  $B_{T\_min} = 10$ ,  $B_{T\_max} = 245$ ,  $B_{T\_step} = 2$  and  $B_D = 5$ . As bomb craters or their shadows, respectively, are generally characterized by darker grey values than those in their surroundings, the procedure should only detect dark blobs. The parameters for filtering are set in a loose way, which allows craters to deviate from a circle. Similarly, the selection of  $B_N$  in the interval  $[B_{N\_m}, B_{N\_M}]$  makes it possible to detect bomb craters with different sizes. Depending on the GSD of the respective image, we set  $B_{N\_m}$  and  $B_{N\_M}$  in a way that blobs in-between diameters of 6 m and 24 m can be detected (note that the GSDs are only roughly known). Although selecting such loose filter restrictions results in many false detections, experiments have shown that a more restrictive choice can exclude the detection of many bomb craters in advance. On the other hand, it is also important that the minimum blob size  $B_{N\_m}$  is not set too small.

We weight the data and prior energy of the MPP equally, i.e.  $\beta$  from equ. 1 is set to  $\beta = 0.5$ . Simulated annealing uses a geometric cooling scheme by reducing the temperature  $T_t$  from equ. 2 using a factor  $f_T$  in the form  $T_t = T_0 \cdot f_T^t$ . We set the start temperature  $T_0 = 100$  and  $f_T = 0.9994$ . The lower and upper limit of the radius (Section 3.1.1) are derived from the minimum and maximum blob radius  $B_r \in \{B_{r\_min}, B_{r\_max}\}$  occurring in the image after blob detection, i.e.  $B_{r\_min} = r_m$  and  $B_{r\_max} = r_M$ . We set the proposition probabilities of the kernels (Section 3.1.2) to  $p_{Q_B} = 0.4$ ,  $p_{Q_D} = 0.4$  and  $p_{Q_T} = p_{Q_M} = 0.1$ . The probabilities for the translation and mark-variation event are relatively low, because a circle no longer has to be shifted significantly due to the size information provided by the blob detector. In order to avoid manual intervention,  $\lambda$  from equ. 4 is set to  $\lambda = \#blobs/20$ , where  $\#blobs$  is the number of blobs. The parameter  $c$  of the first data term ( $U_G(X_t)$ , equ. 6) is set to 1200 and the weight of that term is set to  $f_G = 1$ . The parameters of the term  $U_H(X_t)$  of the data energy in equ. 7 are set to  $f_H = 5$  and  $H_t = 15$ . As craters are not always exactly circular, we set  $H_e = 6$ . Finally, in connection with the prior energy (equ. 8), minor overlap of objects is possible with  $f_p = 10000$ . The initial configuration for the sampling process is an empty set of objects.

With regard to the investigations concerning redundant image information (Section 3.2), we set  $dist_{assign} = 40 \text{ m}$ , to ensure that all possible detections belonging to the same object can be found. Moreover,  $MinNumDet_{PS}$  is set to 3, because very few detections for the same object are more likely to be incorrect.

The bandwidth  $h$  (equ. 3) for kernel density estimation is varied based on the image scale. The threshold  $p$  for the probabilities is set in a way that the area around the centre of an object is always

classified as contaminated within a radius of 20 m for single detections. As areas of bomb craters and their immediate surroundings are likely to contain duds, the radius of 20 m is set relatively small in order to detect those areas that probably need to be probed by experts.

**4.1.3 Evaluation criteria:** As already mentioned, the evaluation of the results is affected by the inaccurate georeferencing for investigations related to Section 3.2. More precisely, the calculated centres of gravity from  $PS$  not including a MD are usually not located at the positions of the craters in the master image. Thus, the centres of the detections do not coincide with those of the respective reference, which would make the evaluation erroneous. To counteract this, these detections are shifted to the references of the master image as explained in the following, whereby detections whose distance to a reference centre is larger than 40 m are not affected. In order to address possible misallocations, a mean displacement vector is determined from  $PS$  with a MD obtained from a certain vicinity (600 m, empirically determined). This vector is used to first move all  $PS$  without MD (more precisely their coordinate centres of gravity) according to the particular offset and finally to the closest reference within 40 m. If only a few  $PS$  with MD are available for the determination of the displacement vector, it is of course less reliable. However, incorrect displacements are more likely to occur in densely bombed areas, so that there are usually sufficient  $PS$  with MD to derive the offset.

The pixel-based evaluation of the results is based on the impact map (Section 3.3) generated from the automatically detected bomb craters using the combined or non-combined detection results. The reference centres of the bomb craters are used for the generation of the reference impact map (same parameter setting as for the generation of the impact map from the object centres). The corresponding impact maps from the reference and the automatic detection are compared and each pixel is classified as being either a True Positive (TP), False Negative (FN), False Positive (FP) or True Negative (TN). A TP is a pixel that was correctly classified as contaminated in both, the reference and automatic detection. FN pixels have been classified as uncontaminated by the automatic detection although they are in fact contaminated. FP pixels were falsely classified as contaminated. Finally, a TN pixel was correctly classified as uncontaminated in both cases. The completeness CP (also known as recall in other areas of the field) is the percentage of the actually contaminated area found by our method, i.e.  $TP / (TP + FN)$ . The correctness CR (also referred to as precision) is the percentage of areas from the automatic detection that lie in areas which are actually contaminated, i.e.  $TP / (TP + FP)$ . The F1-score is the harmonic mean of these two measures, i.e.  $2 \cdot (CP \cdot CR) / (CP + CR)$ .

## 4.2 Validation of the use of redundant image information

The numerical values of completeness (CP), correctness (CR) and F1-score (F1) for the comparison of the non-combined (NON-COMB) and combined (COMB) detection results of the MPP procedure can be found in Table 1. In addition to the total area, the quality measures for each of the 18 investigated images (IMG) coming from three different regions (R) are shown. Furthermore, we show the number of bomb craters (NC) and the mean number of images considered for the combined detection results (NI).

Our results show that an average completeness and correctness of 62 % and 73 %, respectively, can be achieved if the MPP results are combined (columns 8-10). Compared to the non-combined results (columns 5-7) with a completeness and

correctness of 31 % and 55 %, respectively, the F1-score increases strongly. The F1-score of each image improves in 17 out of 18 cases, except for image XVII (region 3). Additionally, the gain in F1-score within this region is small compared to the other two. This can be explained by the comparatively small number of images taken into account (about half as many as in areas 1 and 2). Furthermore, some image areas may only be covered by one or a few other images, which is the case for image XVII. In consequence, craters found at first are removed again, which has a direct negative effect on the completeness. Similarly, when only a relatively small number of additional images is available, craters are also eliminated if they are not (or no longer) representative for the object model (e.g. if craters have been filled up in the meantime and therefore appear bright in the image). The comparatively more complex image content will also affect the results. On the other hand, as already mentioned, the F1-score increases (almost) everywhere, in many cases even enormously (e.g. I, III, VII, X, XV) if redundant image information is taken into account. This is especially the case if the examined image contains image errors, unfavourable shadows, has low contrast or if the craters are not representative for the used object model. While the former two aspects have an impact on the correctness, the latter two influence the completeness. An example related to completeness is illustrated in Figure 2. Figure 2a shows a subset of image I with the reference centres of the bomb craters in turquoise. Here the shadow cast by clouds (mainly the lower and left part in Fig. 2a) leads to poor contrast and furthermore some craters do not appear dark because they are filled with water (e.g. two of the three craters close together to the east of the river, Fig. 2a). The impact maps based on the resultant optimal object configuration after the sampling procedure (centres are shown in yellow) can be found in Figures 2b (non-combined) and 2c (combined). By combining the detection results, the areas falsely classified as uncontaminated almost completely vanish (Fig. 2b-c, red / dark green). The areas falsely classified as contaminated (Fig. 2b-c, pale blue) are marginal in both cases (e.g. such an area is present at the top left side of the crater diagonally right below the TN in Fig. 2b). Thus, and on the basis of the numerical values given in Table 1, our assumptions that results based on individual images are more error-prone and can be stabilized with additional image information, have been essentially confirmed.

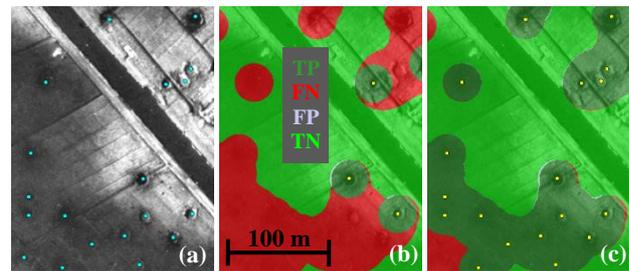


Figure 2: Subset of image I with the reference centres of the bomb craters in turquoise (a) and the superimposition of the corresponding impact map and its evaluation with TP-areas in dark green, FN-areas in red, FP-areas in pale blue and TN-areas in lime green for non-combined (b) and combined (c) detection results.

### 4.3 Considering different degrees of bombing

Unlike in Section 4.2, here we use additional images per region with different degrees of bombing, i.e. the number of bomb craters is not identical for a given area in each image. More precisely, images with fewer or no craters are added. In total, 63 additional images, 43 from region 1, 12 from region 2 and 8 from region 3, are available for this purpose. Table 2 shows the mean number of images taken into account for the combined detection results averaged over the regions (R); on the one hand for an identical level of bombing within the images as in Section 4.2 (NI-I) and for the varying degree (NI-V). Besides, the quality measures for the total areas (TA) based on all 18 images, with an identical (COMB-IDENT) and varying (COMB-VAR) degree of bombing, are shown. Note that the abbreviations regarding the quality measures are the same as in Table 1.

R	NI-I	NI-V	COMB-IDENT			COMB-VAR			
			CP [%]	CR [%]	F1 [%]	CP [%]	CR [%]	F1 [%]	
1	20	34	T A	62.1	72.7	67.0	64.6	61.1	62.8
2	18	20							
3	9	11							

Table 2: Evaluation results for the investigations based on a non-varying and varying degree of bombing.

R	IMG	NI	NC	NON-COMB			COMB		
				CP [%]	CR [%]	F1 [%]	CP [%]	CR [%]	F1 [%]
1	I	22	443	24	86	37	73	85	79
	II	16	114	31	45	37	70	72	71
	III	21	218	24	77	36	76	86	80
	IV	18	191	29	77	42	64	76	69
	V	22	424	61	94	74	80	83	82
	VI	19	43	51	66	58	75	78	76
2	VII	17	137	1	5	2	51	51	51
	VIII	18	33	0	0	n/d	44	40	42
	IX	18	17	13	25	17	58	41	48
	X	17	19	61	10	17	50	46	48
	XI	19	26	80	23	36	47	52	49
	XII	17	248	26	59	36	45	68	54
3	XIII	8	62	5	6	5	13	19	15
	XIV	9	24	2	2	2	4	13	6
	XV	8	52	4	8	5	15	33	21
	XVI	12	50	12	20	15	18	41	25
	XVII	8	11	59	20	30	9	11	10
	XVIII	9	37	20	13	16	26	25	25
<b>Total area</b>				<b>31</b>	<b>55</b>	<b>39</b>	<b>62</b>	<b>73</b>	<b>67</b>

Table 1: Evaluation results for the non-combined and combined detection results; (n/d: not defined).

The results in Table 2 show that if additional 63 images with other degrees of bombing are added to the 123 images, the F1-score of the results does not drop much (from 67.0 % to 62.8 %). However, the correctness decreases by approx. 12 %, while the completeness increases by about 3 %. The latter can be explained by the fact that some of the additionally considered images contain craters that are representative for our model and, thus, may lead to the preservation of such craters. On the other hand, more false detections may now accumulate on objects similar to bomb craters. By increasing the parameter  $MinNumDet_{PS}$  (Section 3.2) to 4, the correctness can be increased to likewise about 73 % (73.0 %), but the completeness decreases to 58.0 % (F1-score = 64.6 %). Thus, the procedure does not seem to be very sensitive to changes in the degree of bombing. Nevertheless, further tests still need to be carried out. In any case, the results are significantly better than those of the single image approach.

### 4.4 Focus on correctness

Similarly to (Kruse et al., 2019), we also want to tune the parameters of the algorithm with respect to the presented application scenario, i.e. the results should have the highest possible correctness, as long as completeness does not suffer too much. This is reasonable because the quality measures shown in

Section 4.2 (Table 1) are in general too low to integrate the results into the workflow of the explosive ordnance disposal service. An average correctness of about 73 % would lead to many areas being unnecessarily probed, resulting in high costs. We vary the parameter  $MinNumDet_{PS}$  in connection with the newly presented approach that combines the individual detection results of the MPPs (Section 3.2). Increasing  $MinNumDet_{PS}$  leads to more and more  $PS$  with a smaller number of detections being removed. This is why we increase  $MinNumDet_{PS}$  starting with  $MinNumDet_{PS} = 3$  (Fig. 3). As before, the quality indices are based on the total area of all 18 images using the same aerial images as in Section 4.2 (identical degree of bombing).

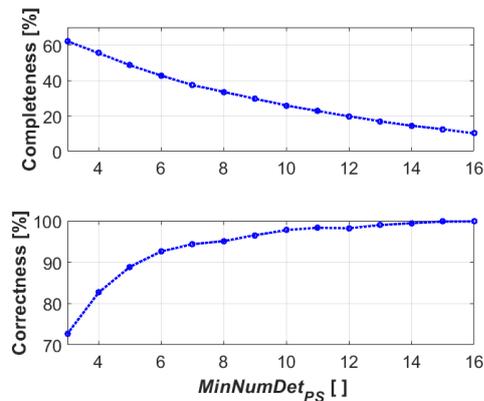


Figure 3: Completeness and correctness as a function of  $MinNumDet_{PS}$  for the total area of all 18 images.

The quality measures completeness and correctness as a function of parameter  $MinNumDet_{PS}$  are shown in Figure 3. It can be seen that the completeness decreases more or less linearly, until  $MinNumDet_{PS} = 7$  slightly more. The correctness is similar to a root function, i.e. initially it increases comparatively more strongly and from approx.  $MinNumDet_{PS} = 5$  the loss in completeness is higher than the gain in correctness. However, having in mind the proposed scenario, e.g. if  $MinNumDet_{PS}$  is set to 6, a correctness of more than 92 % can be achieved with a remaining completeness of nearly 43 %. From about  $MinNumDet_{PS} = 15$  onwards, the correctness reaches (almost) 100 %, whereas the completeness still decreases. While in (Kruse et al., 2019) the completeness was about 15 % for a correctness of approx. 96 %, here it is about twice as high (approx. 30 %) for  $MinNumDet_{PS} = 9$ . Note that the results are not directly comparable, as other images were considered. However, it seems that the images used in this paper are more challenging, as the quality measures based on single images are generally worse than those in (Kruse et al., 2019).

## 5. CONCLUSION AND OUTLOOK

In this paper, MPPs are used for the automatic detection of bomb craters in aerial wartime images. Typically, multiple images of the same area exist. To make use of this additional information, we propose an approach that combines the individual detection results of the MPP procedure. The detections serve as indicators for the areas where duds may be located and, thus, they are used to generate an impact map that provides a quick overview of potentially contaminated areas. The approach was evaluated using 18 coarsely georeferenced panchromatic images. We could show that by combining the individual detection results of images having a same degree of bombing, the F1-score of the results is considerably increased from 39 % (based on single images) to 67 % (based on multiple images). Errors in individual images can

be partially compensated. Thus, our assumption that results based on individual images are more error-prone has been essentially confirmed. If additional images with less or no craters are added, further investigations revealed that the F1-score of the results drops comparatively little to about 63 % and is therefore still significantly better than in the case of the single image approach. In general, the results with a correctness of about 73 % are not good enough for an integration into the workflow of the explosive ordnance disposal service. Too many areas would have to be probed unnecessarily, resulting in high costs. In this context, our final experiments show that the correctness can be increased at the expense of completeness by varying only one parameter within the proposed approach that combines detections. Thus, the procedure is more attractive for the proposed use case. Based on the investigated images, e.g. a correctness of approx. 95 % with a remaining completeness of almost 34 % can be achieved.

As already indicated, the combination of the individual detection results (Section 3.2) can lead to misallocations due to the rough georeferencing of the provided images. This is possible because the assignment process considers detections within a radius of 40 m in order to theoretically be able to detect all possible detections belonging to the same object. However, this can lead to the fact that adjacent detections of a given detection actually represent the detection of another object. In addition, inaccuracies may occur during the subsequent evaluation (Section 4.1.3). In order to counteract these shortcomings, it would be advisable to first improve the co-registration (e.g. Zitová and Flusser, 2003) of the respective images. First investigations have shown that the co-registration process may become more difficult, particularly due to the different appearances (e.g. caused by seasonal changes, noise) and missing camera information of aerial wartime images. In this context (Brenner et al., 2017) propose to perform a sampling of features at a fixed scale, because automatic scale space feature detection is too unstable for the given images. To refine the resulting registration and account for parallax effects, they apply a deformable fine registration approach. However, the paper does not provide any information on how accurate images are aligned afterwards. In addition, the process of co-registration should be robust when considering a potential use case. Therefore, we will address this and the concept from (Brenner et al., 2017) in future investigations. With a successful co-registration in meter range, the search radius  $dist_{assign}$  (Section 3.2) for neighbouring detections could be considerably reduced (e.g. to the radius of the considered detection) and, thus, ambiguities during assignment are avoided. Furthermore, given the fact that certain areas may be covered by hundreds of images, we intend to add more images in future experiments and see how this affects the results. Finally, a more extensive evaluation based on additional master images will be carried out.

## ACKNOWLEDGEMENTS

The authors would like to thank the State Office for Geoinformation and Surveying of Lower Saxony and its Explosive Ordnance Disposal Service as a department of the regional directorate Hamelin-Hanover for providing the data and financial support to this project.

## REFERENCES

Bandeira, L., Ding, W., Stepinski, T.F., 2012. Detection of sub-kilometer craters in high resolution planetary images using shape and texture features. *Advances in Space Research*, 49 (1), pp. 64–74.

- Benedek, C., 2017. An Embedded Marked Point Process Framework for Three-Level Object Population Analysis. *IEEE Transactions on Image Processing: a publication of the IEEE Signal Processing Society*, 26 (9), pp. 4430–4445.
- Brédif, M., Tournaire, O., Vallet, B., Champion, N., 2013. Extracting polygonal building footprints from digital surface models: A fully-automatic global optimization framework. *ISPRS Journal of Photogrammetry and Remote Sensing*, 77, pp. 57–65.
- Brenner, S., Zambanini, S., Sablatnig, R., 2017. Image Registration and Object Detection for Assessing Unexploded Ordnance Risks - A Status Report of the DeVisOR Project. *Proceedings of the OAGM & ARW Joint Workshop*, pp. 109–110.
- Brenner, S., Zambanini, S., Sablatnig, R., 2018. Detection of Bomb Craters in WWII Aerial Images. *Proceedings of the OAGM Workshop*, pp. 94–97.
- Cedilnik, N., Debreuve, E., De Graeve, F., Besse, F., Descombes, X., 2018. SPADE: Small Particle Detection Method Using A Dictionary Of Shapes Within The Marked Point Process Framework. *IEEE International Symposium on Biomedical Imaging (ISBI)*, Washington, DC, United States.
- Chai, D., Förstner, W., Yang, M.Y., 2012. Combine Markov random fields and marked point processes to extract building from remotely sensed images. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 1-3, pp. 365–370.
- Clermont, D., Kruse, C., Rottensteiner, F., Heipke, C., 2019. Supervised detection of bomb craters in historical aerial images using convolutional neural networks. *ISPRS Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-2/W16, pp. 67–74.
- Cohen, J.P., Lo, H.Z., Lu, T., Ding, W., 2016. Crater Detection via Convolutional Neural Networks. <http://arxiv.org/abs/1601.00978> (28 January 2020).
- Craciun, P., Zerubia, J., 2013. Unsupervised marked point process model for boat extraction in harbors from high resolution optical remotely sensed images. *IEEE International Conference on Image Processing (ICIP)*, pp. 4122–4125.
- Daley, D.J., Vere-Jones, D., 2003. *An Introduction to the Theory of Point Processes: Volume 1: Elementary Theory and Methods* (2nd edition). Springer, New York.
- Descamps, S., Descombes, X., Béchet, A., Zerubia, J., 2008. Automatic flamingo detection using a multiple birth and death process. *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1113–1116.
- Descombes, X., Zerubia, J., 2002. Marked point process in image analysis. *IEEE Signal Processing Magazine*, 19 (5), pp. 77–84.
- Descombes, X., 2017. Multiple objects detection in biological images using a marked point process framework. *Methods*, 115, pp. 2–8.
- Ding, M., Cao, Y., Wu, Q., 2013. Novel approach of crater detection by crater candidate region selection and matrix-pattern-oriented least squares support vector machine. *Chinese Journal of Aeronautics*, 26 (2), pp. 385–393.
- Dubosclard, P., Larnier, S., Konik, H., Herbulot, A., Devy, M., 2014. Automatic method for visual grading of seed food products. *International Conference on Image Analysis and Recognition (ICIAR)*, 1, pp. 485–495.
- Emami, E., Ahmad, T., Bebis, G., Nefian, A., Fong, T., 2019. Crater Detection Using Unsupervised Algorithms and Convolutional Neural Networks. *IEEE Transactions on Geoscience and Remote Sensing*, 57 (8), pp. 5373–5383.
- Favreau, J.-D., Lafarge, F., Bousseau, A., Auvolat, A., 2019. Extracting Geometric Structures in Images with Delaunay Point Processes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. DOI: 10.1109/TPAMI.2018.2890586.
- Green, P.J., 1995. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, 82 (4), pp. 711–732.
- Hastings, W.K., 1970. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57 (1), pp. 97–109.
- Jensen, L., Drauschke, M., Förstner, W., 2010. Automatische Detektion von Bombentrümmern in digitalisierten Luftbildern des 2. Weltkriegs. In: *DGPF Tagungsband 19/2010*, pp. 228–235.
- Kirkpatrick, S., Gelatt, C.D., Vecchi, M.P., 1983. Optimization by simulated annealing. *Science*, 220 (4598), pp. 671–680.
- Kruse, C., Rottensteiner, F., Hoberg, T., Ziem, M., Rebke, J., Heipke, C., 2018. Generating impact maps from automatically detected bomb craters in aerial wartime images using marked point processes. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, IV-3, pp. 127–134.
- Kruse, C., Rottensteiner, F., Heipke, C., 2019. Marked point processes for the automatic detection of bomb craters in aerial wartime images. *ISPRS Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-2/W13, pp. 51–60.
- Kumar, S., Hebert, M., 2006. Discriminative random fields. *International Journal of Computer Vision*, 68 (2), pp. 179–201.
- Lafarge, F., Gimel'farb, G., Descombes, X., 2010. Geometric feature extraction by a multimarked point process. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32 (9), pp. 1597–1609.
- Mallick, S., 2015. Blob Detection Using OpenCV (Python, C++). <https://www.learnopencv.com/blob-detection-using-opencv-python-c/> (23 January 2020).
- Meng, D., Yunfeng, C., Qingxian, W., 2009. Method of Passive Image Based Crater Autonomous Detection. *Chinese Journal of Aeronautics*, 22 (3), pp. 301–306.
- Merler, S., Furlanello, C., Jurman, G., 2005. *Machine Learning on Historic Air Photographs for Mapping Risk of Unexploded Bombs*. In: Roli, F., Vitulano, S. (Eds.) *Image Analysis and Processing – ICIAP 2005*. Lecture Notes in Computer Science (LNCS), Springer, Berlin, Heidelberg, 3617, pp. 735–742.
- Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E., 1953. Equation of state calculations by fast

computing machines. *The Journal of Chemical Physics*, 21 (6), pp. 1087–1092.

Pizer, S.M., Amburn, E.P., Austin, J.D., Cromartie, R., Geselowitz, A., Greer, T., Romeny, B.t.H., Zimmerman, J.B., Zuiderveld, K., 1987. Adaptive Histogram Equalization and its Variations. *Computer Vision, Graphics, and Image Processing*, 39 (3), pp. 355–368.

Poulain, E., Prigent, S., Soubies, E., Descombes, X., 2015. Cells detection using segmentation competition. *IEEE International Symposium on Biomedical Imaging (ISBI)*, pp. 1208–1211.

Scott, D.W., 2015. *Kernel Density Estimators*. In: *Multivariate Density Estimation*. John Wiley & Sons, Inc, pp. 137–216.

Solarna, D., Moser, G., Le Moigne, J., Serpico, S.B., 2017. Planetary Crater Detection and Registration Using Marked Point Processes, Multiple Birth and Death Algorithms, and Region-Based Analysis. *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pp. 2337–2340.

Troglio, G., Benediktsson, J., Moser, G., Serpico, S., 2010. Crater Detection Based on Marked Point Processes. *International Geoscience and Remote Sensing Symposium (IGARSS)*, pp. 1378–1381.

Urbach, E.R., Stepinski, T.F., 2009. Automatic detection of sub-km craters in high resolution planetary images. *Planetary and Space Science*, 57 (7), pp. 880–887.

Van Laarhoven, P.J., Aarts, E.H., 1987. *Simulated Annealing: Theory and Applications*. Springer Netherlands, Dordrecht.

Wang, Y., Wu, B., 2019. Active Machine Learning Approach for Crater Detection from Planetary Imagery and Digital Elevation Models. *IEEE Transactions on Geoscience and Remote Sensing*, 57 (8), pp. 5777–5789.

Wenzel, S., Förstner, W., 2016. Facade Interpretation Using a Marked Point Process. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, III-3, pp. 363–370.

Wetzler, P.G., Honda, R., Enke, B., Merline, W.J., Chapman, C.R., Burl, M.C., 2005. Learning to Detect Small Impact Craters. *IEEE Workshops on Applications of Computer Vision (WACV/MOTION)*, pp. 178–184.

Yu, Y., Li, J., Guan, H., Wang, C., Cheng, M., 2012. A marked point process for automated tree detection from mobile laser scanning point cloud data. *International Conference on Computer Vision in Remote Sensing (CVRS)*, pp. 140–145.

Zhang, J., Sohn, G., Brédif, M., 2013. Single tree detection from airborne laser scanning data using a marked point process based method. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, II-3/W1, pp. 41–46.

Zitová, B., Flusser, J., 2003. Image registration methods: a survey. *Image and Vision Computing*, 21 (11), pp. 977–1000.