

THERMAL ANOMALY DETECTION BASED ON SALIENCY ANALYSIS FROM MULTIMODAL IMAGING SOURCES

A. Sledz¹, C. Heipke

Institute of Photogrammetry and GeoInformation, Leibniz University Hannover, Germany
sledz@ipi.uni-hannover.de

KEY WORDS: Thermal infrared imaging, thermal anomaly, saliency map, Dempster-Shafer theory, Random Forest, distributed heating systems

ABSTRACT

Thermal anomaly detection has an important role in remote sensing. One of the most widely used instruments for this task is a Thermal InfraRed (TIR) camera. In this work, thermal anomaly detection is formulated as a salient region detection, which is motivated by the assumption that a hot region often attracts attention of the human eye in thermal infrared images. Using TIR and optical images together, our working hypothesis is defined in the following manner: a hot region that appears as a salient region only in the TIR image and not in the optical image is a thermal anomaly. This work presents a two-step classification method for thermal anomaly detection based on an information fusion of saliency maps derived from both, TIR and optical images. Information fusion, based on the Dempster-Shafer evidence theory, is used in the first phase to find the location of regions suspected to be thermal anomalies. This classification problem is formulated as a multi-class problem and is carried out in an unsupervised manner on a pixel level. In the following phase, classification is formulated as a binary region-based problem in order to differentiate between normal temperature variations and thermal anomalies, while Random Forest (RF) is chosen as the classifier. In the second phase, the classification results from the previous phase are used as features along with temperature information and height details, which are obtained from a Digital Surface Model (DSM). We tested the approach using a dataset, which was collected from a UAV with TIR and optical cameras for monitoring District Heating Systems (DHS). Despite some limitations outlined in the paper, the presented innovative method to identify thermal anomalies has achieved up to 98.7 percent overall accuracy.

1. INTRODUCTION

Anomaly detection is the identification of unusual elements, occurrences or observations, which raise suspicions because of a significant deviation from the majority of data or expected behaviour. In the context of thermal anomaly detection, the goal is to identify regions with irregular temperatures that deviate from their surroundings. Despite the fact that cold spots may also be viewed as thermal anomalies, this study concentrates exclusively on hot spot detection as thermal anomalies.

This study is based on the assumption that a hot spot, a region (group of pixels) in a Thermal InfraRed (TIR) image with higher average temperature than a surrounding area, is salient, i.e. attracts attention of the human eye in thermal images. On the other hand, an entity in an optical image, being a physical body or a region, that stands out relative to its neighbourhood, should be salient also. The background of this study is the detection of leakages of underground district heating systems (DHS). Such leakages produce hot areas underneath the surface and possibly thermal anomalies on the surface, which typically cannot be detected in the visible spectrum from above. We note, in passing, that a poorly insulated roof of a building is also being detected using our approach. It is, however, easy to separate the two cases, as we assume to have access to height data also.

The main hypothesis proposed here is therefore that a hot spot, that appears as a salient region only in the TIR image and is not salient in the optical image, is a thermal anomaly. As seen in the following chapters, this hypothesis tends to be valid with a number of exceptions only. Chapter 3 shows that our saliency map (a modified version of the saliency model proposed by Itti et al. (1998a)) may be considered as a belief function in the sense of the Dempster-Shafer Theory (DST, Dempster, 1967; Shafer,

1976). One of the key advantages of DST is that it can explicitly deal with class unions in a classification problem. Rottensteiner et al. (2005) was one of the first to demonstrate that DST can be used in an unsupervised pixel classification of high-resolution remotely sensed data of different origins. In the context of this work, DST is used to identify candidates for thermal anomalies by combining information from TIR and optical images in the form of their corresponding saliency maps on a pixel-per-pixel basis. To reduce the resulting high number of false alarms, we then carry out a second binary, supervised region-based classification, based on the Random Forest (RF) classifier (Breiman, 2001). During this phase, thermal anomaly candidates detected by DST and their respective surroundings are treated as regions. The classification is based on the temperature measured with the TIR camera, height information in terms of a Digital Surface Model (DSM) and the classification results of DST. The rationale for using height information is twofold: (a) persons, cars and other objects which can cause false alarms stand out with respect to the ground surface and can thus be detected in the height data, (b) in the presence of height discontinuities, e.g. at the border of buildings, a region detected by DST typically has a mixed thermal signal coming from both, the roof and the neighbouring ground. As a consequence, thermal anomalies in such parts of the image are bound to be false alarms, too.

The paper is organized in the following way: Chapter 2 gives an overview of the related work on leakage detection in DHS and how saliency analysis is applicable in such particular case. Chapter 3 introduces the developed method, while Chapter 4 presents the experiments, the dataset and the findings. The final chapter sets out conclusions and open questions for future work.

¹ Corresponding author

2. RELATED WORK

A traditional District Heating System (DHS) uses the concept of co-generation and distribution of heat produced as waste heat during the generation of electricity. A maintenance plan of a DHS is focused on a statistical evaluation of the hazards to the network. The problem of underground leakage can be remedied by thermal infrared (TIR) airborne imaging of DHSs, which prevents interference with the operating cycle. This technology makes it possible to detect surface temperature variations. UAV-based thermography provides a rapid response for thermal anomaly detection and localization with relatively low cost compared to manned flight thermography. A common problem of both, airborne and UAV-based thermography is that besides leakages a number of other objects also exhibit local temperature maxima. Examples are recently parked cars, chimneys on a roof, pedestrians or street lamps. Thus, when using only TIR images, the probability of high false alarms is rather high.

First methods for large-scale monitoring by manned flight thermography were already studied many decades ago, e.g. by Ljungberg et al. (1988) and Axelsson (1988). Friman et al. (2014) present a system for the automated analysis of TIR images to find leaks in DHS pipes. Leakages are located by choosing the warmest, in the range of 0.005% to 0.5%, of all pixels over the pipe. As the authors suggest, one way to minimize the number of false alarms is to eliminate candidate leak detections in close proximity to buildings, accounting to about 20% of all false alarms. In order to reduce the false alarm rate associated with abnormal temperatures around buildings, buildings are masked out using a watershed transformation followed by a classification step based on Adaboost. Also Berg et al. (2014) address the problem of reducing the false alarm rate among potential leakages in district heating networks, detected in airborne TIR images. A region-based shape representation of the detection and proximity to buildings are used for false alarm reduction. Nonetheless, false alarms in existing methods remain the most difficult problem.

Zhong et al. (2019) present a saliency-based DHS leakage detection method; an infrared saliency map is created to enhance the leakage targets, while the pipeline location comes from a Geographic Information System (GIS). A local saliency map shows small image regions distinct with respect to their local neighbourhoods, where intensity and orientation features are adopted in the saliency analysis. On the other hand, a global saliency map represents saliency by calculation of the seldomness of features across the entire scene. In the final step, adaptive target segmentation by maximum entropy permits the automatic detection of potential leakage targets in the fused saliency map.

Researchers have used various theories or methods to describe and create saliency maps for different applications. Saliency models analyse the distinctiveness of image regions with respect to their local neighbourhoods (Borji, 2012). The first saliency computational model was developed by Itti et al. (1998a), who built a saliency map by implementing local centre-surround operations on low-level visual features. Harel et al. (2006) used graph algorithms and dissimilarity measurements to derive their graph-based visual saliency (GBVS) model. The attention by information maximization model (AIM) developed by Bruce and Tsotsos (2005) is calculated by quantifying the auto-information on each local image patch. Qi et al. (2013) were of the first to establish the link between thermal target detection and saliency detection. In their work, the authors formulate the problem of small infrared target detection as salient region detection, which was inspired by the fact that a small target can often attract attention of human eyes in infrared images.

There are two reasons for using the saliency model suggested by Itti et al. (1998a) in this study. The first is the ease with which the original model could be changed to serve the purpose of detecting hot spots. The second justification is based on the advantage that the size of the area to be detected can be controlled by choosing the centre and surrounding scales of the model: a detailed discussion can be found in the following sections. In addition, the current study demonstrates how the combination of TIR and optical image data, in form of their saliency maps, can be used to detect thermal anomalies.

3. ANOMALY DETECTION

3.1. Terminology and overview

This chapter introduces and discusses a method for thermal anomaly detection based on saliency analysis with TIR and optical images as information sources. Prior to the discussion of the details of the method, basic terminology and assumptions are presented:

- Hot spot – region (connected group of pixels) in a TIR image that has higher average temperature than the surrounding area. Taking into account that part of image processing is done on raw pixel values, a hot spot is also a region with higher intensity values than its surroundings, since the TIR camera in use is configured to display hotter temperatures as higher intensity.
- Cold spot – region (connected group of pixels) in a TIR image that has lower average temperature than the surrounding area.
- Salient region – a certain part of an image, that may depict an entity or be a region, that appears to an observer to be distinct from its surroundings. Following this definition, cold and hot spots are salient regions.
- Thermal anomaly – a hot spot that occurs as a salient region only in the TIR image and not in the optical image.

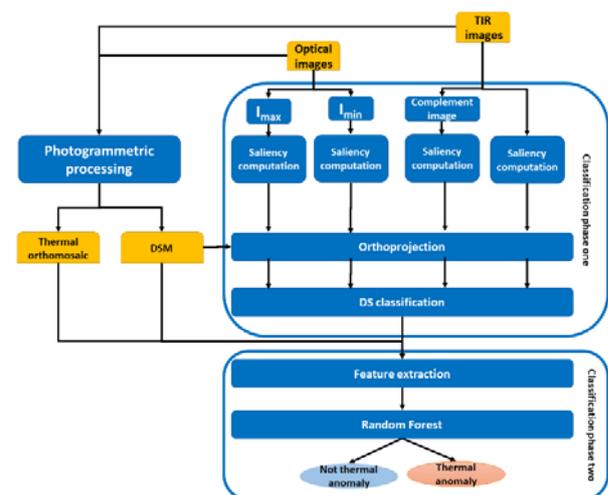


Figure 1: Method overview (see text for details)

The developed method, as outlined in Figure 1, consists of three main components:

- Photogrammetric processing – the processing was carried out employing rigorous photogrammetric techniques (see Sledz et al., 2018; 2020 for details) for the TIR and optical images captured from an UAV. The products of photogrammetric processing are: a thermal orthomosaic containing intensity values (raw sensor data), which are then converted to temperature values; an optical orthomosaic with the same Ground Sampling Distance (GSD) as the thermal

one and a DSM, which is derived from optical images, in a raster form with the same GSD as the thermal orthomosaic.

- Classification phase one – this step consists of a saliency computation and a multi-class DST-based per-pixel classification. Note, that while this separation of the different functional components suggests that saliency maps are computed from the orthomosaics, we actually derive saliency maps from the images instead, and then project the saliency maps into object space. The outcome of this step are regions containing thermal irregularities, which need be further investigated. The details are presented and discussed in the section 3.3.
- Classification phase two – this step comprises a binary region-based classification of the thermal irregularities from the preceding stage, making use also of a DSM, see section 3.4. The main purpose of this second stage is the reduction of false alarms.

3.2. Saliency analysis

3.2.1 The method according to Itti et al. (1998): Saliency analysis highlights regions in a picture that stand out from their neighbours. A saliency map is the result of an image transformation that assigns a unique quality to each pixel in an image based on the degree to which the pixel varies in its neighbourhood.

The model to compute a saliency map relates to the so-called "feature integration theory", which describes human visual search strategies. Multiple spatial locations compete for saliency within each map, and only locations that locally stand out from their surroundings will sustain. The computation consists of a number of steps: (i) feature representation in the form of image pyramids of colour, intensity and orientation (obtained by Gabor filters with different orientation angles); (ii) a centre-surrounding difference between pyramid entries (see eq. (1)), (iii) cross-scale addition and normalization for the construction of the final saliency map.

Nine spatial scales, $\sigma \in [0, \dots, 8]$, are created by progressively low-pass filtering and subsampling the input image, yielding horizontal and vertical image-reduction factors ranging from 1:1 (scale zero, $\sigma = 0$) to 1:256 (scale eight, $\sigma = 8$) in eight octaves. Each feature is computed by a set of "centre-surround" differences (see eq. (1)). Centre-surround difference \ominus is computed by interpolating the image at scale s to the finer scale c , represented by $I^c(s)$ in eq. (1), and then subtracting the images. The centre is a pixel at scale c , $c \in \{2, 3, 4\}$, and the surround is the corresponding pixel at scale $s = c + \delta$, with $\delta \in \{3, 4\}$. Using several scales not only for c but also for s yields truly multiscale features, by including different size ratios between the centre and surround regions.

$$I(c, s) = |I(c) \ominus I(s)| = |I(c) - I^c(s)| \quad (1)$$

One challenge in integrating the different feature maps of colour, intensity and orientation is that they represent non-commensurable modalities with different dynamic range. In addition, salient regions, which appear strongly in only a few maps, may be obscured by noise or by the less-salient regions in the remaining maps, and thus get lost. To address this difficulty, Itti et al (1998a) employs a map normalization operator $N(\cdot)$ described by eq. (2). Such an operator globally promotes maps in which there are a small number of high peaks, while globally suppressing maps containing multiple comparable responses.

$$N(I_{map}) = (I_{map})_0^M \cdot (M - \bar{m})^2 \quad (2)$$

where $(I_{map})_0^M$ stands for the normalization of the values between $[\min(I_{map}), \max(I_{map})]$ of the map to a fixed range of

$[0, \dots, M]$, in order to eliminate modality-dependent amplitude differences (without loss of generality M can be set to 1). \bar{m} is the average of all local maxima in a 3*3 pixel neighbourhood.

Feature maps of colour, intensity and orientation are combined into three results (called "conspicuity maps" by Itti et al., 1998) at scale $\sigma = 4$. They are obtained through across-scale addition, which consists of downscaling or upscaling each map to scale 4 and point-by-point addition. In the end, the saliency map is calculated as an average of normalized conspicuity maps.

3.2.2 The method used in this work: First, since the TIR images are single-channel images, the saliency map is determined solely on the basis of intensity and orientation features, colour features are not used. Also, while the original approach aims to identify regions that are more likely to attract visual attention (these could be bright or dark), the purpose of saliency analysis usage in this work is to find hot spots. Hot spots are characterised by a higher temperature (not a lower one) and therefore appear brighter than its surroundings. However, the scale-difference operation, shown in eq. (1), uses absolute differences to be sensitive to positive (bright) and to negative (dark) changes. In order to only detect bright regions, the equation is modified as shown in eq. (3), introducing a threshold th_{diff} . To extract only bright regions the value for th_{diff} is set to 0. When th_{diff} is set to minus infinity eqs. (1) and (3) are identical. Figure 2 demonstrates the effect of th_{diff} during saliency computation. The picture in the centre shows the saliency map overlaid on the input image with $th_{diff} = -\infty$, being compatible with the original method. It can be seen that, beside the hot street lamp, that is highlighted by red circles, the cold garbage bins, that are highlighted by yellow circles, are also salient. This is the direct influence of the absolute operator introduced in the eq. (1). However, when $th_{diff} = 0$ (see right image at Figure 2) the garbage bins are no longer salient. On the other hand, the surface under the garbage bins is hotter and, as the consequence, it is detected as salient.

$$I(c, s) = |\max(I(c) - I^c(s), th_{diff})| \quad (3)$$

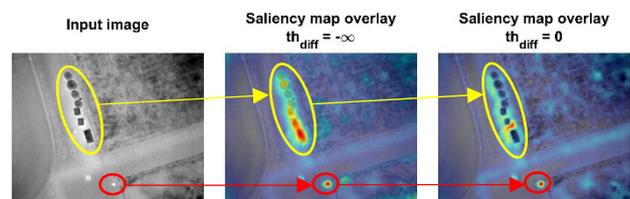


Figure 2: Influence of th_{diff} on saliency computation. The saliency map overlay, shown in the centre and right images has the following interpretation: no colour represents zero saliency, blue-greenish represents weak saliency, red represents high saliency.

Another extension of the original method concerns the normalization: the original approach attempts to identify a region that is most likely to attract visual attention, in terms of its importance related to other objects in a scene. This is done by map normalization, i.e. by comparing the global maximum M of the entire map with the average of all local maxima \bar{m} . When the difference is significant, M stands out, and during the subsequent combination with other maps contributes the most to the final saliency result, i.e. the map is strongly promoted. When the difference is negligible, the map does not contain anything special and therefore will only marginally contribute to the final saliency result.

In contrast, our aim is to also preserve smaller differences in temperature, which tend to be small local peaks in the maps. This goal can be reached by restricting the initial normalization range

from $[min(I_{map}), max(I_{map})]$ to $[pr(I_{map}, p_{min}), pr(I_{map}, p_{max})]$, where pr stands for percentile and p_{max} and p_{min} are the percentile values of the I_{map} . While the use of p_{max} less than 100 is intended to promote local peaks, the use of p_{min} higher than 0 results in suppressing noise. Figure 3 demonstrates the influence of the various limits involved in normalization (the meaning of the colours is the same as in Figure 2). It should be noted, that for the results shown in Figure 3, th_{diff} is set to zero. The upper left picture is the input image with four hot entities: the group of people (top, in the middle), the square manhole (top, left), the round manhole (centre) and a hot spot (centre, right) that was caused by the DHS. The upper right picture shows the outcome of the usage of $N(\cdot)$ as proposed by Itti et al. (1998a): the outcome is that only the group of people is salient and the rest of the hot entities is not. The left picture at the bottom shows the outcome for $p_{max} = 99\%$. The outcome shows the desired result: all four entities are salient. The right picture at the bottom row shows the outcome where, in addition, $p_{min} = 1\%$. All four entities are still salient and the result contains less noise, visible in terms of less bluish effects of the overlay. It should be noted that the saliency map is thus very sensitive to the selection of p_{max} . Typically, a change of p_{max} by 1 percent can lead to a significant change. The choice of p_{min} , however, is much less critical.

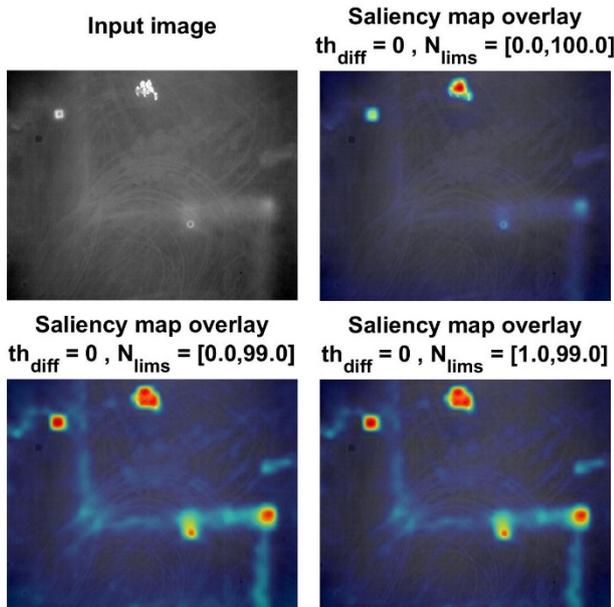


Figure 3: Effect of the normalization limits on the saliency computation (for the meaning of colour, see caption of Figure 2).

During this study, it was found that optical images should also be viewed differently than in the original approach from the saliency analysis point of view. While Itti et al., (1998) seek to identify locations that should attract visual attention, one of the purposes of this study is to find the same salient regions in TIR and optical images. Itti's approach handles all three types of features (colour, intensity and orientation) in the same way. The consequence is that if one of the features obtains a highest score, colour for example, the others will be suppressed. Figure 4 demonstrates the example of the scene where two cars are observed. When the colour features are in use as seen in the central image of Figure 4, the red car has the highest score and thus the grey car is not considered to be salient. Our goal, on the other hand, is to have both cars as salient regions. To address this issue, this study suggests the use of the max. and min. intensity across all three colour channels, I_{max} and I_{min} as shown in eq. (4) and (5). I_{max} is responsible for the detection of bright regions, while I_{min} is responsible

for the identification of dark regions. The right image in Figure 4 shows the effect when I_{max} is used: both cars are salient in the same way.

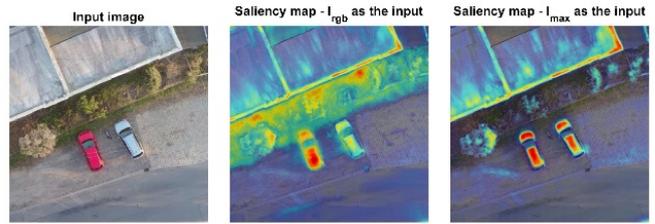


Figure 4 Saliency map comparison of the effect of the different input types. For the meaning of colour, see caption of Figure 2.

$$I_{max}(x, y) = \max_{c \in \{Red, Blue, Green\}} I_{op}(x, y, c) \quad (4)$$

$$I_{min}(x, y) = \overline{\min_{c \in \{Red, Blue, Green\}} I_{op}(x, y, c)} \quad (5)$$

where (x, y) are the image coordinates and $\overline{}$ stands for the complement operator. The complement of an image is computed by subtracting pixel values from the maximum possible value of the image.

3.3. Classification phase one

3.3.1 Introduction to Dempster-Shafer Theory: The detection and localization of thermal anomalies is considered as a classification problem based on saliency maps as input. The classification relies on a fusion of information sources based on the Dempster-Shafer Theory (DST, Dempster, 1967; Shafer, 1976). DST has been introduced as an expansion of probabilistic estimation that can accommodate imprecise and incomplete knowledge as well as data conflicts. An important property of this theory is its capability to consider also unions of classes. We use N mutually exclusive classes C_i with $1 \leq i \leq N$:

$$\Theta = \{C_1, C_2, \dots, C_N\} \quad (6)$$

with Θ the so-called frame of discernment. In the current work, N is equal to four and Θ is described by:

- C_a – class that represents a thermal anomaly candidate.
- C_h – class that represents a hot spot.
- C_c – class that represents a cold spot.
- C_b – class that represents a thermal background.

In DST, the state space is defined as the power set of Θ . This set, usually referred as 2^Θ , contains all subsets of Θ and the empty set \emptyset . In DS theory, a probability mass $m(A)$ is then assigned to every element $A \in 2^\Theta$ by an information source such that $0 \leq m(A) \leq 1$, $m(\emptyset) = 0$ and $\sum_{A \in 2^\Theta} m(A) = 1$. Imprecise information can be handled by assigning a non-zero probability mass to the union of two or more classes. If P data sources are available, the probability mass $m_i(B_i)$ has to be defined for each data source i with $1 \leq i \leq P$ and for all elements $B_i \in 2^\Theta$. The DST allows the fusion of these probability masses from a variety of data sources to compute a combined probability mass for each $A \in 2^\Theta$ according to eq. (7):

$$m(A) = \frac{\sum_{B_1 \cap B_2 \cap \dots \cap B_p = A} (\prod_{1 \leq i \leq p} m_i(B_i))}{1 - \sum_{B_1 \cap B_2 \cap \dots \cap B_p = \emptyset} (\prod_{1 \leq i \leq p} m_i(B_i))} \quad (7)$$

Two parameters, support and plausibility (see eqs. (8) and (9)), can be defined for all $A \in 2^\Theta$. The support for A aggregates all probability masses that directly provide evidence for A , while the plausibility aggregates all probability masses that do not directly provide evidence against A .

$$sp(A) = \sum_{B \in A} m(B) \quad (8)$$

$$pl(A) = \sum_{B \cap A \neq \emptyset} m(B) = 1 - sp(\bar{A}) \quad (9)$$

The classification requires a hard decision for only one class C_i , which implies a mapping from the state space 2^Θ back to Θ . A number of different functions have been defined in the literature for such mapping. The three most prominent mapping functions are the maximum support rule (eq. 10), the maximum plausibility rule (eq. 11) and the maximum mean rule (eq. 12).

$$D = \underset{A \in \Theta}{\operatorname{argmax}} sp(A) \quad (10)$$

$$D = \underset{A \in \Theta}{\operatorname{argmax}} pl(A) \quad (11)$$

$$D = \underset{A \in \Theta}{\operatorname{argmax}} (sp(A) + pl(A)) \quad (12)$$

3.3.2. DST approach used in this work: Three information sources are introduced in this work for the given problem of thermal anomaly detection:

- S_{hot} – an information source that represents hot temperatures, described by a belief function $P_h = SM_{TIR}$, whereas SM_{TIR} is the saliency map of the TIR orthomosaic.
- S_{cold} – an information source that represents cold temperatures, described by a belief function $P_c = SM_{\bar{TIR}}$, whereas $SM_{\bar{TIR}}$ is the saliency map of the complement of the TIR orthomosaic.
- S_o – an information source that represents salient regions from the optical orthomosaic. It is described by a belief function given in eq. (13).

$$P_o = \max(SM_{max}, SM_{min}) \quad (13)$$

where SM_{max} and SM_{min} are:

- SM_{max} – a saliency map that represents bright regions in the optical orthomosaic, that is calculated from the I_{max} .
- SM_{min} – a saliency map that represents dark regions in the optical orthomosaic, that is calculated from the I_{min} .

Table 1 shows all subsets of 2^Θ and their probability masses. The justification for the assignment of the belief functions for each subset of the 2^Θ is as follows:

- High values of P_o are an evidence for a salient region in the optical image. Such a region can be hot (e.g., a manhole that is marked by “2” in Figure 5), cold (e.g., garbage bins that are marked by “4” in Figure 5) or a part of thermal background (e.g., a road marking, that has the same temperature as the surroundings, that is marked by “3” in Figure 5). Therefore, P_o is assigned as the probability of the union of C_h , C_c and C_b . Low values of P_o provide evidence for thermal background or a thermal anomaly. Thus, $(1 - P_o)$ is assigned as the probability of the union of C_a and C_b . As can be seen from the numerator of the combined probability mass of C_b in Table 1, S_o does not provide any evidence for a decision regarding thermal background.
- High values of P_c are an evidence for a cold region (e.g., garbage bins that are marked by “4” in Figure 5), thus P_c is assigned to C_c . Low values of P_c provide evidence for thermal background, a hotspot or a thermal anomaly. Then, $(1 - P_c)$ is assigned as the probability of the union of C_a , C_h and C_b .
- High values of P_h are an evidence for a hot region, that can come from a hotspot (a manhole that is marked by “2” in Figure 5) or a thermal anomaly (heat signature that is marked by “1” in Figure 5). Therefore, P_h is assigned as the probability mass of the union of the C_h and C_a . Low values of the P_h provide evidence for thermal background or a cold region, thus $(1 - P_h)$ is assigned as the probability of the union of the C_c and C_b .

Given the fact that the combined probability masses for all subsets of 2^Θ in Table 1, that include a union of one or more classes from Θ , are equal to zero, support and plausibility of all classes are equal. This means that the three decision rules described by eqs. (10) to (12) yield the same outcome. After per-pixel classification, connected components of the C_a class pixels are generated and are considered as candidate regions for thermal anomaly. Regions with a size of less than 50 pixels are rejected, because we found empirically that they are too small to provide enough details, particularly with regard to their surroundings.

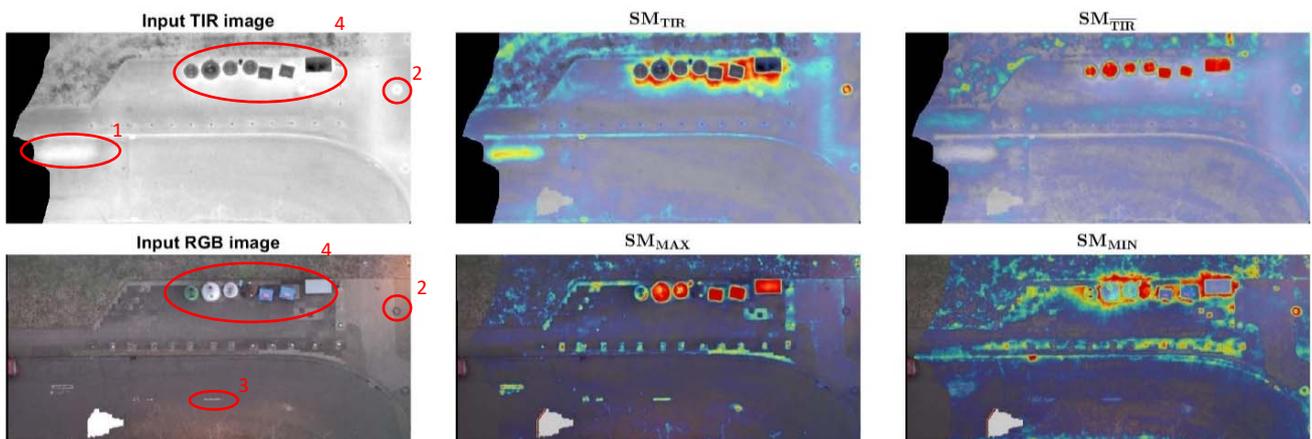


Figure 5 Saliency map examples (for the meaning of the colours, see caption of Figure 2)

Table 1 Probability masses of $A \in 2^{\theta}$

Source Element of 2^{θ}	S_{hot}	S_{cold}	S_o	Combined probability mass $m(A)$
C_a	0	0	0	$\frac{P_h(1-P_c)(1-P_o)}{1-(P_hP_c+(1-P_h)P_c(1-P_o))}$
C_h	0	0	0	$\frac{P_h(1-P_c)P_o}{1-(P_hP_c+(1-P_h)P_c(1-P_o))}$
C_c	0	P_c	0	$\frac{(1-P_h)P_cP_o}{1-(P_hP_c+(1-P_h)P_c(1-P_o))}$
C_b	0	0	0	$\frac{(1-P_h)(1-P_c)}{1-(P_hP_c+(1-P_h)P_c(1-P_o))}$
$C_a \cup C_h$	P_h	0	0	0
$C_a \cup C_c$	0	0	0	0
$C_a \cup C_b$	0	0	$(1-P_o)$	0
$C_h \cup C_c$	0	0	0	0
$C_h \cup C_b$	0	0	0	0
$C_c \cup C_b$	$(1-P_h)$	0	0	0
$C_a \cup C_h \cup C_c$	0	0	0	0
$C_a \cup C_h \cup C_b$	0	$(1-P_c)$	0	0
$C_a \cup C_c \cup C_b$	0	0	0	0
$C_h \cup C_c \cup C_b$	0	0	P_o	0
$C_a \cup C_h \cup C_c \cup C_b$	0	0	0	0

3.4. Classification phase two

While the first classification phase may be seen as a detection and localization stage with the main purpose is not to miss any thermal anomalies, the second classification phase is important to reduce the resulting false alarms among thermal anomaly candidates. We define the problem as a binary classification in order to distinguish only between thermal anomalies and normal temperature variations. In comparison to the first phase, the candidates for thermal anomaly are now viewed as regions rather than single pixels. Each thermal anomaly candidate, which is detected in the classification phase one, is characterised by a set of features that are extracted from different sources of information. The information sources for the second classification phase are the classification results of phase one, the TIR orthomosaic, now with temperature values, and the DSM. The DSM comes into play to provide information on height variations, which may indicate people or cars and thus false alarms, or height discontinuities, e.g. at the outline of a building, where the temperature signal is a mixture of (at least) two surfaces and thus unreliable.

3.4.1 Feature extraction: While modern deep learning methods learn the features to be used for classification, such approaches need a lot of training data, which in our case are not available. Therefore, in this work, features are defined in a more classical way. Region-based features are calculated for each thermal anomaly candidate detected in the previous stage and its surroundings. The surrounding area is calculated by morphological filtering as follows:

$$surr = (obj \otimes se_1) \& \overline{(obj \otimes se_2)} \quad (14)$$

where obj is a binary image of a single thermal anomaly candidate, \otimes is morphological dilation, $\&$ stands for the binary “and” operation and $\overline{}$ stands for binary complement. se_1 and se_2 are circular structuring elements with $r_1 = 1.5 \cdot r_{min}$ and $r_2 = 3 \cdot r_{min}$. r_{min} is the smaller principle axis of the candidate region.

The usage of se_2 ensures that there is no overlap between object and its surroundings. The surroundings can thus be considered as a, possibly deformed, ring. The surroundings are then divided into 8 segments by intersection with 8 rays evenly spaced with respect to their direction, and each segment is represented by its average source value. Examples for the area of the surroundings and the resulting segments are shown in Figure 6 in yellow in the bottom left and right sub-images.

The following group of features is defined so that a thermal anomaly candidate, which is characterized by its average temperature value t_{obj} , stands out against its surroundings, characterized by the vector \mathbf{t}_{surr} of the average temperatures of the surrounding segments. The employed features are $t_{diff-max}$, $t_{diff-min}$ and $t_{diff-dsm}$. $t_{diff-max}$ stand for maximum temperature difference between the area of the thermal anomaly candidate and its surroundings and is described by eq. (15). $t_{diff-min}$ stands for minimum temperature difference between the area and its surroundings and is described by eq. (16). $t_{diff-dsm}$ describes the temperature difference between the thermal anomaly candidate and the surrounding segment, for which the DSM value ($dsm_{surr}(j)$) is closest to the average height of the thermal anomaly candidate area (dsm_{obj}). The formulation of this feature is shown in eq. (17). This feature encourages the fact that the difference in temperature should be measured on the same height as far as possible, as it makes little sense to take into account the difference in temperature between, e.g., the ground surface and the roof of a building.

$$t_{diff-max} = t_{obj} - \mathbf{t}_{surr}(i), \text{ where } \quad (15)$$

$$i = \underset{j \in [0,1,\dots,7]}{\operatorname{argmin}}(\mathbf{t}_{surr}(j))$$

$$t_{diff-min} = t_{obj} - \mathbf{t}_{surr}(i), \text{ where } \quad (16)$$

$$i = \underset{j \in [0,1,\dots,7]}{\operatorname{argmax}}(\mathbf{t}_{surr}(j))$$

$$t_{diff-dsm} = t_{obj} - t_{surr}(i), \text{ where} \quad (17)$$

$$i = \operatorname{argmin}_{j \in [0,1,\dots,7]} |dsm_{surr}(j) - dsm_{obj}|$$

As discussed before (see left image in Figure 2), one of the possibilities that triggers false alarms is a cold body of class C_c next to a surface of average temperature (thus hotter than the cold body). This surface will then be salient and will be considered a thermal anomaly candidate. To identify such false alarms we use the distance between the thermal anomaly candidate and the nearest cold object. First, for each pixel of the candidate, we compute the distance x to the nearest pixel belonging to a cold object C_c . In order to normalise the distances we then apply the sigmoid function (see eq. 18) to all x . In the sigmoid function, α is the slope and b represents the value of x where the function has a value of 0.5. In order to give smaller x values a higher weight, we choose a value $a < 0$. Finally, the feature ($d_{cold-obj}$), that reflects the proximity of the thermal anomaly candidate to a cold region, is computed as the average of the values $f(x)$ for all pixels belonging to the candidate.

$$f(x) = \frac{1}{1 + e^{-a(x-b)}} \quad (18)$$

Ideally, a thermal anomaly is surrounded only by a uniform temperature environment, which should be defined by DST as thermal background C_b . However, there are instances where other classes occur in the surroundings of the thermal anomaly candidate. The pixel-wise class distribution (h_{class}) of the surroundings of the candidate region, represented as the normalized histogram, provides such an information and is therefore defined as yet another feature (note, that here, the surrounding area is viewed as a whole entity and not as a set of segments).

The last source of information used to extract features is the DSM. As was discussed above, temperature differences between areas belonging to different heights can trigger false alarms. Therefore, we require thermal anomaly candidates to lie on a plane. The main aim of the next feature is to characterize thermal anomaly candidates detected by DST, in terms of height variation of the candidate area and its surroundings. To describe these variations, we compute point, line and area regions from the DSM according to Förstner (1994). Then, the number of pixels belonging to each region type in the candidate area and its surroundings, respectively, which come in the form of normalised histograms $h_{dsm-obj}$ and $h_{dsm-surr}$, is considered as the feature to be used in the classification. Figure 6 gives an example of false alarm (region marked with red in the bottom left image), which is caused by a difference in temperature between the ground surface and the building roof. In such case, $h_{dsm-obj}$ ideally includes entries only for area regions (see the bottom right image in Figure 6), while in comparison, $h_{dsm-surr}$ also has entries for line regions.

Altogether, each thermal anomaly candidate is described by a feature vector. Table 2 summarize the structure of the feature vector.

Table 2 Feature vector

Feature	Size	Feature	Size
$t_{diff-max}$	1	$h_{class-surr}$	4
$t_{diff-min}$	1	$h_{dsm-obj}$	3
$t_{diff-dsm}$	1	$h_{dsm-surr}$	3
$d_{cold-obj}$	1		

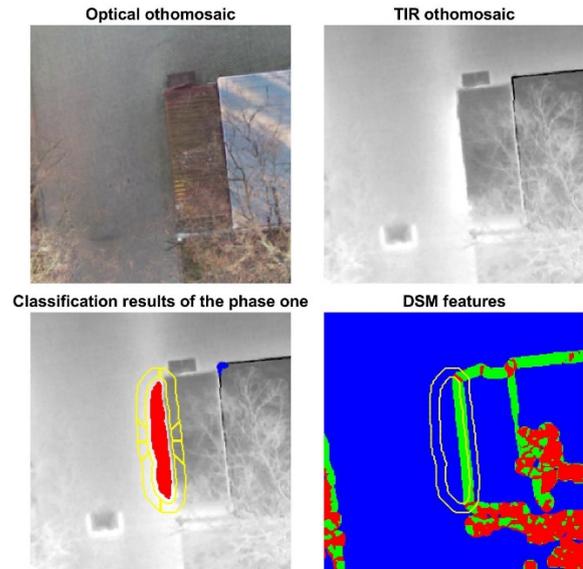


Figure 6: Feature selection - example of a false alarm: thermal anomaly candidate close to a building. The bottom right image shows the description of height variations: area regions are marked in blue, line regions in green and point regions in red.

3.4.2 Classification: The supervised classification itself is formulated as a binary problem in order to differentiate between thermal anomalies and normal temperature variations. We use the standard Random Forest (RF) classifier, as it has achieved promising results in preliminary tests. The dataset is formed by all thermal anomaly candidates from the phase one classification. For each of these candidates, a label has been assigned manually, showing whether it is a thermal anomaly or it may be categorized as a normal temperature variation.

4. EXPERIMENTS AND RESULTS

4.1. Objectives

The main goal of the experimental part is, of course, to confirm the main hypothesis on which this study is based. In addition, it is important to understand the effect of the saliency model parameters and their relationship to the TIR and the optical images, the influence of the different training methods for the supervised learning of the second phase, and the influence of the temperature of the thermal anomaly candidates on the results of the classification.

The criteria to measure the success of the experiments obviously depend on the application. In DHS monitoring, the cost of missing an existing pipe leakage is very high. On the other hand, the cost of false positives, the time spent by a human operator to validate the origin of an observed thermal anomaly, could be considered to be low.

The proposed method is based on two classification phases, and the success criteria for each one are slightly different. In the first phase, ideally all real thermal anomalies should be detected and a higher number of false positive alarms is acceptable. The criteria for the second classification phase is to reduce these false positive alarms as much as possible, while ideally keeping all real anomalies.

4.2. Data and processing strategy

The data was acquired by a DJI M200 UAV, a vertical take-off and landing (VTOL) quadcopter fitted with a GNSS receiver, an inertial measurement unit (IMU) and a barometer. The camera

system in use was the XT2 DJI Zenmuse. It is a gimbal-stabilized camera setup, which rigidly combines a dual-sensor design of FLIR radiometric thermal imager and a CMOS optical camera.

A total of 11 flights have been carried out in various regions of Hannover, Germany. Flight planning was based on the TIR camera because of a lower GSD. Based on a pixel size of $17 \mu\text{m}$ and a focal length of 13 mm, the flight height was set to 40 meters above ground, resulting in a GSD of 5.2 cm for the TIR images (the corresponding GSD of the optical camera was 0.95 cm).

Rigorous photogrammetric processing was carried out using Agisoft Metashape² including GNSS measurements for the positions of the image projection centres. A detailed explanation of the photogrammetric processing and the results can be found at Sledz et al. (2020).

As noted by Sledz et al. (2020), this particular dataset lacks real water leakages. On the other hand, the working hypothesis is true not only of DHS-related leaks. It was therefore decided to concentrate on the detection of all available thermal anomalies in the current dataset. As the consequence, in this work no GIS data were used to limit the search space for thermal anomalies around the pipe locations, as was done by Sledz et al. (2020).

The Saliency Model Implementation Library for Experimental Research (SMILER) by Wloka et al. (2018) is a software package, which provides an open, standardized, and extensible framework for computational saliency models. This software package was used in all relevant parts of this study to compute the saliency map and served a starting point for the required modifications.

4.3. Results

4.3.1 Saliency analysis and classification phase one: The size of the region detected by the saliency analysis depends on the "centre-surround" difference step in the saliency model. Itti et al. (1998a) use central scales of $c \in \{2, 3, 4\}$, and a surrounding scale $s = c + \delta$, with $\delta \in \{3, 4\}$. As Zhong et al. (2019) also pointed out, the values for c and s influence the size of detectable regions. By selecting a larger value for c , larger objects become salient, and small objects are not salient anymore. The following configurations have been used in the current study:

- For the TIR images, which come with a resolution of 512×640 pixels, saliency scales are selected as $c \in \{1, 2, 3, 4\}$ and $\delta \in \{3, 4\}$.
- For the optical images, which come with a resolution of 3000×4000 pixels, the selected values $c \in \{3, 4, 5, 6\}$ and $\delta \in \{3, 4\}$.

The choice of the larger values of c during saliency map calculation for optical images is based on the fact, that the optical camera has higher GSD. The need to align the TIR and optical representations in object space entails such the selection.

After a series of experiments, it was concluded that the best way to calculate a saliency map of an orthomosaic is to use orthoprojection of saliency maps of single images. Such an approach ensures that a region with smaller temperature changes do not compete with a region from a different section of the entire scene, thus preserving local saliency, i.e. saliency per input image.

Figure 7 and Figure 8 show examples of the phase one classification results. The figures are constructed in a following manner: to the left is part of an optical orthomosaic, the corresponding of the TIR orthomosaic is shown in the centre, classification results are shown as an overlay on the TIR orthomosaic to the right. The

classification results are indicated by the following colours: thermal anomaly candidates C_a are shown in red, C_a smaller than 50 pixels in magenta, cold spots C_c in blue, hot spots C_h in yellow; thermal background C_b is not shown in colour. It can be seen, that the thermal anomalies (marked by red dotted circles), as defined by the working hypothesis, are indeed identified correctly. Cold entities such as the garbage bins in Figure 7 or the car roofs in Figure 8 are also identified correctly as C_c . Hot items such as the manhole in Figure 7 or the car's engine hoods in Figure 8 are classified correctly as C_h . Areas without major temperature fluctuations are classified as thermal background C_b .

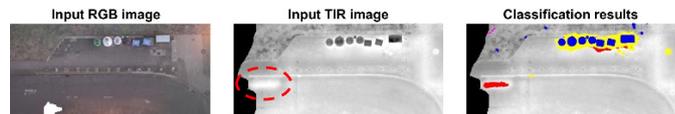


Figure 7: Phase one classification results – scene 1

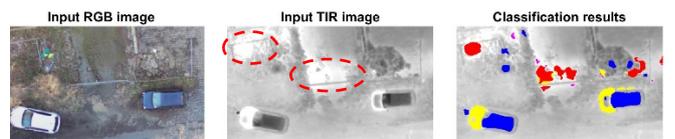


Figure 8: Phase one classification results – scene 2

Classification evaluation of DST was performed manually: each thermal anomaly candidate was visually inspected using the TIR and the optical image. It should be noted that as we are only interested in thermal anomalies, only the results for C_a were validated, while the remaining parts of the images were checked for non-detected thermal anomalies. Overall, the findings are satisfactory, from the 1390 detected thermal anomaly candidates 59 out of 60 true thermal anomalies were found, resulting in a recall of 98% and a precision of 4% for the thermal anomaly class. The low precision follows our expectation and underlines the need of the second classification phase in order to eliminate the false alarms.

In spite of this positive outcome, the suggested DST classification has its own limitations:

- The change in intensity of the optical image induced by shadows may be considered as salient, if it has appropriate size. A thermal anomaly underneath this region is then incorrectly classified as a hot spot (C_h). Figure 9 shows an example of a bad roof insulation that creates a heat escape and at the same time part of the roof looks much brighter than the surroundings due the partial shadow in the optical image, which is considered salient in the optical image. The consequence is that the true thermal anomaly is incorrectly classified as a hot spot (marked as yellow) and not as thermal anomaly.
- The suggested technique relies on high-quality optical images in terms of contrast, so that objects on the surface stand out as salient regions. However, this is not always the case. Figure 10 shows an example of a manhole, which is visible in the TIR image, but not in the corresponding optical one due to low illumination. As a result, this area is only salient in the TIR image and thus classified incorrectly as thermal anomaly, thus it is a false alarm.

² <https://www.agisoft.com/>

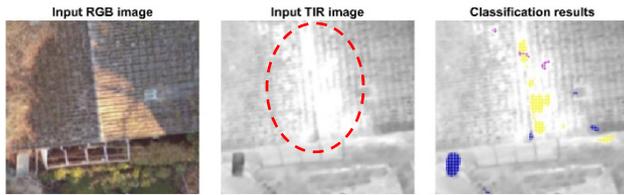


Figure 9: False classification by DST caused by the shadows.

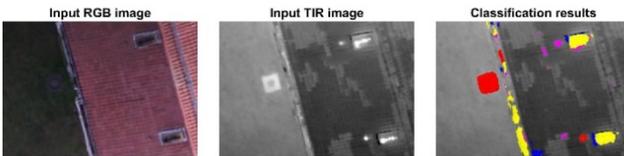


Figure 10: False classification by DST as a consequence of low contrast in the optical image

4.3.2 Classification phase two: The thermal anomaly candidates, which were detected by DST, form the dataset for the second classification phase. This phase distinguishes between thermal anomalies, which are defined as class 1, and the rest of the cases, which are defined as class 0. Table 4 depicts the number of cases for the two classes depending on the temperature difference between the region and the surroundings, based on the feature $t_{diff-min}$; it can be seen that the two classes occur with significantly different frequency. The RF parameters are chosen by random search in parameter space: number of trees with a span of [5,250] and step of five, depth of a tree with a span of [10,220] and step of five, minimum number of samples required for a leaf node with a span of [2,25] and step of two, number of features to consider when looking for the best split with a span of [2,25] and step of two. The evaluation was carried out by k-fold cross validation, with five folds. One of the benefits of RF is that the probability of a classified entity (a region is our case) to belong to each of the classes can be (indirectly) estimated. In the case of a binary classification, the selection of the resulting class is made by selecting the one with the highest probability. In other words, a hard threshold of 0.5 is being used. However, a better outcome may be obtained by adjusting this threshold. Every classification result for the test portion was fine-tuned by adjusting the probability threshold of class 1 (anomaly class) to the point where False Positive Rate (FPR) and False Negative Rate (FNR) are equal. Despite what was mentioned earlier for the case of DHS, the intersection between FPR and FNR is chosen as the optimal point here due to the lack of knowledge of the actual costs of each misclassification.

In order to investigate the influence of the unbalanced dataset, two separate experiments were conducted, as can be seen in Table 5. Class 0 was under-sampled in both experiments: 20% was used for training and 80% for testing (instead of the opposite, which is considered as no under-sampling). In contrast to class 0, class 1 was treated differently: in experiment #1 80% was used

for training and 20% for testing, while in experiment #2 for training 80% was selected and then over-sampled by the Synthetic Minority Oversampling Technique (SMOTE, Chawla et al., 2002) to reach the number of the class 0 training samples. The remaining 20 percent of class 1 were used as the test portion in both experiments.

In order to determine the effect of the thermal anomaly temperature threshold on the ability to yield correct results, different datasets were prepared with regard to the feature ($t_{diff-min}$), as shown in Table 4. These thresholds were chosen on the basis of the particular dataset at hand, whereas they could differ in other scenarios. As can be seen this dataset is highly unbalanced regardless of the difference used.

Table 4 Classes distribution in the datasets

Anomaly threshold	Non-Anomaly (class 0)	Anomaly (class 1)
$t_{diff-min} > 0^{\circ}\text{C}$	1331	59
$t_{diff-min} > 0.5^{\circ}\text{C}$	1332	58
$t_{diff-min} > 1^{\circ}\text{C}$	1339	51
$t_{diff-min} > 1.5^{\circ}\text{C}$	1351	39

Table 5 Cross validation configuration of the two experiments

		Experiment #1	Experiment #2
Train	Class 0	20%	20%
	Class 1	80%	SMOTE(80%)
Test	Class 0	80%	80%
	Class 1	20%	20%

Table 3 presents RF classification results of experiment #1 and #2. The results are shown as an average of the 5-fold cross validation. It can be seen that with higher anomaly threshold the true positive rate (TPR) of the classification reaches 100%. It should also be noted that for thermal anomalies with $t_{diff-min} > 0.5^{\circ}\text{C}$, TPR stands at least at 92 percent. FPR is also closely associated with $t_{diff-min}$ in a way that it is decreased with a higher threshold. From the point of view of TPR, SMOTE only has a marginal advantage, visible with higher thermal anomaly threshold only.

Table 3 Cross validation classification results

Anomaly threshold	Experiment #1			Experiment #2		
		<i>Predicted 0</i>	<i>Predicted 1</i>		<i>Predicted 0</i>	<i>Predicted 1</i>
$t_{diff-min} > 0^{\circ}\text{C}$	<i>Actual 0</i>	TNR = 89.5%	FPR = 10.5%	<i>Actual 0</i>	TNR = 88.3%	FPR = 11.7%
	<i>Actual 1</i>	FNR = 15.6%	TPR = 84.4%	<i>Actual 1</i>	FNR = 13.3%	TPR = 86.7%
$t_{diff-min} > 0.5^{\circ}\text{C}$	<i>Actual 0</i>	TNR = 94.7%	FPR = 5.3%	<i>Actual 0</i>	TNR = 96.5%	FPR = 3.5%
	<i>Actual 1</i>	FNR = 4.0%	TPR = 96.0%	<i>Actual 1</i>	FNR = 8.0%	TPR = 92.0%
$t_{diff-min} > 1.0^{\circ}\text{C}$	<i>Actual 0</i>	TNR = 98.6%	FPR = 1.4%	<i>Actual 0</i>	TNR = 99.0%	FPR = 1.0%
	<i>Actual 1</i>	FNR = 5.0%	TPR = 95.0%	<i>Actual 1</i>	FNR = 5.0%	TPR = 95.0%
$t_{diff-min} > 1.5^{\circ}\text{C}$	<i>Actual 0</i>	TNR = 95.4%	FPR = 4.6%	<i>Actual 0</i>	TNR = 98.7%	FPR = 1.3%
	<i>Actual 1</i>	FNR = 0.0%	TPR = 100.0%	<i>Actual 1</i>	FNR = 0.0%	TPR = 100.0%

5. CONCLUSION AND DISCUSSION

This paper suggests a two-phase classification approach for thermal anomaly detection from a combination of thermal, optical and height data. In the core of this work stands the hypothesis that thermal anomalies occur as a salient region only in the TIR image and not in the optical image. In general, thermal anomalies are typically rare events. As the consequence, there is an issue with the high false alarm rate, which we tackle successfully.

The first classification phase is based on DST classification, where the belief functions come in the form of saliency maps of the TIR and the optical image data. It was shown that a modification of the saliency model proposed by Itti et al. (1998) provides an effective tool for hot spot detection. In the second classification phase false alarms are reduced: anomaly candidates from the first phase and their surroundings are viewed as independent entities. Measured temperature values, height information derived from a DSM and the classification results of the DST step are used in a binary Random Forest classification to distinguish thermal anomalies from normal temperature variations.

It is shown that saliency maps of the TIR and optical images, when used as belief functions in DST, can provide a valuable classification tool. Nevertheless, despite the good results of the DST classification, the proposed approach has its own limitations outlined in 4.3.1. In order to address these limitations, the optical camera should be of high quality in terms of contrast, as it may need to be operated in low illumination conditions.

In future work, we will investigate the use of Convolutional Neural Nets (CNNs), which have been shown to be a very effective tool for a variety of classification tasks, but require vast amounts of training data. Also, semantic segmentation of optical images, which can be used to help identify objects such as cars, chimneys, manholes, and other structures will be investigated. Afterwards the heat signature of these objects can be confirmed by thermal data. Following the current working hypothesis, heat signatures that are not assigned to any observed object can then be considered a thermal anomaly.

ACKNOWLEDGMENT

The work is supported by the Arbeitsgemeinschaft industrieller Forschungsvereinigungen (AiF) under IGF-grant no. 19768 N. This support is gratefully acknowledged. The authors would like to thank our partners Fernwärme-Forschungsinstitut (FFI) GmbH, Hemmingen (Germany) and Enercity AG, Hannover (Germany), and in particular Volker Herbst (FFI) and Werner Manthey (Enercity) for their strong support.

REFERENCES

Axelsson, S., 1988. Thermal modelling for the estimation of energy losses from municipal heating networks using infrared thermography, *IEEE Transactions on Geoscience and Remote Sensing*, vol. 26, no. 5, pp. 686–692.

Berg, A., Ahlberg, J., 2014. Classification of leakage detections acquired by airborne thermography of district heating networks. 8th IAPR Workshop on Pattern Recognition in Remote Sensing, 10.1109/PRRS.2014.6914288.

Borji, A., 2012. Exploiting local and global patch rarities for saliency detection. In: *Computer Vision and Pattern Recognition*, pp. 478–485.

Breiman, L., 2001. Random Forests, *Machine Learning* 45, 5–32. <https://doi.org/10.1023/A:1010933404324>.

Bruce, N.D.B., Tsotsos, J.K., 2005. Saliency based on information maximization. In: *Neural Information Processing Systems*, pp. 155–162.

Chawla, N., Bowyer, K., Hall, L., Kegelmeyer, W.P., 2002. SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*, 16, 321–357. 10.1613/jair.953.

Dempster, A.P., 1968. A generalization of Bayesian inference. *Journal of the Royal Statistical Society. Series B (Methodological)* 30, pp. 205–247.

Förstner, W., 1994. A framework for low level feature extraction, in: J.O. Eklundh (Ed.), *Computer Vision—ECCV '94*, vol. II, 5th ICCV, Boston, MA, pp. 383–394.

Friman, O., Follo, P., Ahlberg, J., Sjökvist, S., 2014. Methods for large-scale monitoring of district heating systems using airborne thermography, *IEEE Trans. Geo-science and Remote Sensing*, 52(8):5175–5182.

Harel, J., Koch, C., Perona, P., 2006. Graph-based visual saliency. in *Advances in Neural Information Processing Systems 19: Proceedings*, MIT Press, 2007, pp.545–552.

Itti L., Koch C., Niebur E., 1998. A model of saliency-based visual attention for rapid scene analysis, in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 11, pp. 1254–1259.

Ljungberg, S-A., Rosengren, M., 1988. Aerial and mobile thermography to assess damages and energy losses from buildings and district heating networks - operational advantages and limitations. In: *Int. Arch. Photogramm. Remote Sens.*, XXVII-B7, pp. 348–359.

Qi, S., Ma, J., Tao, C., Yang, C., Tian, J., 2013. A Robust Directional Saliency-Based Method for Infrared Small-Target Detection Under Various Complex Backgrounds, in *IEEE Geoscience and Remote Sensing Letters*, vol. 10, no. 3, pp. 495–499, doi: 10.1109/LGRS.2012.2211094.

Rottensteiner, F., Trinder, J., Clode, S., Kubik, K., 2005. Using the Dempster-Shafer method for the fusion of LIDAR data and multi-spectral images for building detection. *Information Fusion*. 6. 283–300. 10.1016/j.inffus.2004.06.004.

Shafer, G. 1976. *A mathematical theory of evidence*. Princeton University Press.

Sledz A., Unger J., Heipke C., 2018. Thermal IR imaging: image quality and orthophoto generation. In: *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLII-1, 413–420, <https://doi.org/10.5194/isprs-archives-XLII-1-413-2018>.

Sledz A., Unger J., Heipke C., 2020. UAV-based thermal anomaly detection for distributed heating networks. In: *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLIII- B1-2020, 499–505, <https://doi.org/10.5194/isprs-archives-XLIII-B1-2020-499-2020>.

Wloka., Kunić., Kotseruba., Fahimi., Frosst., Bruce., Tsotsos., 2018. SMILER: Saliency Model Implementation Library for Experimental Research. arXiv preprint arXiv:1812.08848

Zhong, Y., Xu, Y., Wang, X., Jia, T., Xia, G., Ma, A., Zhang, L., 2019. Pipeline leakage detection for district heating systems using multisource data in mid- and high-latitude regions. *ISPRS Journal of Photogrammetry and Remote Sensing*. (151) 207–222. 10.1016/j.isprsjprs.2019.02.021.