Enhanced analysis of thermographic images for monitoring of district heat pipe networks

Amanda Berg, Jörgen Ahlberg and Michael Felsberg

The self-archived postprint version of this journal article is available at Linköping University Institutional Repository (DiVA):

<http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-133004>

N.B.: When citing this work, cite the original publication.

Berg, A., Ahlberg, J., Felsberg, M., (2016), Enhanced analysis of thermographic images for monitoring of district heat pipe networks, *Pattern Recognition Letters*, 83(2), 215-223.

<https://doi.org/10.1016/j.patrec.2016.07.002>

Original publication available at:

<https://doi.org/10.1016/j.patrec.2016.07.002>

Copyright: Elsevier

<http://www.elsevier.com/>

Research Highlights (Required)

To create your highlights, please type the highlights against each `\item` command.

It should be short collection of bullet points that convey the core findings of the article. It should include 3 to 5 bullet points (maximum 85 characters, including spaces, per bullet point.)

- False alarm reduction of automatically detected leakages in district heating networks
- We propose a method for characterization of leakages over time
- We provide extensive experimental analysis on real-world data
- The described system can be used to plan for long-term maintenance
- We propose a visualization scheme exploiting the consecutive data collections

Enhanced analysis of thermographic images for monitoring of district heat pipe networks

Amanda Berg^{a,b,**}, Jörgen Ahlberg^{a,b}, Michael Felsberg^a

^aComputer Vision Laboratory, Department of Electrical Engineering, Linköping University, SE-581 83 Linköping, Sweden

^bTermisk Systemteknik AB, Diskettgatan 11 B, SE-583 35 Linköping, Sweden

ABSTRACT

We address two problems related to large-scale aerial monitoring of district heating networks. First, we propose a classification scheme to reduce the number of false alarms among automatically detected leakages in district heating networks. The leakages are detected in images captured by an airborne thermal camera, and each detection corresponds to an image region with abnormally high temperature. This approach yields a significant number of false positives, and we propose to reduce this number in two steps; by (a) using a building segmentation scheme in order to remove detections on buildings, and (b) to use a machine learning approach to classify the remaining detections as true or false leakages. We provide extensive experimental analysis on real-world data, showing that this post-processing step significantly improves the usefulness of the system. Second, we propose a method for characterization of leakages over time, *i.e.*, repeating the image acquisition one or a few years later and indicate areas that suffer from an increased energy loss. We address the problem of finding trends in the degradation of pipe networks in order to plan for long-term maintenance, and propose a visualization scheme exploiting the consecutive data collections.

© 2018 Elsevier Ltd. All rights reserved.

1. Introduction

District heating networks distribute heat through underground pipes carrying media, *i.e.*, hot water or steam, from a central power plant. Heat leakages due to damaged insulation or media leakages due to cracks are common problems. The pipes degenerate with time (Olsson (2001)) and in some cities the pipes have been used for several decades. Loss of media (water/steam) or energy is expensive and has negative impact on the environment (Fröling (2002)). It is therefore of great interest to the network owners to find methods to detect and localize the leakages reliably. The fact that the pipes are placed underground increases the need for correct localization. Moreover, major leakages (in the order of 50–150 m³ of media per day) may cause the ground to collapse due to erosion, whereby large amounts of water at boiling temperature are exposed.

Potential district heat pipe leakages can be detected by analysing imagery captured by airborne thermography. While

this method has been very successful in recent years, it suffers from large numbers of false detections, since there are several types of objects and phenomena that are likely to be detected as well. Examples are areas that, for some reason, are warmer than their surroundings, such as, chimneys, cars, and heat leakages from buildings. In a large city, there might be several thousands of false detections. Another problem is that thermography gives a snapshot of the network's status, from which it is difficult to see the trends in network degradation that is needed for long-term maintenance planning.

In this paper, we present a method to reduce the number of false detections while maintaining the true positive rate at a fixed level. In order to achieve this, we follow a two-step classification procedure:

- Extract building locations from publically available geographic information, and remove all detections located on buildings.
- Extract image features and use a machine learning method to classify detections as true (media/energy) or false detections.

Moreover, we propose a novel method for temporal charac-

^{**}Corresponding author

e-mail: amanda.berg@liu.se (Amanda Berg),
jorgen.ahlberg@liu.se (Jörgen Ahlberg), michael.felsberg@liu.se
(Michael Felsberg)

terization and visualization of the energy loss of the network. Long-term degradation of a pipe can not be detected as a single leakage. Instead, we analyze larger areas and compare the radiated energy at two different times, separated by one or a few years. The area covering the district heating network is divided into square cells and the comparison of energy loss is done for each cell individually.

1.1. Related work

Various methods for monitoring of district heating networks have been developed over the years, for example methods based on frequency response or change in electrical impedance for a thread installed inside the pipe insulation. It is also common to measure the flow of water or steam in the inlet and outlet. If it differs, there is a leakage somewhere along the pipe. Such methods can be used to detect the presence of a leakage and sometimes its approximate location (which pipe segment). However, the exact location, needed for digging, is not revealed.

Methods for large-scale monitoring by aerial thermography, that is remote sensing from an aircraft using a thermal camera, was investigated already in the 80's by Ljungberg (1987) and Axelsson (1988). The results are somewhat antiquated due to the drastic development of thermal cameras during the last two decades. Also, ground-based thermography using hand-held cameras has been investigated by Bohm and Borgström (1996) and Zinko et al. (1996). Compared to aerial thermography, this has several drawbacks, such as restricted access to many areas of interest and less scaleability.

The first system with automatic image analysis was presented by Friman et al. (2013). The system uses anomaly detection in order to detect abnormally warm areas along the pipes. However, the problem is the large number of false alarms since there are many areas that, for one reason or another, are warmer than the surroundings. To reduce the number of false alarms, buildings are segmented in order to avoid detections due to, e.g., chimneys when the pipes pass under buildings.

Temporal characterization of remote sensing data can be regarded as a form of change detection, which has been extensively studied. Applications include various kinds of environmental monitoring (e.g., land use and land cover (LULC) change, deforestation and crop monitoring; see D. Lu and Moran (2003) for a review of such applications), urban change (Li et al. (2012)), and military target detection (Meola et al. (2011); Dekker et al. (2013)). The employed methods usually assume multispectral, sometimes even hyperspectral data, or SAR data (Li et al. (2012)). Methods vary greatly, depending of the type of change to be detected, and they can be pixel-based (Meola et al. (2011); Theiler (2008); Blaschke (2010)) or object-based (Blaschke (2010)).

1.2. Contribution

The main contribution of the present work is to *characterize* detections obtained using the method presented by Friman et al. (2013), and then to *classify* them as real leakages or false detections. The two subproblems of feature selection and choice of classification methods are both addressed. Second, we propose to do the *building segmentation* differently compared to Friman

et al. (2013). Third, we propose a method for *temporal characterization* and *visualization* of district heating network energy loss.

1.3. Outline

The outline of this paper is as follows. In Section II, the acquisition of data and the resulting data sets are described. Section III describes our method for false alarm reduction and how it adds on to existing methods. Section IV describes the proposed method for temporal analysis. Experiment and results are described in Section V, and, finally, Section VI contains our conclusions.

2. Data acquisition and leakage detection

This section briefly describes the data acquisition process, pre-processing of the thermal images and the employed detection method. This corresponds to the scheme described in Friman et al. (2013), but is included here in order to make the paper self-contained.

2.1. Image data

The thermal images are acquired from an aircraft. GPS and IMU are used to record the position and orientation of the aircraft in order to facilitate georeferencing. The imagery is georeferenced using semi-automatic commercial off-the-self software and stored in GeoTIFF format.

The thermal camera is a cooled mid-wave infrared FLIR SC7000 Titanium with a resolution of 640×512 pixels and a field of view of 11°. At an altitude of 800 m, this yields a pixel footprint of 25×25 cm. The mid-wave camera was chosen because of its fast, cooled detector that allows to capture images without motion blur at the present ground speed.

In order to minimize the number of false detections, data collection is mainly done during the night or at dawn during spring or autumn. At this time, neither vegetation or snow is blocking the view, the effect from sun heating is minimal, and the streets are not covered with cars blocking the view (Sjörkvist et al. (2012)).

The number of night flights required in order to cover the whole area depends on the size of the area. For a medium-sized Swedish city (150.000 inhabitants), about three flights are needed.

2.2. GIS data

The network owner provides pipe location information in the form of vector maps. This information is projected on top of the georectified images creating a rasterized pipe mask for each image, Fig. 1. The mask is then used to limit the search for unexpectedly high temperatures to areas where pipes are buried.

Fig. 1. An example of a georectified thermal image and a pixel mask of the district heating pipes obtained by rasterizing a GIS layer of the district heating system.

2.3. Detections

A detection is in this context an area with a certain shape and location pointed out as abnormally warm. That is, it is an extended object, not just a map coordinate. In order to extract the detections from the images, we use the anomaly detection method from Friman *et al.* (2013). Statistics of the ground temperature inside the pipe mask are calculated from all images captured during one flight and the most deviating pixels above certain thresholds in the high end of the distribution (*i.e.* the "warmest pixels") are marked as detections. The percentage thresholds are; 0.05%, 0.1%, 0.5%, 1%, 3% and 5%, resulting in six different layers of detections.

2.4. Ground truth data

Acquisition campaigns during the last couple of years have resulted in thousands of thermal images from 17 Scandinavian towns and cities. Three of the most recent acquisitions were selected for this study. The selection was based on the fact that the customers for these flights could provide ground truth, *i.e.*, information about which detections had been investigated further and proven to be real (or false) leakages.

Detections from these cities have been manually labelled as media leakages, energy leakages, or false detections. The compilation of manual annotations was realized through interviews with the network owners. All interviews were conducted several months after delivery of the detection results, *i.e.*, when the true status of the pipes in many cases have been investigated (by digging). Also, many detections can be pointed out as false by the network owners or by ourselves.

In total, we have 1585 labeled detections, of which 1173 are false. The number of samples for each class and layer can be seen in Table 1.

3. False alarm reduction

Our approach is illustrated in Fig. 2. The green boxes are identical to the scheme by Friman *et al.*, while the blue ones are added or modified. Below we will describe the proposed building segmentation scheme, and then the feature extraction and classification.

Fig. 3. An example of a building mask (right) generated from the colour raster map OpenStreetMap (left).

3.1. Removing false detections using building segmentation

A common source for false alarms are detections of objects at rooftops with unnaturally high temperatures, *e.g.*, chimneys and atriums. These false alarms appear because the pipes sometimes pass beneath buildings. Since we know that real leakages of the district heating network never can appear at rooftops, information on building locations can be used to remove false detections. Friman *et al.* implemented a building segmentation scheme based on the watershed transform and AdaBoost classification in order to automatically extract building information from thermal imagery.

The proposed building segmentation method is based on OpenStreetMap¹. A binary building mask is generated using a simple color segmentation scheme, since buildings in these images have a specific set of colors; an example can be seen in Fig. 3. By thresholding the three RGB channels, each pixel is classified as belonging to a building or not.

OpenStreetMap-images as well as the thermal images are stored in GeoTIFF format with world coordinate information related to each pixel, information that is used for image registration.

3.2. Removing false detections using a classifier

The basic assumption for our approach is that distinguishing features of the different types of detections do exist. Such features could be extracted from the imagery or from the detections themselves (shape descriptors). The labelled examples should be studied to find an initial set of discriminating features which is later refined. The extracted features are then used to evaluate a number of classifiers. Below, we describe the selection of features and classifiers to use in the final system.

3.2.1. Features

Features were found by studying the labeled samples. The initial selection consisted of 18 scalar features based on (thermal) intensity distribution within the detection, shape and propagation of the detection, and proximity information. All initial features are listed and described in Table 2.

Feature selection was done by calculating the Mahalanobis distance, d , between class means, μ_1, μ_2 , for each feature.

$$d = \sqrt{(\mu_1 - \mu_2)^T S^{-1} (\mu_1 - \mu_2)} \quad (1)$$

¹<http://openstreetmap.org>

Table 1. Number of ground truth samples for each layer and class. The concept of layers and percentage thresholds is explained in Sec. 2.3.

Layer no.	1	2	3	4	5	6
Threshold	0.05%	0.1%	0.5%	1%	3%	5%
True (Media/Energy)	34	39	71	89	99	80
False	71	75	148	237	294	348

Fig. 2. The proposed approach. Green boxes are identical to the scheme by Friman *et al.*, while the blue ones are added or modified.

where S is the covariance matrix of the second distribution. Assuming Gaussian distribution and conditional independence among samples, the lowest possible error rate for a given class can be calculated with the Bayes error rate (Duda et al. (2001)):

$$P_e = 1 - \sum_{i=1}^c \int_{R_i} p(x|\omega_i)P(\omega_i)dx \quad (2)$$

P_e is the probability of error, R_i are the regions which the space has been divided into by the classification boundary, x is a sample and ω_i are the true labels of the samples. (2) can be rewritten as (Duda et al. (2001)):

$$P_e = \frac{1}{\sqrt{2\pi}} \int_{d/2}^{\infty} e^{-u^2/2} du \quad (3)$$

where d is the Mahalanobis distance. As the Mahalanobis distance increases and approaches infinity, the probability of error decreases. Therefore, the Mahalanobis distance is a common choice for measuring the "goodness" of a feature (Jain and Zongker (1997)).

3.2.2. Classifiers

Two linear and three nonlinear classifiers were chosen for evaluation. The linear ones are Linear Discriminant Analysis (LDA) (Duda et al. (2001)) and Linear Support Vector Machines (Hastie et al. (2008)). These were mainly included in the experiments to exclude linear methods from further evaluation. Three different nonlinear classifiers were evaluated; the Radial Basis Function Support Vector Machine (Hastie et al. (2008)), AdaBoost (Bishop (2006)), and Random Forest (Breiman (2001)). We used the implementations from PRTtoolbox (Duin et al. (2007)).

3.3. Evaluation and selection methodology

There are three main types of detections; media leakages, energy leakages and false detections. Since both media and energy leakages are interesting for the network owner, we chose to

Fig. 4. A part of the grid covering the area. The detections from the old acquisition campaign are marked with red, dashed lines and the new detections are green and filled.

combine the media and energy samples into one class, hereafter called true detections. Furthermore, incorrect classification of one of these classes as the other one is not as critical as incorrect classification between media/false or energy/false.

As mentioned before, the cost for classifying a true detection as a false one is much higher than classifying a false detection as a true one. Therefore, the true positive rate was set to a specific value, namely 99%, and the error measurement used for evaluation was the false positive rate.

4. Temporal analysis

If the acquisition of thermal imagery covering a city is repeated one or a few years later, it is possible to compare the status of the network at the first acquisition with the status at the second acquisition. An automatic comparison method and a visualization technique have been developed for this purpose.

First, a grid consisting of cells, size 50×50 m, is created for the covered area, see Fig. 4. The grid has M rows and N

Table 2. Initial selection of features

Group	Initial features	Description	Final feature
Intensity	Median intensity	Median intensity within the detection.	Yes
	Mean intensity	Mean intensity within the detection.	No
	Standard deviation	Standard deviation of the intensity within the detection.	Yes
	Skewness	Skewness of the intensity within the detection.	No
	Kurtosis	Kurtosis of the intensity within the detection.	No
	Flatness	Flatness of the maximum intensity peak.	No
Shape & Propagation	Area	Area of the detection.	No
	Circumference	Circumference of the detection.	No
	Compactness	$\frac{\text{circumference}^2}{\text{area}}$	No
	Coverage	Ratio of detection area inside heat pipe mask.	Yes
	Eccentricity	Rate of the longest chord within the detection to the length of the longest chord perpendicular to the first one.	No
	Elongatedness	$\frac{\text{area}}{4d^2}$, where d is the number of erosions needed to make the detection disappear.	Yes
	Rectangularity	Ratio of the detection area to the area of the minimum bounding rectangle.	No
	Circularity	Ratio of the detection area to the area of a circle having the same perimeter.	No
	Concentricity	Measurement of how central the maximum intensity value is within the detection.	Yes
Proximity	Connected components	Number of other detections which lie within a certain radius from the detection.	Yes
	Border average	Mean intensity within an area around the detection.	Yes
	Distance to building	Distance from maximum intensity value to the wall of the closest building.	Yes

columns (depending on the size of the area). For each acquisition a and cell (m, n) , a total radiated power, $P_{m,n}^a$, is calculated ($a = 1, 2; m = 1, 2, \dots, M$ and $n = 1, 2, \dots, N$). Acquisition $a = 1$ is the earliest of the two acquisitions. Since we know the estimated ground temperature, we can compute the radiated power for each detection. For this purpose, Stefan-Boltzmanns law

$$\frac{dQ}{dt} = \epsilon \sigma (T^4 - T_0^4) A \quad (4)$$

is used. $\sigma = 5.67 \cdot 10^{-8} \text{ W m}^{-2} \text{ K}^{-4}$ is the Stefan-Boltzmann constant, A is the area, and ϵ is the emissivity of the object, which in this case mainly consists of ground in different forms. Soil, grass, and asphalt typically have an emissivity in the interval $0.90 - 0.96$. Information on the distribution of different ground types was not available for the calculations, hence, a constant value ($\epsilon = 0.92$) had to be used for all detections even though the different materials have slightly different radiative abilities. $T[K]$ is the mean temperature of all pixels within the detection. T_0 represents the background temperature and is estimated as the mean temperature of all pixels within the current heat pipe mask and flight. Ground temperatures were found by exporting intensity images with the Altair software, taking atmospheric effects etc. into consideration enabling comparison of different acquisitions. The total radiated power (TRP) of cell (m, n) and acquisition a , $P_{m,n}^a$, is calculated as in (5). For each layer l , the sum of the radiated power of all detections that have their centroids in cell (m, n) is computed. The TRP of a cell is then defined as the mean of this sum over layers, *i.e.*,

$$P_{m,n}^a = \frac{\sum_{l=1}^L \sum_{k \in S_{m,n,l,a}} \epsilon \sigma (T_k^4 - T_0^4) A_k}{L} \quad (5)$$

$S_{m,n,l,a}$ is the set of detections in acquisition a and layer l that have their centroids inside the boundaries of cell (m, n) . $l = 1, 2, \dots, L$ is the layer and, as previously mentioned, the total number of layers in this case is $L = 6$. The difference of radiated power, $\Delta_{m,n}$, for each cell represents the change in radiated power from the previous acquisition until the current one. It is calculated as the difference of TRPs for the two acquisitions $\Delta_{m,n} = P_{m,n}^2 - P_{m,n}^1$. When the comparison results are visualized to the operator, each cell is colored according to its calculated TRP difference, $\Delta_{m,n}$, and the grid is overlaid on top of a mosaic of the thermal images, see Fig. 5. Red indicates an increase in radiated power, $\Delta_{m,n} > 0$, and green indicates a decrease. Transparency is used to visualize the degree of change. The cells with the largest increase/decrease are assigned zero transparency and the cells with lowest full transparency. The transparency scale is then linearly distributed for all cells with TRP differences in between. If there is no change at all, *i.e.* $\Delta_{m,n} = 0$, the cell will be colorless. This visualization technique with red and green squares gives the operator a quick overview of the status of the network. He or she can soon pinpoint the most critical areas.

Note that given the rough assumption that all ground materials have equal emissivity, the visualization is invariant to the value of ϵ . In (5), ϵ is a constant and the visualization is relative to the largest increase/decrease in difference of the calculated radiated power. We chose $\epsilon = 0.92$ because it is a sufficiently good approximation, but any value for ϵ would have resulted in the same coloured boxes visualized to the operator.

Fig. 5. Visualization of changes in radiated power. A red square indicates that the area suffers from an increased energy loss while a green square means that the energy loss within the area has decreased.

It should be emphasized that the presented method for temporal analysis is an approximation. Since no information on the distribution of different ground materials could be provided, the emissivity is assumed to be constant even though the different materials have slightly different radiative properties. Further, the analysis provides a measurement of the radiated power at ground level, but how the heat transfers from the pipe through the soil remains unexplored. It is, however, clear that the properties of the pipe, material, depth, insulation etc., and the soil composition affects how much radiated power that reaches the ground surface. The full implications of these error sources remain unknown.

5. Experimental results

5.1. Building segmentation

Two building segmentation schemes were evaluated. The scheme proposed by Friman et al. (2013) using the thermal imagery, and the scheme using OpenStreetMap. It was observed that both schemes sometimes result in errors, but the errors were of different character. The method in Friman et al. (2013) sometimes suffer from non-building areas being classified as buildings, which could potentially make true detections (*i.e.*, real leakages) be discarded. In contrast, the OpenStreetMap-based scheme sometimes suffer from missing buildings, which might result in false detections classified as true detections. As mentioned, the cost for the latter is much lower, thus we selected the OpenStreetMap-based scheme. Examples of the performance of both methods can be seen in Fig. 6.

Using the OpenStreetMap-based scheme described, removing all detections which lie 100% on top of a building, 19% of the false detections in the data set could be removed without removing any of the energy or media samples. There is, however, a bias in the data set. The rate of false detections lying on top of buildings to the total number of false detections in the data

Fig. 6. (a) shows an example of the performance of the building segmentation method proposed by Friman et al. (2013). The green boundaries indicate the areas that have been classified as buildings. The arrows show the problem areas. Unacceptable classification of ground as buildings which could lead to missed real leakages. (b) shows instead an example of a successfully classified false detection, indicated with red boundaries, when the OpenStreetMap-based scheme is used.

set is probably a bit higher than the actual rate. This is due to manual annotation and the fact that this type of false detections are easier to find than others. Therefore, it should be observed that when the building segmentation approach is applied to data outside the data set, the percentage of removed false detections will probably be lower than 19%. Moreover, the total area of ground above district heating pipes located beneath buildings varies between different cities.

5.2. Classification

First of all, parameter values for all nonlinear classifiers are chosen, Sec. 5.2.1. Then, these values are used for classifier evaluation and selection, Sec. 5.2.2. The full feature set with 18 features is used for both parameter value and classifier selection. In Sec. 5.2.3, all features are ranked based on the Mahalanobis distance and a subset of eight features is selected. Finally, the possibility of improving results by voting or layer invariant classification is evaluated in Sec. 5.2.4.

5.2.1. Parameter selection

For each nonlinear classifier, parameters were evaluated and chosen by running each algorithm on a set of different parameter values. The parameter value that resulted in the smallest false positive rate, averaged (over all layers), was chosen. The true positive rate was fixed to 99%. Plots of averaged false positive rates can be seen in Fig. 7 and choice of parameter values in Table 3. The specific choices are discussed below.

- Radial Basis Support Vector Machine (RBF-SVM): Different values of the variance σ^2 were evaluated. Fig. 7a shows that there is an optimum in the vicinity of $\sigma^2 = 200$, and this value was thus chosen for the further evaluation. The regularization parameter c was set to the default value of PRTools. Other kernels (like, e.g., Gaussian) were not evaluated.
- AdaBoost: The commonly used weak classifier *decision stump* was chosen and a set of different values for the number of weak classifiers was tried. As shown in Fig. 7b, the

Table 3. Classifier parameters

Method	Parameters
RBF-SVM	Variance: $\sigma^2 = 200$ Regularization parameter: $c = 1$
AdaBoost	Weak classifier: Decision stump No. of weak classifiers: 192
Random Forest	No. of decision trees: 120 No. random features tried during training for splitting at each node: 1

false positive rate does not improve when the number of weak classifiers is larger than around 200. A local minimum could be found in 192 and that number was thus chosen.

- Random Forest: We used a truly randomized forest, i.e., a one random feature was assigned to each node without comparison to other features. Thus, the training procedure was fast, but might require more trees, and consequently the performance for different number of trees was evaluated, see Fig. 7c. The false positive rate seems to stabilize around 100 trees, and the value 120 was chosen as it gave the best result in the test.

5.2.2. Classifier selection

Each classifier was evaluated using 10-fold cross-validation and the full set of 18 features. The classifiers were trained and tested on each layer individually.

The averaged results over all layers are shown, for each individual classifier, in Fig. 8. The Random Forest classifier clearly outperforms the others. Combining the results from the building based rejection and the classification, the weighted (with respect to number of false samples in each layer) averaged false positive rate across layers is 58%. Since the Random Forest

Fig. 7. Averaged false positive rate (over layers) for different parameter values and classifiers.

Fig. 8. Averaged false positive rate for all evaluated classifiers and layers. The red line shows the median, the blue box the 25th and 75th percentile and the whiskers mark the most extreme samples not considered as outliers. Outliers are plotted as red crosses.

so clearly outperformed the other classifiers, it was deemed unnecessary to investigate slight variations of AdaBoost and SVM (such as Gaussian SVM kernels or other boosted weak classifiers).

5.2.3. Feature selection

Feature selection was conducted using a ranking based on the Mahalanobis distance, as described in Section 3. Mahalanobis distances for all features can be seen in Fig. 9. In order to exclude features not contributing significantly to the results, five different experiments were performed.

1. Use all **18 features** for classification.
2. Use the **13 features** with the largest Mahalanobis distance.
3. Use the **10 features** with the largest Mahalanobis distance.
4. Use the **3 features** with the largest Mahalanobis distance.
5. Remove the **3 features** with the largest Mahalanobis distance from the set.

Results of these experiments are shown in Fig. 10. Reducing the number of features from 18 to 10 does not make a significant difference to the classification results. However, reducing

Fig. 9. Averaged ranking of features. The features to the left with the highest Mahalanobis distances are the ones most important for discrimination. The blue vertical lines indicate the standard deviation.

from 10 to three yields a considerable drop in performance. Approximately the same performance is achieved when the three most discriminating features are removed, indicating their importance.

Both mean and median intensity are among the 10 most discriminating features. Since we anticipate them both to be discriminating in the same way, one of them, mean intensity, was removed. Further, also flatness could be removed without a significant loss of performance (58% false positive rate compared to 57% when classification results are combined with the building based rejection). Therefore, the eight chosen features are; coverage, border average, concentricity, median intensity, connected components, standard deviation, elongatedness and distance to building. These features are further described in Table 2. The ten removed features did not contribute significantly to the result and could thus be disregarded.

The features corresponding to shape and propagation of the detections are clearly under-represented, a somewhat expected result since the shape and propagation is heavily dependent on other factors than the leakage itself (such as soil physics and environmental conditions).

Fig. 10. Averaged false positive rate for all layers for the Random Forest classifier on different sets of features. The red line shows the median, the blue box the 25th and 75th percentile and the whiskers mark the most extreme samples not considered as outliers. Outliers are plotted as red crosses.

5.2.4. Classification and detection layers

As mentioned, the detection method results in several layers of detections, where the first layer contains the most obvious detections only (the ones with the most anomalous temperatures). In the previous sections, training and classification were done separately for each layer. We have investigated whether these classifiers can be combined, since (i) if a detection is present in several layers, adding information from the other layers might improve robustness, and, (ii) the differences in classification output from the different layers indicate the robustness of the classifiers. Both experiments below were conducted on the eight feature subset that was the result of feature selection.

Voting Since detections corresponding to the same area are often present in several layers, the possibility of improving the result by voting between layers has been investigated.

The idea of using a committee of classifiers is widely used and a common method is known as *Bootstrap aggregating* or *Bagging* (Bishop (2006)). Of course, there has to be some variability among the different classifiers for the voting to make sense. In this case, the variability is introduced by the same area having different detection appearances at different thresholds. On the other hand, a low variability would tell us that the classification is independent of the threshold level, *i.e.*, the classification would be more reliable. When evaluating, the samples were 10-fold cross-validated and at each classification, all classified labels from the different thresholds belonging to the same detection were found. If one of these labels was true, then the final voted label of the sample was set to true. The classifier used was the Random Forest classifier with 120 trees.

No improvement was found at any of the percentage thresholds. In fact, no occasion of dissimilar labelling by different threshold classifiers was observed at all. Accordingly, the classification of the detections across layers was proven to be consistent.

Layer invariant classification For the evaluation of layer invariant classification, all the 1585 detections from the six layers were combined to form one large data set.

Since the Random Forest classifier has shown to achieve the best general performance for all percentage thresholds, it was chosen to be the classifier also for the threshold invariant classification. Combining the results from the building based rejection and the classification, the false positive rate was 42%. Thus, to our surprise, this combined classifier gives bet-

Table 4. Summary of false alarm reduction results

Method	Average fpr
Building segmentation	81%
Build seg + classification	58%
Build seg + voting	58%
Build seg + layer invariant classification	42%

ter performance (on average) than a set of specialized classifiers. However, the results for the set of specialized classifiers is somewhat unreliable for layer 1 and 2 due to the small number of samples.

5.3. Summary of false alarm reduction results

Results for all evaluated false alarm reduction methods have been summarized in Table 4. The method with the smallest average false positive rate was the combination of the OpenStreet-map based building segmentation (Sec. 5.1) and layer invariant classification (Sec. 5.2.4). The proposed building segmentation scheme is preferable compared to the scheme proposed by Friman *et al.*, not due to its superior performance in general, but due to the different type of errors produced by the two methods (Sec. 5.1).

5.4. Temporal analysis

The temporal analysis had, at the time of writing, been used in the described form in one city. That city was not one of the three for which ground truth samples for classifier evaluation had been collected. However, some confirmed leakages have been provided by the network owner allowing us to draw some conclusions about the result. In Fig. 11 and Fig. 12, examples are shown of how the visualization clearly indicates cells containing confirmed media leakages as suffering from an increased energy loss. One particularly interesting example where the comparison acts as a complement to detection ranking based on radiated power is presented in Fig. 13. Here, a major media leakage of 70 m³/day gave rise to some headaches for the network owner who had searched unsuccessfully for the leakage for several years. The district heating pipe laid on top of a bed of gravel and beneath the said pipe was a larger stormwater pipe along which all media from the district heating pipe ran. Due to this choice of path by the media, the temperature difference that could be measured at ground level was only 3°C, placing the detection far down the detection ranking list. Nevertheless, the change in radiated power since the last acquisition was noticeable and in the visualization the cell containing the detection was marked in red.

In order to make an assessment of the overall status of the network at the current versus the previous acquisition, a scatter plot as the one in Fig. 14, can be used. Each scatter point corresponds to one cell and each cell has two TRPs, $P_{m,n}^1$ and $P_{m,n}^2$, here the x- and y-axis respectively. A red line has been fitted to the points through a least squares fit. If the energy loss has not changed, the angle α between the line and the x-axis is 45°. A smaller α , as in Fig. 14, indicates that in general, there has been

Fig. 11. A cell (a) that has been marked as suffering from an increased energy loss that proved to contain a true media leakage (b). The yellow arrow indicates the position of the leakage within the cell and the yellow number is the ID of the detection.

Fig. 12. Another example of a cell (a) that has been marked as suffering from an increased energy loss that proved to contain a true media leakage (b). The yellow arrow indicates the position of the leakage within the cell and the yellow number is the ID of the detection.

Fig. 13. Example of major leakage made visible by the comparison acting as a complement to detection ranking based on radiated power.

Fig. 14. Scatter plot of the TRPs for each cell (black dots). The red line is a line that has been fitted to the points. It indicates whether or not the overall energy loss of the network has decreased or increased.

a decrease in energy loss. In fact, this conclusion coincided with the network owners feeling about the networks status.

6. Conclusion

We have presented an improvement to a previously published method for finding leakages in district heating networks by classifying its resulting detections using trained classifiers. We have evaluated various features and five different classifiers. The classifier that generally produced the smallest false positive rate while maintaining a true positive rate of 99% is a Random Forest classifier with 120 trees, an average tree depth of 10 and splitting at nodes based on one randomly selected feature. The classification of different layers of detections have been investigated, and the results indicate that the classification is consistent over layers.

We have also shown that a pre-processing step for building extraction using publicly available GIS data is preferable compared to the previously used method, not due to its superior performance in general, but due to the different type of errors produced by the two methods.

In the end, we were able to achieve a false positive rate of 42%, that is, we can discard 58% of the false detections given by the previous system, thus significantly enhancing the usability.

The proposed temporal analysis improves usability of the system and the visualization allows the operator to get a quick overview of what areas that should be studied more carefully.

Acknowledgments

This work has been partly funded by the Swedish Research Council (Vetenskapsrådet) through the project Learning systems for remote thermography, grant no. 621-2013-5703, and by the The Swedish Research Council through a framework grant for the project Energy Minimization for Computational Cameras (2014-6227).

References

- Axelsson, S.R.J., 1988. Thermal modeling for the estimation of energy losses from municipal heating networks using infrared thermography. *IEEE Transactions on Geoscience and Remote Sensing* 26, 686692.
- Bishop, C.M., 2006. *Pattern Recognition and Machine Learning*. First ed., Springer Science.
- Blaschke, T., 2010. Object based image analysis for remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing* 65, 2–16.
- Bohm, B., Borgström, M., 1996. A comparison of different methods for in-situ determination of heat losses from district heating pipes. Technical Report. Dept. of Energy Engineering, Technical University of Denmark.
- Breiman, L., 2001. Random forests. *Machine learning* 45, 5–32.
- D. Lu, P. Mausel, E.B., Moran, E., 2003. Change detection techniques. *Int. Journal of Remote Sensing* 25(12), 2365–2407.
- Dekker, R.J., Schwing, P.B., Benoist, K.W., Pignatti, S., Santini, F., 2013. Lwlr hyperspectral change detection for target acquisition and situation awareness in urban areas. *Proc. SPIE* 8743, 153–158.
- Duda, R.O., Hart, P.E., Stork, D.G., 2001. *Pattern Classification*. Second ed., John Wiley & Sons.
- Duin, R., Juszczak, P., Paclik, P., Pekalska, E., de Ridder, D., Tax, D., Verzaikov, S., 2007. PR-Tools4.1, a Matlab toolbox for pattern recognition. <http://www.prtools.org>. Version 4.2.4.
- Friman, O., Follo, P., Ahlberg, J., Sjökvist, S., 2013. Methods for large-scale monitoring of district heating systems using airborne thermography. *IEEE Transactions on Geoscience and Remote Sensing*.
- Fröling, M., 2002. Environmental and thermal performance of district heating pipes. Ph.D. thesis. Chalmers university of technology.
- Hastie, T., Tibshirani, R., Friedman, J., 2008. *The elements of statistical learning*. Second ed., Springer.
- Jain, A., Zongker, D., 1997. Feature selection: Evaluation, application, and small sample performance. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 19, 153–158.
- Li, X., Zhang, L., Guo, H., Sun, Z., Liang, L., 2012. New approaches to urban area change detection using multitemporal radarsat-2 polarimetric synthetic aperture radar (sar) data. *Canadian Journal of Remote Sensing* 38(3), 253–266.
- Ljungberg, S.A., 1987. Aerial thermography - a tool for detecting heat losses and defective insulation in building attics and district heating networks, in: *Proceedings of SPIE Thermosense IX: Thermal Infrared Sensing for Diagnostics and Control*, p. 257265.
- Meola, J., Eismann, M., Moses, R., Ash, J., 2011. Detecting changes in hyperspectral imagery using a model-based approach. *IEEE Trans. Geoscience and Remote Sensing* 49(7), 2647–2661.
- Olsson, M., 2001. Long-term thermal performance of polyurethane-insulated district heating pipes. Ph.D. thesis. Chalmers university of technology.
- Sjökvist, S., Wren, J., Ahlberg, J., 2012. Kvantifiering av värmelckage genom flygburen IR-teknik en förstudie. Technical Report 2012:17. Fjärrsyn.
- Theiler, J., 2008. Quantitative comparison of quadratic covariance-based anomalous change detectors. *Applied Optics* 47(28), F12–F26.
- Zinko, H., Björklev, J., Björström, H., Borgström, M., Bohm, B., Koskelainen, L., Phetteplace, G., 1996. Quantitative Heat Loss Determination by Means of Infrared Thermography - The TX model. Technical Report. International Energy Agency.