MSER-based Framework for Classification of Objects in Thermal Images

Alia Aljasmi[1] and Andrzej Śluzek[1,2][a]
[1]*Khalifa University, Abu Dhabi, U.A.E.*
[2]*Warsaw University of Life Sciences-SGGW, Warsaw, Poland*

Keywords: Thermal Images, MSER, Object Detection, Shape Descriptors, Object Classification.

Abstract: In this paper, the problem of multi-class object recognition in thermal images is discussed. An alternative model of thermal objects is investigated, where an object is represented by multiple shapes extracted by MSER detectors. The shapes are nested within the largest MSER outlining the object (which might be the actual outline of the object, the outline of its thermal footprint or the outline of its largest prominent fragment). We show, using a multi-class dataset of thermal images captured in indoor environments, that the proposed methodology is a feasible solution for various object classification problems in thermal imaging. In particular, no object-specific algorithms are needed, so that the method is applicable to most of typical applications of thermal cameras (subject to general limitations of data captured by thermal imaging devices). The presented work is considered a preliminary feasibility study exploring potentials an limits of thermal image classification in more sophisticated machine vision problems.

1 INTRODUCTION

Thermal images are an alternative representation of visually difficult environments (poor illumination, foggy/smoky conditions, confusing patterns/camouflage etc.) which nevertheless contain objects of distinctive temperature profiles. The most popular applications in visual surveillance and monitoring tasks include, see (Gade and Moeslund, 2014), detection and tracking of moving objects (humans, animals, vehicles, e.g., (Wang et al., 2010; Fernandez-Caballero et al., 2014; Zhou et al., 2009; Christiansen et al., 2014; Iwasaki et al., 2013), etc.), inspection, security and quality control, and other selected industrial applications (e.g., (Sirmacek et al., 2011; Vidas et al., 2013; Ginesu et al., 2004; Ng et al., 2007; Meriaudeau et al., 2010), etc.).

However, applications of thermal imaging in typical problems of multi-class object identification are rather limited. This can be attributed to the following factors. First, the spatial resolution of thermal cameras is still low, compared to standard cameras. Secondly, the visual distinctiveness in thermal images is rather poor due to heat radiation an dissipation. Therefore, objects in thermal images are typically blurred and poorly contrasted, where regions (often with boundaries only approximately delimited) are the sole available representation of those objects. Correspondingly, very few experimental works have been reported on classification of several types of objects within the same task, where only thermal imaging is used (e.g. (Meis et al., 2003)). The majority of thermal imaging applications applications focus on object detection and subsequent tracking. Not surprisingly, the diversity of features used in such works is also limited (mostly binary regions and/or characteristics of their boundaries) and the reported results are not very impressive, even with features hand-crafted for specific problems and a limited number of considered classes (as in (Meis et al., 2003)).

In this paper, object classification in thermal imaging for is discussed from a more general perspective, even though we (indirectly) focus on indoor tasks (e.g. visual surveillance in dark premises). Primarily, we investigate an alternative model of objects in thermal images (where each instance of an object is represented by multiple regions extracted by MSER detectors (Matas et al., 2002; Nistér and Stewénius, 2008), as explained in Section 2). Subsequently, in Section 3, we use a simple classification method to:

- Identify 3D objects from a range of diversified classes using regions extracted by MSER detec-

[a] https://orcid.org/0000-0003-4148-2600

Aljasmi, A. and Śluzek, A.
MSER-based Framework for Classification of Objects in Thermal Images.
DOI: 10.5220/0008116105660572
In *Proceedings of the 16th International Conference on Informatics in Control, Automation and Robotics (ICINCO 2019)*, pages 566-572
ISBN: 978-989-758-380-3

tor from thermal images.

- Distinguish (from a sequence of thermal frames) between rigid (fixed-geometry) and articulated (e.g. animals, humans, walking toys, etc.) objects. This is a supplementary objective.

The experimental results are also discussed in Section 3. Finally, the paper is briefly summarized in Section 4.

2 MSER-BASED MODELS OF THERMAL OBJECTS

In thermal images, even objects with distinctive temperature profiles are normally seen as diluted silhouettes on images of rather poor quality. Such images can be subsequently binarized (into objects and background) for further analysis and processing. Unfortunately, standard image thresholding algorithms (e.g. (Sezgin and Sankur, 2004; Puneet and Garg, 2013)) cannot handle typical effects of thermal images (as illustrated in Fig. 1) so that accurate outlines of the actual objects may not be reliably extracted.

Therefore, we propose an alternative approach based on *maximally stable extremum regions* (MSERs). MSERs are the image fragments which are least sensitive to the binarization threshold variations (see (Matas et al., 2002; Nistér and Stewénius, 2008)). Therefore, MSER detector can identify in infrared images even poorly contrasted fragments, as long as these fragments have the most distinctive thermal profiles within the processed image. Actually, MSERs can be detected at very small computational costs (including on-chip implementation of the detector (Sluzek et al., 2019)) which makes them particularly attractive for low-cost systems (i.e. IoT devices).

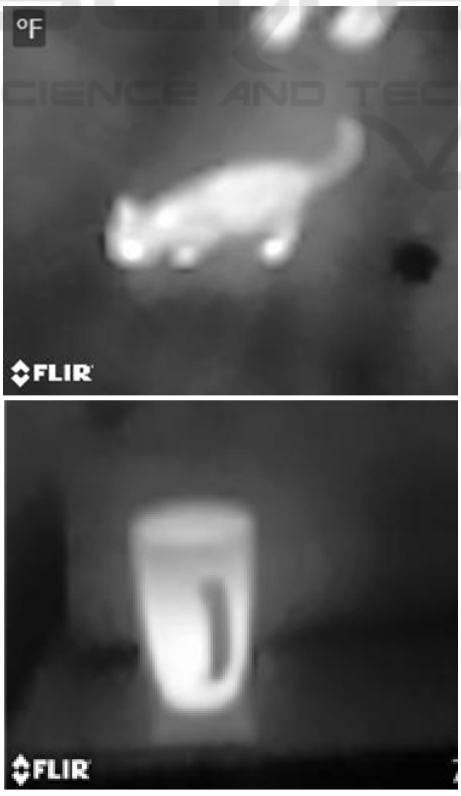

Figure 1: Examples of raw-data thermal images of objects.

Formally, binary MSER regions $Q(t)$ (where t indicates the threshold level) are detected as local minima of the growth rate function $q(t)$ defined by the derivative of the region's area over the threshold values:

$$q(t) = \frac{\frac{d}{dt}|Q(t)|}{|Q(t)|}, \tag{1}$$

where $|\cdot|$ represents the area of a region.

MSER detector has been reported in some applications of thermal images, e.g. (Lahouli et al., 2018). Nevertheless, the most significant advantage of MSER in thermal images, i.e. their ability to extract not only the outlines of objects, but also distinctive internal fragments of objects or their thermal shadows on the surrounding scenes, etc. (as illustrated in examples in Fig. 2) is apparently not fully exploited yet in the available literature. Therefore, we propose to represented thermal objects by the family of MSERs nested within the largest MSER outlining the object.

This largest MSER can be the actual outline of the object shape, the outline of its thermal footprint (i.e. incorporating the heat radiation effects on the object neighborhood) or the outline of the largest prominent fragment within the object (if the whole object is indistinguishably blended with its background). Certain practical constraints are obviously applied, namely removal of too small/large MSERs from the processed thermal images or rejection of MSERs which cannot (for various reasons) represent physical objects.

An example of a simple thermal object represented by two binary shapes of its MSERs is given in Fig. 3.

The practicality of this methods has been tested in two experiments:

- In the first experiment, the objective is to classify detected in thermal images objects from a collection of exemplary classes of 3D objects.
- In the second experiment, we attempt to classify moving objects into either *rigid* category (bodies of fixed geometry) or *articulated* category (i.e.

mechanical or biological bodies changing their configuration while in motion).

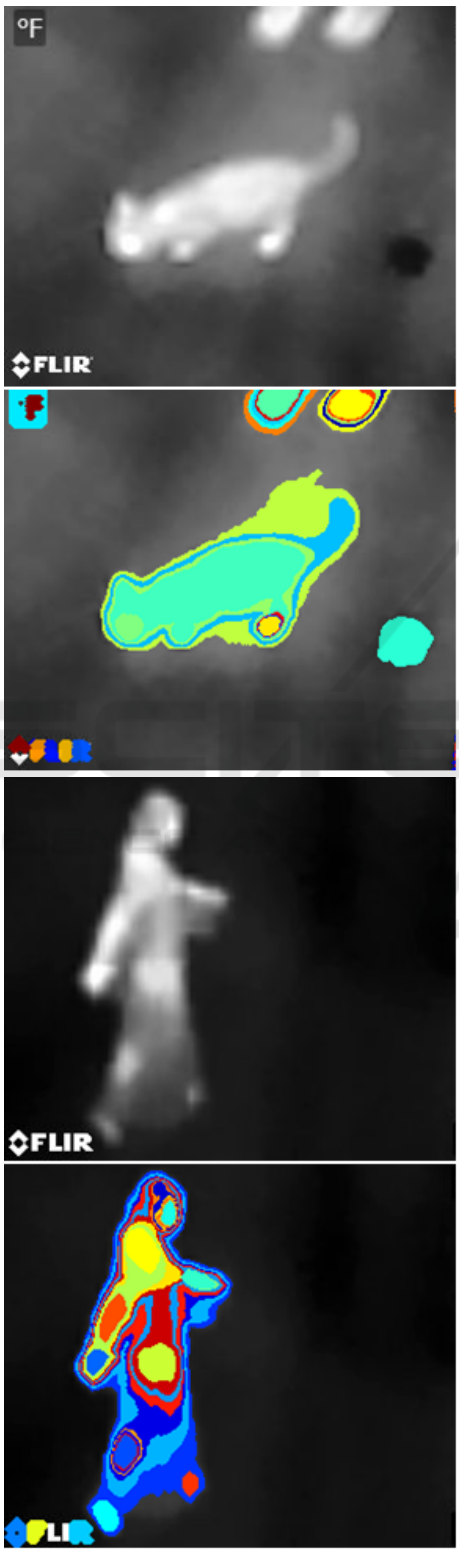

Figure 2: Examples of MSER detection in thermal images.

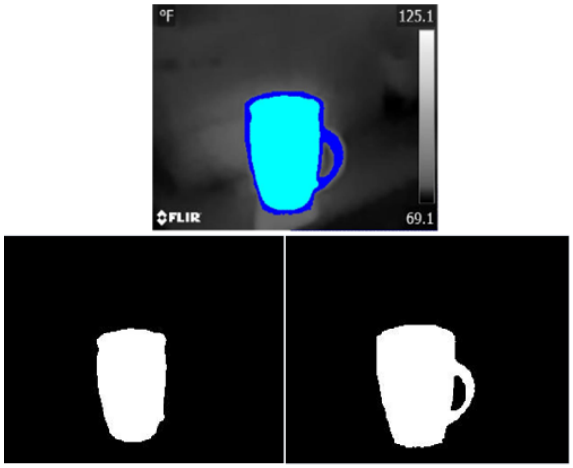

Figure 3: Example of a thermal object represented by two shapes of its MSERs.

All dataset and test images for the experiments are captured in natural indoor environments at 640×480 resolution, using FLIR C2 thermal camera.

2.1 Dataset for Multi-class Recognition

For the multi-class recognition experiment, a dataset of over 700 images has been collected for 13 diversified objects, including 8 rigid objects (*glass_plant*, *cup*, *bottle_of_water*, *bottle_of_juice*, *iron*, *plate*, *stapler* and *kettle*) and 5 fully or partially articulated objects (*natural_flower*, *pot_plant*, *woman_in_abaya*, *bicycle* and *teddybear*). For rigid objects, the images were captured from sufficiently diversified viewpoints, while for articulated objects various geometric configurations were additionally taken into consideration. Examples of the dataset images (for two different objects) are shown in Fig. 4.

Eventually, each class C_i is modeled by a collection of *all* binary MSER regions $\{MSER_1(C_i), ..., MSER_{n_i}(C_i)\}$ extracted from all ground-truth examples of the corresponding category.

2.2 Datasets for *Rigid-articulated* Categorization

In this experiment, no permanent dataset is used. Instead, temporary reference datasets are dynamically updated from the most recent 5 frames $\{t, ..., t-4\}$ containing the object of interest O. Binary MSER shapes (extracted from the object in the same way as above) are grouped within the corresponding frames, i.e.

$$Frame0 \rightarrow \{MSER_1(O_t), ..., MSER_{n_0}(O_t)\}$$
$$...$$
$$Frame4 \rightarrow \{MSER_1(O_{t-4}), ..., MSER_{n_4}(O_{t-4})\}$$

The union of these groups of shapes is considered the currently used representation of the analyzed object.

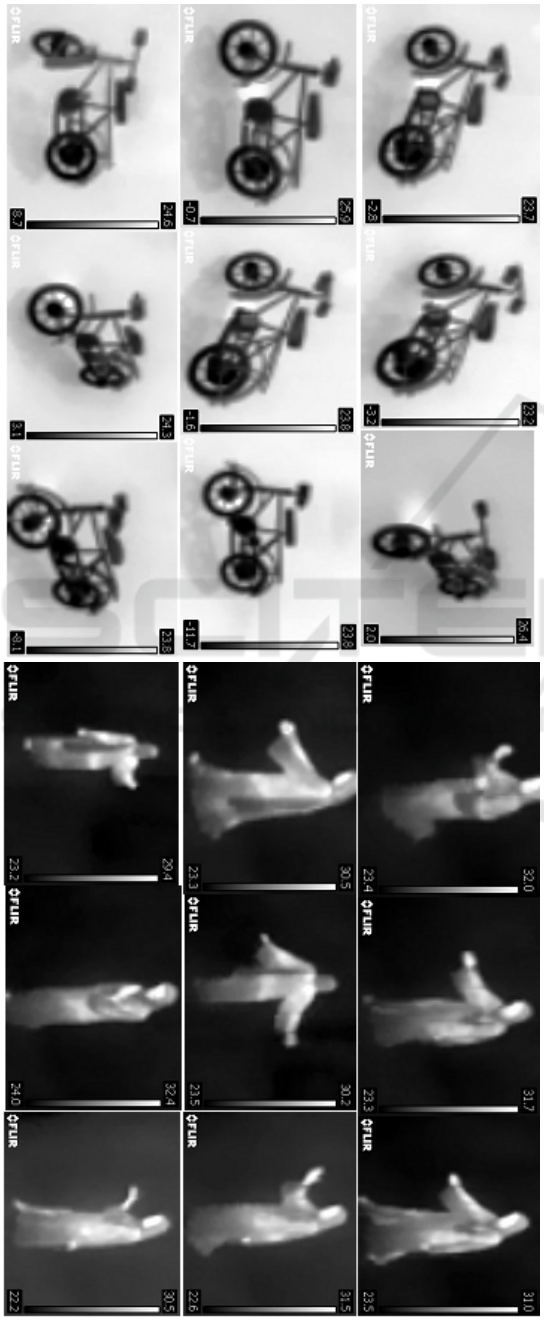

Figure 4: Examples of thermal dataset images for two classes.

3 OBJECT CLASSIFICATION BY SHAPE DESCRIPTORS

Since (as discussed in Section 2) the thermal objects of interest (and classes of objects) are eventually represented by collections of binary shapes, only binary shape descriptors can be used for object classification. Because regions extracted from thermal images are generally smooth and without intricate shape details, we preliminarily selected very popular Hu moment invariants, e.g. (Sluzek, 1995; Flusser, 2000), to represent MSER regions by 7D vectors (with some refinements as specified below):

$$V7 = \{I_1, ..., I_7\} \tag{2}$$

where I_i are the original invariants ϕ_i from (Hu, 1962) normalized to have the same mean and standard deviation values. The normalization was done on a popular benchmark dataset of binary shapes MPEG7[1], and the first Hu invariant was used as the reference.

It was later experimentally verified that vectors of invariant with the dimensionality reduced (by using PCA) to 3D provide practically the same performances, so that the regions are alternatively represented by $V3$ vectors

$$V3 = \{I(PCA)_1, I(PCA)_2, I(PCA)_3\} \tag{3}$$

Now, similarity between MSER regions (which is needed for the region-based classification of thermal objects) can be defined as follows. Given a dataset region R_D and an object region R_O, the level of similarity between R_O and R_D is straightforwardly defined as:

$$Sim(R_O, R_D) = 1 - \frac{\|V7(R_O) - V7(R_D)\|}{\|V7(R_D)\|} \tag{4}$$

or

$$Sim(R_O, R_D) = 1 - \frac{\|V3(R_O) - V3(R_D)\|}{\|V3(R_D)\|} \tag{5}$$

where $\|\cdot\|$ is the vector norm.

Eventually, we consider two regions similar if their level of similarity by Eq. 4 or Eq. 5 exceeds the predefined (established experimentally) threshold.

Two examples of diversified similarity levels between binary regions are given in Fig. 5.

[1] http://www.dabi.temple.edu/~shape/MPEG7/MPEG7dataset.zip

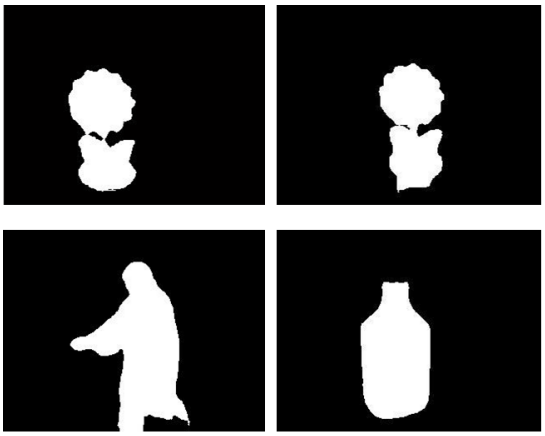

Figure 5: Regions with 0.94 similarity (top row) and with 0.54 similarity (bottom row).

3.1 Multi-class Recognition of Thermal Objects

Given a thermal object O (represented by a number of binary MSER shapes $\{MSER_1(O),...,MSER_m(O)\}$) and the class C_i (modeled, as defined in Subsection 2.1, by binary MSER shapes $\{MSER_1(C_i),...,MSER_{n_i}(C_i)\}$) we assume that O is similar to C_i class (i.e. it can potentially be a member of this class) if:

1. Several shapes from $\{MSER_1(O),...,MSER_m(O)\}$ set are similar to some $\{MSER_1(C_i),...,MSER_{n_i}(C_i)\}$ shapes. In practice, the minimum number of similarities may be required.
2. If O object is similar to too many classes, only the classes with the highest numbers of inter-shape similarities (e.g. top three) are eventually accepted. This assumption is applied in our tests.

For the actual tests, we selected only five classes from the developed dataset, namely *glass_plant*, *woman_in_abaya*, *bottle_of_juice*, *teddybear* and *pot_plant* (the remaining classes acting as confusion data only). Fig. 6 shows the confusion matrix of the classification statistics. Unfortunately, we did not find any suitable benchmark to compare to, but the obtained results can be approximately compared to (Meis et al., 2003) (which is the only similar example we found of multi-class recognition in thermal objects) and the outcome should be considered satisfactory.

		CLASSIFIED OBJECTS				
		GLASS_PLANT	WOMAN_IN_ABAYA	BOTTLE_OF_JUICE	TEDDYBEAR	POT_PLANT
GROUND TRUTH	GLASS_PLANT	45.5%	8.3%	15.8%	0%	30.4%
	WOMAN_IN_ABAYA	0%	44.4%	33.4%	22.2%	0%
	BOTTLE_OF_JUICE	7.69%	15.38%	46.17%	0%	30.76%
	TEDDYBEAR	0%	28.57%	14.28%	57.15%	0%
	POT_PLANT	23.07%	21.02%	7.56%	0%	49.35%

Figure 6: The confusion matrix for 5-class test results.

3.2 Recognition of Rigid and Articulated Thermal Objects

In the second experiment, we tested the method's ability to distinguish between rigid and articulated objects. Objects of both categories can move (with respect to he camera) but over short periods of time only the articulated objects are expected to significantly change their shapes in he captured thermal images. Thus, as explained in Subsection 2.2, the sequence of most recent five images (frames) is used to identify the object category. As shown in Fig. 7, we build a matrix of inter-frame similarities, where $\times$ marker indicates that similar MSER regions are found in the corresponding pair of frames. If less than 8 entries (i.e. 40%) are marked, the object is considered *articulated*, and if the number of marked entries exceeds 12 (i.e. 60%) the object is recognized as *rigid*. Otherwise, no decision is made.

	Frame 0	Frame 1	Frame 2	Frame 3	Frame 4
Frame 0		X			
Frame 1	X			X	
Frame 2	X				X
Frame 3		X			
Frame 4					

Figure 7: Exemplary similarity matrix between 5 subsequent frames (the content of this matrix represents an articulated object).

Performances of this approach are illustrated by the confusion matrix in Fig. 8. Though some rigid objects are wrongly classified as articulated (as they may move/rotate relatively to the camera, or the thermal conditions of the scene change) we have not found any case of an articulated object (performing the actual motion) recognized as rigid. The percentage of *unknown* decisions is moderate. Altogether, we consider these results satisfactory, at least within the classes of tested objects.

GROUND TRUTH	CLASSIFICATION			
		RIGID	ARTICULATED	UNKNOWN
	RIGID	71.2%	19.3%	10.5%
	ARTICULATED	0%	86.2%	13.8%

Figure 8: The confusion matrix for the results of rigid-articulated classification.

Examples of rigid and articulated sequences are given in Fig. 9, where only the external outline MSERs are shown.

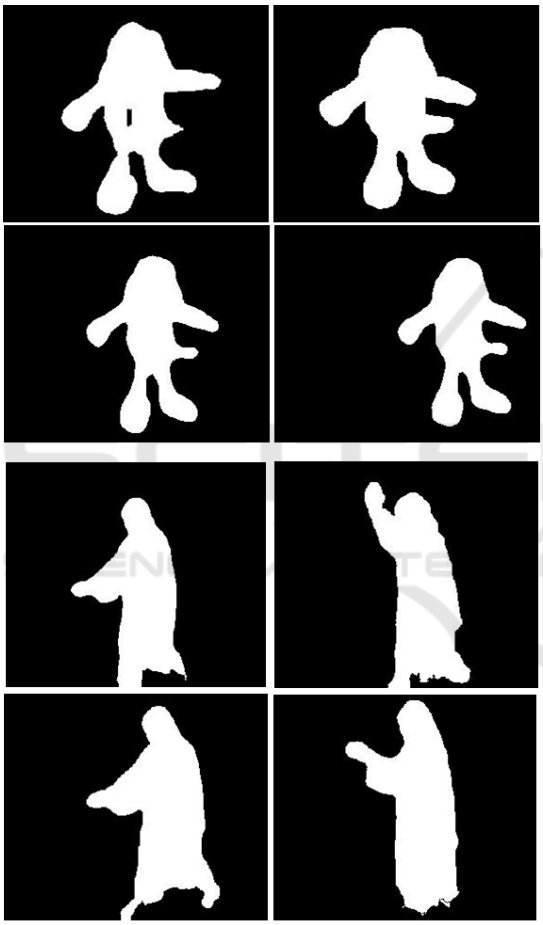

Figure 9: Examples of sequences showing *rigid* (two top rows) and *articulated* (bottom rows) objects.

4 CONCLUDING REMARKS

The presented work is a preliminary feasibility study exploring potentials of thermal imaging in more sophisticated applications than typical detection and tracking tasks. In particular, we consider the prospective needs of visual surveillance and monitoring systems in environments which should remain dark. In many such problems, the major task is not to detect the presence of thermally distinctive objects, but rather to classify them in terms of their identity and/or behavior (e.g. to identify dangerous or critical scenarios).

Our results indicate that such results are practical (subject to well-known limitations of thermal imaging) under some constraints, e.g. with rather limited numbers of object classes, non-overlapping objects, etc.

We can also conclude that for thermal images suitable representation of objects is, in general applications, more critical than the specific features/descriptors. The presented results have been obtained usign (deliberately) simplified shape descriptors. Because of such a simplification, the presented algorithms are suitable for low-cost solutions (including IoT devices, small robotic systems, etc.).

REFERENCES

Christiansen, P., K.A. Steen, R. J., and Karstoft, H. (2014). Automated detection and recognition of wildlife using thermal cameras. *Sensors*, 14(8):13778–13793.

Fernandez-Caballero, A., Lopez, M., and Serrano-Cuerda, J. (2014). Thermal-infrared pedestrian roi extraction through thermal and motion information fusion. *Sensors*, 14(4):6666–6676.

Flusser, J. (2000). On the independence of rotation moment invariants. *Pattern Recognition*, 33:1405–1410.

Gade, R. and Moeslund, T. (2014). Thermal cameras and applications: a survey. *Machine Vision & Applications*, 25(1):245–262.

Ginesu, G., Giusto, D., Margner, V., and Meinlschmidt, P. (2004). Detection of foreign bodies in food by thermal image processing. *IEEE Trans. on Industrial Electronics*, 51(2):480–490.

Hu, M. (1962). Visual pattern recognition by moment invariants. *IRE Transactions on Information Theory*, 8(2):179–187.

Iwasaki, Y., Misumi, M., and Nakamiya, T. (2013). Robust vehicle detection under various environmental conditions using an infrared thermal camera and its application to road traffic flow monitoring. *Sensors*, 13(6):7756–7773.

Lahouli, I., Haelterman, R., Chtourou, Z., Cubber, G., and Attia, R. (2018). Pedestrian detection and tracking in thermal images from aerial mpeg videos. In *Proc.13 Int. Joint Conf. VISIGRAPP 2018*, volume 5(VISAPP), pages 487–495.

Matas, J., Chum, O., Urban, M., and Pajdla, T. (2002). Robust wide baseline stereo from maximally stable extremal regions. In *Proc. British Machine Vision Conference*, pages 384–393.

Meis, U., Ritter, W., and Neumann, H. (2003). Detection and classification of obstacles in night vision traffic

scenes based on infrared imagery. In *Proc. 2003 IEEE Int. Conf. on Intelligent Transportation Systems*, pages 1140–1144.

Meriaudeau, F., Secades, L., Eren, G., Ercil, A., Truchetet, F., Aubreto, O., and Fofi, D. (2010). 3-d scanning of nonopaque objects by means of imaging emitted structured infrared patterns. *IEEE Trans. on Instrumentation and Measurement*, 59(11):2898–2906.

Ng, Y.-M., Yu, M., Huang, Y., and Du, R. (2007). Diagnosis of sheet metal stamping processes based on 3-d thermal energy distribution. *IEEE Trans. on Automation Science & Eng.*, 4(1):22–30.

Nistér, D. and Stewénius, H. (2008). Linear time maximally stable extremal regions. In *Proc. 10th European Conf. ECCV 2008*, pages 183–196.

Puneet, G. and Garg, N. (2013). Binarization techniques used for grey scale images. *Int. Journal of Computer Applications*, 71(1):8–11.

Sezgin, M. and Sankur, B. (2004). Survey over image thresholding techniques and quantitative performance evaluation. *Journal of Electronic Imaging*, 13(1):146–168.

Sirmacek, B., Hoegner, L., and Stilla, U. (2011). Detection of windows and doors from thermal images by grouping geometrical features. In *Proc. 2011 Joint Urban Remote Sensing Event*.

Sluzek, A. (1995). Identifcation and inspection of 2-d objects using new moment-based shape descriptors. *Pattern Recognition Letters*, 16(7):687–697.

Sluzek, A., Saleh, H., Mohammad, B., Al-Qutayri, M., and Ismail, M. (2019). Mser-in-chip: An efficient vision tool for iot devices. In Elfadel, I. and M.Ismail, editors, *Innovations in Intelligent Image Analysis*, pages 245–259. Springer.

Vidas, S., Moghadam, P., and Bosse, M. (2013). 3d thermal mapping of building interiors using an rgb-d and thermal camera. In *Proc. 2013 IEEE Robotics & Automation Conf. (ICRA)*, pages 2311–2318.

Wang, W., Zhang, J., and Shen, C. (2010). Improved human detection and classification in thermal images. In *Proc. 2010 IEEE ICIP Conference*, pages 2313–2316.

Zhou, D., Dillon, M., and Kwon, E. (2009). Tracking-based deer vehicle collision detection using thermal imaging. In *Proc. IEEE Int. Conference on Robotics and Biomimetics (ROBIO)*.