ISSN 2464-4617 (print) WSCG 2016 - 24th Conference on Computer Graphics, Visualization and Computer Vision 2016
ISSN 2464-4625 (CD-ROM)

Infrared-based Object Classification for the Surveillance of Valuable Infrastructure

Christos Palaskas[1] Savvas Rogotis[1] Dimos Ioannidis[1,2] Dimitrios Tzovaras[1] Spiros Likothanassis[2]

[1]Information Technologies Institute – Centre for Research and Technology Hellas
57001,Thermi,Thessaloniki, Greece
{chris.palaskas,srogotis,djoannid, dimitrios.tzovaras}@iti.gr

[2]Pattern Recognition Laboratory – Computer Engineering and Informatics – University of Patras
26500,Rio, Patras, Greece
{djoannid,likothan}@ceid.upatras.gr

ABSTRACT

The surveillance of valuable infrastructure, such as photovoltaic parks, is considered of fundamental importance for their proper function and maintenance as well as the avoidance of criminal damage incidents. At the same time, the privacy of employees working in the same area should not be jeopardized and their personal data should always be protected. The use of thermal cameras presents a solution to both of the above issues by offering an unobtrusive surveillance approach with the ability to supervise industrial premises under a wide range of environmental and situational conditions. The current paper proposes an algorithm for the classification of moving objects that aims to increase the efficiency of surveillance methodologies by shifting the focus on high-risk classes, such as humans instead of animals. The proposed methodology utilizes an automated decision framework that determines when textural features are fit to be used, based on the discriminative power of the texture of the object. Many texture descriptors were tested, including Local Phase Quantisation and Histograms of Oriented Gradients, resulting in the use of a lately proposed combination of these descriptors. This new multi-class object classification approach introduces the use of confidence values and a voting system to achieve a more accurate selection of the appropriate class. The velocity was also used as a discriminative feature, especially to help distinguish between humans and motorcycles. Several algorithms have been used to validate the results of the experimental studies with special focus on the classification accuracy. The experimental results were obtained from a series of scenarios demonstrated in four different condition sets (different temperature-humidity-illumination), that exposes the advantages and disadvantages of the proposed unimodal classification method in infrared imagery. The dataset is also benchmarked against another state-of-the-art approach.

Keywords

Thermal Imaging; Multi-Class Classification; Shape Descriptor; Texture Descriptor; Local Phase Quantization; Histogram of Oriented Gradients; Contour Point Distribution; Surveillance

1. INTRODUCTION

Monitoring and surveilling facilities and estates has been a primary need of human society since the dawn of time. There is a number of ways to obtain information to support the protection of a facility, such as color cameras, depth cameras, thermal cameras, noise sensors, CO_2 emission sensors and so on. Over the last few years, an explosion of camera invasion in everyday life was witnessed: streets, airports, hospitals, shopping malls, office buildings are loaded with cameras, surveilling and tracking anything that moves within their field of view. This camera-ubiquity amplified the necessity for more privacy preserving techniques, but without compromising the level of security. Such a technology is the monitoring of facilities using infrared cameras, since no visual features of the human face are obtained. Although there are some disadvantages using such a camera instead of a visual one, such as the low resolution of IR images and the high market price, there are also advantages, such as the ability to operate even at challenging conditions (e.g. nighttime, through fog, smoke). The use of such technology has released lots of human resources in the field of security and surveillance,

replacing guards with intelligent surveillance systems.

The main input instrument in the process of surveilling, the camera, can function in various wave ranges: optical, long wave infrared, short wave infrared, medium wave infrared etc. Many works have been published dealing with the fusion of more than one modes of input. In general, the tradeoff between optical imagery and thermal – IR imagery is that the first gives more information in specific environmental sets but almost no information in other sets, i.e. optical imagery works only with optical light but yields no information in darkness, whereas IR imagery, which depends on the emission of IR light due to the temperature of the bodies, is not affected by light, making it a more efficient method to monitor during nighttime.

This paper presents an improved method for classifying moving objects from infrared images into 4 classes, namely human, car, motorcycle and animal, using spatial features (shape, texture) and temporal (velocity) and feeding them to a voting system. The textural features are used only when the image of the object has enough colour discrimination, otherwise the classification is based only on the shape and velocity. The algorithm has been tested in multiple and demanding weather sets and illumination conditions. The features (shape and texture) have been selected based on performance, after testing and discarding many other features. Also the proposed classifier has been tested using the same features on another classifier. Finally the whole approach was benchmarked against another state-of-the-art algorithm, using the same challenging dataset with encouraging results.

The remainder of the paper is organized as follows: Section 2 reports some of the related work in the field. In Section 3 the implemented method is described, namely the feature extraction of the object, the training of the model and the final classification process. Experiments and discussions are provided in Section 4, while Section 5 concludes the paper.

2. RELATED WORKS

Effective monitoring and surveillance is prerequisite for adequate object detection [Jo13a]. Other ways are background subtraction [Bar11a] [Van12a], or segmentation [Arb11a] [Alp12a] followed by tracking [Bab11a] within the camera's field. As the object has been discriminated from the background it is feasible to extract its features and move on to its classification.

Many works dealing with classification methods have been published, but most of them concern the optical input mode (RGB and grayscale). In one of these works [Lia15a] a two-level Haar wavelet transform is applied to the bounding window of the object in an RGB colored image, and from these two level bands the local shape features and the Histogram of Oriented Gradients (HOG), are extracted. These are fed to a Support Vector Machine (SVM) which has been trained from a data set, so as to classify the object into one of four classes (human, bicycle, motorcycle, car).

There have also been some approaches [Ku10a] [Shu11a] where only the shape was used to classify objects in infrared imagery. On the other hand, many works use only textural descriptors, as will be presented below, but nothing significant was found using both and also determining when it is profitable to do so, as is the focus of this work.

In a recent work which aimed at detecting pedestrians at night [Joh15a] an adaptive fuzzy C-means clustering was adopted to segment the IR images and retrieve the candidate pedestrians. A convolutional neural network was then used to simultaneously learn relevant features and perform the binary classification.

In another recent work [Wan15a] a spatiotemporal saliency model based on three-dimensional Difference-of-Gaussians filters was proposed for small moving object detection in infrared videos. First, instead of utilizing the spatial Difference-of-Gaussians (DoG) filter which has been used to build saliency models for static images, they proposed the extension of the spatial DoG filter to construct three-dimensional (3D) Difference-of-Gaussians filters for measuring the center-surround difference in the spatiotemporal receptive field. After that, an effective spatiotemporal saliency model was generated based on those filters.

Attempting to detect humans in infrared imagery using a gradient-based technique, the authors in [Olm12a] introduced the exploitation of local information histogram of orientations of phase coherence. Thus, they obtained a scale, brightness and contrast invariant descriptor that can detect pedestrians in IR images. In another work [Che12a] the geometrical-based features and the texture-based features are extracted and then are fed to two trained classifiers: the kNN and the SVM. The geometrical features used are: aspect ratio, compactness, fill ratio, and Hu's invariant moments [Hua10a]. The textural features are: smoothness, uniformity and skewness (a metric that indicates where the bulk of the distribution histogram lies, i.e. to the left or to the right and to what extent. The features are range-normalized and then a Principal Component Analysis (PCA) is performed, to maintain the most significant features of the classification.

A shape-based fuzzy network [Jua08a] has also been used for moving object classification. The distance of the center of the object to every point of the contour is calculated and smoothed, and the coefficients obtained from their discrete Fourier transform are used for the feature vector, and so is the aspect-ratio. The feature vector in its turn is used to construct a Self-cOnstructing Neural Fuzzy Inference Network (SONFIN) used for object recognition.

In another approach [Asp14a] the object was divided into ringlets, each one contributing to the creation of a histogram. The value of each pixel was weighted according to a Gaussian distribution of the distance from the ringlet. This feature turned out to be rotation invariant and centered weighted, since the significance of the ringlets closer to the center are more important than the outer ones in some cases.

All said, there seems to be a difficulty in the discrimination between humans and motorcycles / bicycles using shape descriptors, because the top half of both classes is identical, and many times in infrared imagery there is a partial detection. There also seems to be an inadequacy in the decision of the use of shape and/or texture descriptors robustly. This would provide better classification results, aiding the surveilling purpose, by allotting more resources to the surveillance of particular classes, which are considered a higher threat. For example, a human poses more of a threat than a dog in an outdoor surveillance system.

3. METHODOLOGY

In an effort to better distinguish humans from motorcycles, a new algorithm for the classification of moving objects in infrared imagery into multiple classes is proposed, using both textural and shape descriptors, along with the velocity of the object. The result of the algorithm from every frame of the sequence must be used with a tracker in a voting system, allowing for greater accuracy. An object detection algorithm based on ViBE [Bar11a] [Van12a] and a moving object tracker based on the work proposed on [Tor12a] have been implemented for the first stage of surveillance, prior to classification, but their results are not being discussed, as their interest lies out of the purpose of this work. Every ROI from the segmentation is tracked even when it stops for a while. When two or more objects enter the scene they are tracked and classified separately, except the case when they come extremely close. In this case classification stops, and continues only if the objects separate again. When this happens, another texture based classification takes place, in order to give them the right IDs, the ones that were provided before the merge. After this, object classification continues normally.

3.1 Feature Extraction

The features selected for the classification process regard the following categories: Shape descriptors, Texture descriptors and Velocity. The features were selected according to their ability to describe objects of the same class (small intraclass variation) and at the same time differentiate objects from different classes (large interclass variation). A 3-D representation of the dispersion of the training set after multidimensional scaling (MDS) can be seen in Figure 1 and Figure 2. These features were selected after testing the discriminative power of many other features, and particularly horizontal and vertical projections [Gur11a], Gaussian Ringlet Intensity Distribution features [Asp14a], aspect ratio, fill ratio, uniformity, skewness, smoothness, compactness, Hu's moments [Hu62a], Local Binary Patterns (LBP) [Oja02a], Histogram of Oriented Gradients (HOG) [Tsa10a] and Local Phase Quantization (LPQ) [Jia14a].

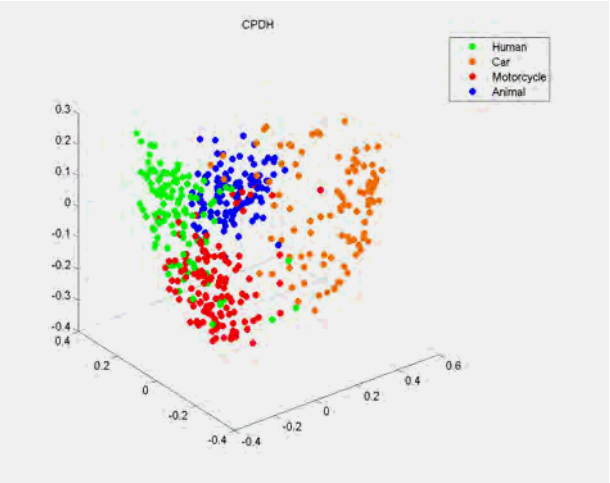

Figure 1. Dispersion of the four classes after MDS using the CPDH shape descriptor

3.1.1 Shape Descriptors

The description of the shape of the object is performed using the Contour Points Distribution Histogram (CPDH) algorithm [Gur11a], which is implemented by measuring the distribution of the contour points in a predefined topography.

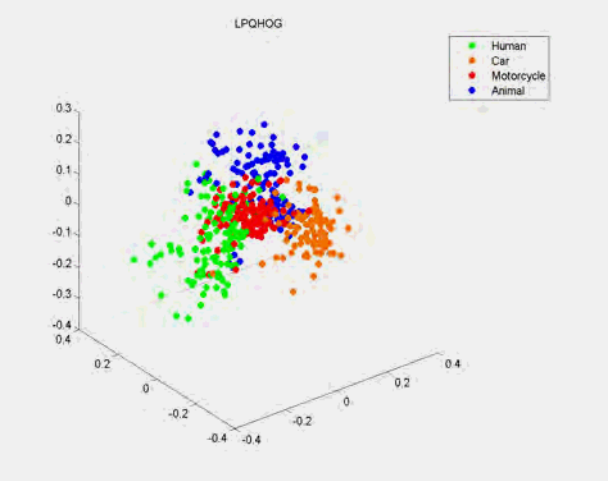

Figure 2. Dispersion of the four classes after MDS using the LPQHOG texture descriptor

The object is divided into 36 segments – 3 zones around the object by twelve 30 degree sectors. By counting the points of the contour that belong to each segment, the histogram of the distribution is produced (Figure 3). This histogram is used as a shape descriptor feature. Its discriminative ability is shown in the distance matrix of the training set in Figure 5, where the 4 classes of 130 objects each, are distinguished adequately. Blue represents zero (0) distance that goes towards red which represents one The main diagonal is of course blue, which means that every histogram has zero distance from itself. The distance between objects of the same class is small (blue), while it is large (red) between objects of different classes. The similarity between the first and third class (human / motorcycle) can be distinguished using the velocity feature. (1).

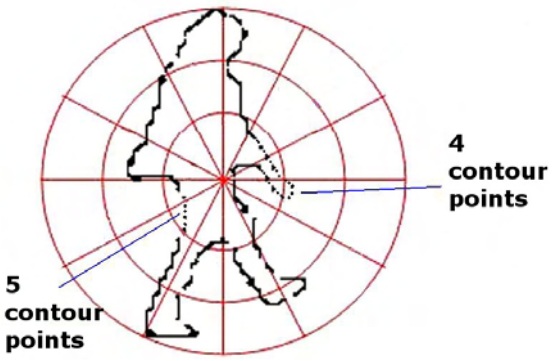

Figure 4. Contour Points Distribution into geometric segments.

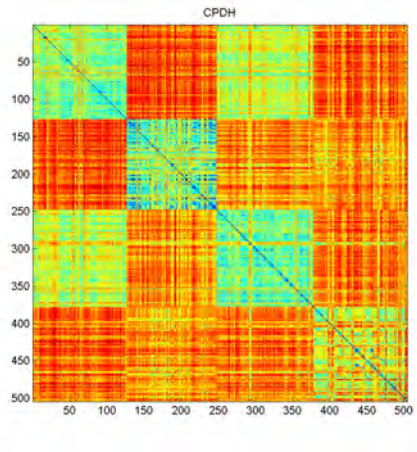

Figure 5. Distance matrix of training data with CPDH histograms

3.1.2 Texture Descriptors

The Local Phase Quantization histogram concatenated with the Histogram of Oriented Gradients first introduced in [Rog15a] was used as a texture descriptor. The gradient based properties of the HOG descriptor are enriched with information derived from the frequency domain of the LPQ descriptor with better results than each descriptor on its own.

The LPQ descriptor is performed in local areas where quantization of the Fourier transform phase takes place. Phase information is extracted using a 2-D Discrete Fourier Transform extracted from a rectangular N-by-N neighborhood N_x on every pixel x of the image $f(x)$ defined by

$$F(\mathbf{u},\mathbf{x}) = \sum_{\mathbf{y}\in N_x} f(\mathbf{x}-\mathbf{y}) e^{-j2\pi \mathbf{u}^{\mathrm{T}}\mathbf{y}} = \mathbf{w}_{\mathbf{u}}^{\mathrm{T}}\mathbf{f}_{\mathbf{x}} \qquad (1)$$

where $\mathbf{w_u}$ is the basis vector of the 2-D DFT at frequency $\mathbf{u}$, $\mathbf{f_x}$ is the vector containing all N^2 samples from N_x, and j the imaginary unit (j^2=-1). The local Fourier coefficients are computed at four frequency points: $u_1 = [a, 0]^T$, $u_2 = [0, a]^T$, $u^3 = [a, a]^T$, *and* $u_4 = [a, -a]^T$, where a is a sufficiently small scalar (a = 1/7 in our implementation), and T the transpose of a vector, which is denoted by a bold symbol.

The histogram of oriented gradients is a powerful texture descriptor created by calculating the gradient values of the pixel's intensity using a discrete derivative mask in horizontal and vertical direction, or even the sum of the vectors in all directions (Figure 5). The vector is the difference in intensity and direction between the central pixel and its neighbor. As such, the feature captures the

distribution of intensity gradients within the object of interest and expresses it by a single histogram.

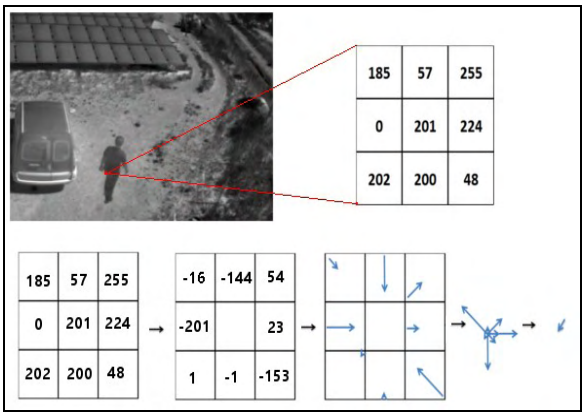

Figure 5. Gradient extraction of a pixel

The discriminative power of the texture descriptor is shown in Figure 6. In general, texture variation is limited in infrared imagery in comparison to RGB images, which explains why the shape descriptor yields better results. But it will be shown that the use of both texture and shape descriptors yields better results.

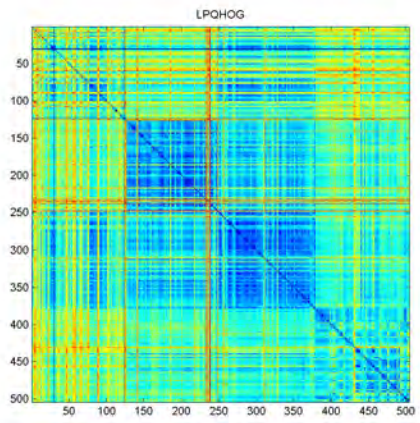

Figure 6. Distance matrix of training data with LPQHOG histograms

3.1.3 Velocity

In order to distinguish between humans and motorcycles and between cars and animals the velocity of the object can be used as a feature. For a mounted camera on a pole the objects appear larger and moving faster as they approach, while they appear smaller and moving slower when they are far from the camera. It was found that the ratio of the distance in pixels per frame over the height of the object is proportional to the actual speed of the object, so it was used as a feature. The velocity is calculated by dividing the Euclidean distance of the center of mass of the object between n frames by the height of the object:

$$u_i = \frac{\|C_i - C_{i-n}\|}{h_i} \quad (2)$$

u_i is the velocity of the object at frame i, C_i is the center of mass of the moving object at frame i, and h_i is the height of the object at frame i.

Figure 7. Examples of the templates of the training database

3.2 Training – Database description

The training dataset (Figure 7) was collected during a span of more than half a year so that images from very different environmental conditions would be stored, using a stationary long-wave infrared camera providing 320X256 pixel images. During the data collection, about 31,600 images of humans were obtained, 15,400 car images, 6,400 motorcycle images and 4,400 images of dogs. The images were obtained during day and night, during sunny, cloudy and rainy days, in hot and cold weather. Different types of cars and motorcycles were used during the course of collecting the dataset. Images of humans were captured in various poses, and wearing a variety of clothes. The predescribed features were extracted from all the obtained images and two SVM classifiers were created. Finally, the training set was refined, selecting one hundred and thirty images from each class as training data, based on the distance of their features in the SVM space, both textural and shape based.

Another database was recorded for the experimental evaluation of the proposed algorithm, called testing dataset. The following scenarios were repeated in four different weather conditions, i.e. early in the morning, on a sunny afternoon, at night and on a rainy morning.

1. Human walks and runs
2. Car drives
3. Motorcycle drives
4. Animal wanders

About 15,000 images were recorded under all weather conditions, with a total of 62,784 images.

3.3 Classification

From every object that is recognized as foreground in a frame, the shape descriptor and the velocity described above are extracted and projected on the feature space of the SVM. The velocity is the last dimension of the SVM hyperspace. The distance to the nearest support vector of the shape descriptor classifier is calculated for every class. After that, the image's median is calculated to decide whether its textural data will help discriminate its class. The optimal value (med_o) was determined by testing the accuracy of the algorithm on a wide range of values, using only images from the testing dataset that were pertinent to the experiment, i.e. that had small textural discrimination (Figure 9). For images with enough textural data the texture descriptor described above is extracted and projected on the feature space of the SVM. (Figure 8)

Therefore, four or eight distances are calculated for every object: four are the distances of the object to each class' boundary in the SVM hyperspace of the shape descriptors and the velocity and four are the corresponding distances of the texture descriptor. All distances are normalized with the minimum and maximum value of the corresponding training set, i.e. the distance of an object to the human shape descriptor SVM is normalized using the distances of all training data to the same SVM. These distances are considered as confidence values of the classifier: the greater the distance from the border of the two classes, the greater the confidence. The use of these confidence values was validated by running the entire testing dataset over a wide range of confidence level values (i.e. from zero to one, using a 0.05 step), and discarding the votes that had lower confidence than the value (Figure 10). The diagram is indicative of the intrinsic capability of the proposed algorithm to cope with results of low confidence, maintaining in this manner the overall accuracy of the algorithm at 0.83 by avoiding false negatives. Furthermore, a decrease in accuracy is only visible after the exclusion of solutions with confidence above 0.4 which affects the number of the false positives. So if an object that is very different from the four classes enters the scene, it will remain unclassified, as it's confidence value will be low.

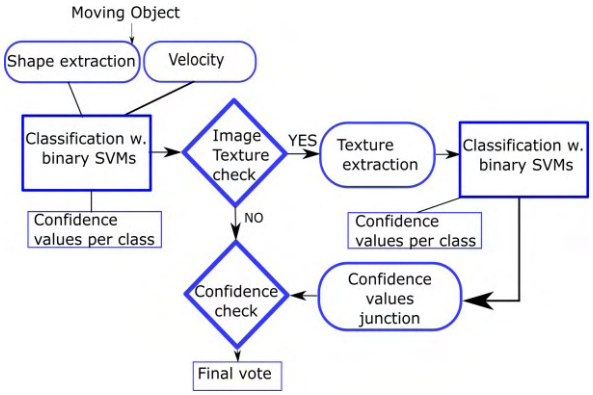

Figure 8. Classification procedure-block diagram

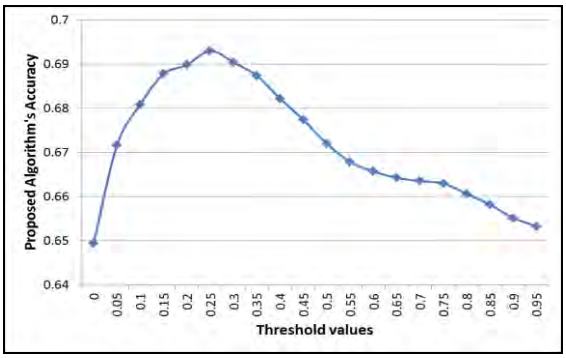

Figure 9. The accuracy of the algorithm over a range of texture discrimination values applied on a portion of the dataset with low texture images

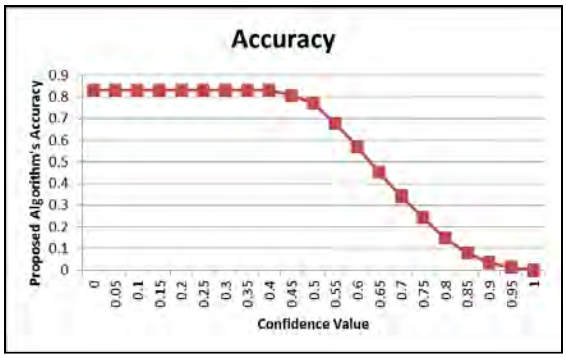

Figure 10. The accuracy of the algorithm over a range of confidence values applied on the whole dataset

In case the object's texture does not have enough discriminative power, which usually happens in infrared imagery when the object emits significantly more radiation than its background, only the shape descriptors and the velocity are used to classify the object. In all other cases, both shape/velocity and texture are used for the classification by fusing the results of the classifiers, by adding the normalized confidence value of each decision. The confidence

that the object belongs to the returned class is given by the formula:

$$C_{class} = \begin{cases} \max\begin{cases} C_{SH} \\ C_{SC} \\ C_{SM} \\ C_{SA} \end{cases}, & if\,(med > med_o) \\ \max\begin{cases} \frac{C_{SH}+C_{TH}}{2} \\ \frac{C_{SC}+C_{TC}}{2} \\ \frac{C_{SM}+C_{TM}}{2} \\ \frac{C_{SA}+C_{TA}}{2} \end{cases}, & else \end{cases} \quad (3)$$

where C_{XY} is the normalized distance of the X descriptor (Shape/Velocity or Texture) to the Y SVM (Human, Car, Motorcycle, Animal), *med* the median of the object and med_o the optimum value of the median, that was defined experimentally, as mentioned earlier.

The object is tracked in time and a vote is added to the class it is classified at every frame, if the confidence is above the confidence value. The object is finally classified into the class that has the most votes. For the purposes of this paper, and in order to address the possibility of inputs outside the spectrum of the datasets that were used for training and testing, a threshold of 0.25 was selected.

Figure 11. A human that can be discriminated through the use of texture descriptors

Figure 12. A dog in challenging condition where it fails to be distinguished from its surrounding background street

Figure 13. A human emitting significantly more radiation than the background

4. EVALUATION RESULTS - DISCUSSION

As was expected, using only shape descriptors did not yield encouraging results, especially in classifying humans from motorcycles, though it recognized cars and motorcycles adequately. (Table 1) This was a driving factor to the proposed work, which included textural descriptors, temporal features (velocity) and a decision on the use of textural features that improved significantly the classification results.

The proposed algorithm works well when the object's shape is well defined and there is textural discrimination, i.e. when the background emits more or less radiation than the object, but not excessively (Figure 11). When they emit the same IR radiation it is very difficult to differentiate between background and foreground (Figure 12). In case the object emits much more radiation than the background, only shape descriptors are used, lowering the accuracy of the decision (Figure 13).

The algorithm's accuracy has been estimated over the demanding testing dataset, and has an overall

accuracy of 83%. The precision and recall of the classifier on different weather sets are shown in Figure 14 and Figure 15. Unfortunately there was no data with animals during the rainy morning session. The precision for the Motorcycle class is low because in some cases it was mistakenly classified as human. This happened especially when the road was as hot as the motor, so the only part left distinguishable was the rider.

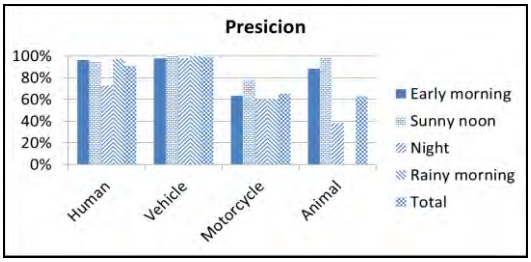

Figure 14. Precision per class

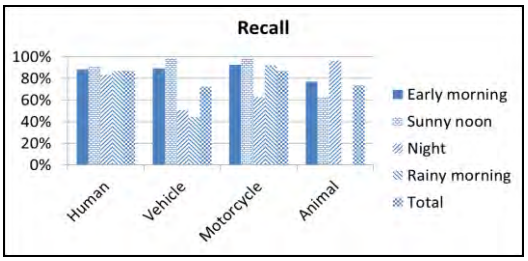

Figure 15. Recall per class

The descriptors were tested on the same data base (all weather conditions) using the Random Forest Classifier [Liv05a] (Table 2) instead of the proposed classifier (Table 3), but the results imply that the proposed classifier works better. The Random Forest Classifier scored 78.2% on accuracy. It was utilized in both feature sets, i.e. texture and shape providing two votes, although in some occasions, when the textural discrimination was low, it returned only one vote, based solely on the shape. The proposed algorithm does not return any votes if the confidence of the two binary classifiers is low.

The implemented method for benchmarking was found in the bibliography and was used to validate the proposed method using the same data set. In the work of [Wan10a] a Shape Context Descriptor is proposed where the log-polar histogram of the shape is exported in 5 bins for *log r* (radius) and 12 bins for θ (angle). Then a Shape context based Adaboost cascade classifier is used to classify the shape. The average time for a frame to be processed completely (segmentation, classification, tracking) using the proposed method is 46.9msec, while the benchmarking method needs 33.4msec. The times were computed on a Intel Core i7-4790K processor at 4.00GHz with 8GB RAM. The reason for this difference is that the proposed method performs almost the same tasks as the benchmark method, plus the texture feature extraction.

Wang's Algorithm		Actual classes: Human	Car	Moto	Animal
Predicted classes (votes)	Human	**299**	9	117	57
	Car	108	**1417**	44	146
	Moto	2486	30	**943**	16
	Animal	208	141	44	**166**
				Overall Accuracy	**45.3%**

Table 1. Benchmarking algorithm's accuracy

Proposed Algorithm		Actual classes: Human	Car	Moto	Animal
Predicted classes (votes)	Human	**2633**	140	132	10
	Car	0	**811**	11	0
	Moto	389	36	**954**	85
	Animal	19	138	0	**266**
				Overall Accuracy	**83%**

Table 2. Proposed algorithm's accuracy

Random Forest Classifier		Actual classes: Human	Car	Moto	Animal
Predicted classes (votes)	Human	**2047**	4	126	7
	Car	22	**1071**	5	5
	Moto	934	31	**964**	33
	Animal	38	19	2	**316**
				Overall Accuracy	**78.2%**

Table 3. Accuracy of proposed method using random forest classifiers instead of SVMs

5. CONCLUSIONS

This paper introduced a new approach in multi-class object classification, using confidence values in each decision, and deciding whether to use textural

features along with a shape descriptor and velocity on each frame. In the highly demanding field of infrared thermography, there must be enough flexibility to choose the best features to use for classification, either texture and shape, or only shape. The algorithm reached an overall accuracy of 83%, but can climb up to 91% under certain climate conditions (e.g. a sunny afternoon). The structure of the approach allows for many improvements in future works: from choosing different descriptors to replacing the binary classifiers with other methods. The field is open for more research especially regarding the extraction of features from objects that are partially detected due to heavy occlusion. The size of the testing and training dataset may also grow in future work to allow for greater validity of the results and also for more discussion.

6. ACKNOWLEDGEMENTS

This work was partially supported by the EU funded PREACT Capability Project (CP) (FP7-607881).

7. REFERENCES

[Alp12a] Alpert, Sharon, et al. "Image segmentation by probabilistic bottom-up aggregation and cue integration", IEEE Transactions on Pattern Analysis and Machine Intelligence, 34.2, p.315-327, (2012).

[Arb11a] Arbelaez, Pablo, et al. "Contour detection and hierarchical image segmentation", IEEE Transactions on Pattern Analysis and Machine Intelligence, 33.5, p.898-916, (2011).

[Asp14a] Aspiras, Theus H., Vijayan K. Asari, and Juan Vasquez. "Gaussian ringlet intensity distribution (GRID) features for rotation-invariant object detection in wide area motion imagery", IEEE International Conference on Image Processing (ICIP), p.2309-2313, (2014).

[Bab11a] Babenko, B., Ming-Hsuan Y., and Belongie, S., "Robust object tracking with online multiple instance learning", IEEE Transactions on Pattern Analysis and Machine Intelligence 33.8, p.1619-1632, (2011).

[Bar11a] Barnich, Olivier, and Marc Van Droogenbroeck. "ViBe: A universal background subtraction algorithm for video sequences", IEEE Transactions on Image Processing, 20.6, p.1709-1724, (2011).

[Che12a] Chen, Eli, Oren Haik, and Yitzhak Yitzhaky. "Classification of moving objects in atmospherically degraded video", Proceedings of SPIE-The International Society for Optical Engineering, 51.10, p.1-14, (2012).

[Gur11a] Gurwicz, Yaniv, Raanan Yehezkel, and Boaz Lachover. "Multiclass object classification for real-time video surveillance systems", Elsevier, Pattern Recognition Letters 32.6, 805-815, (2011).

[Hu62a] Hu, Ming-Kuei. "Visual pattern recognition by moment invariants", IRE Transactions on Information Theory, 8.2, p179-187, (1962).

[Hua10a] Huang, Zhihu, and Jinsong Leng. "Analysis of Hu's moment invariants on image scaling and rotation", IEEE 2nd International Conference on Computer Engineering and Technology (ICCET), 7, p.476-480, (2010).

[Jia14a] Jiang, Bihan, et al. "A dynamic appearance descriptor approach to facial actions temporal modeling", IEEE Transactions on Cybernetics, 44.2, p.161-174, (2014).

[Jo13a] Jo, Ahra, et al. "Performance improvement of human detection using thermal imaging cameras based on mahalanobis distance and edge orientation histogram", Information Technology Convergence, Springer Netherlands, Lecture Notes in Electrical Engineering, 253, p.817-825, (2013).

[Joh15a] John, Vijay, et al. "Pedestrian detection in thermal images using adaptive fuzzy C-means clustering and convolutional neural networks", IEEE International Conference on Machine Vision Applications (MVA), 14th IAPR, p246-249, (2015).

[Jua08a] Juang, Chia-Feng, and Liang-Tso Chen. "Moving object recognition by a shape-based neural fuzzy network", International Conference on Artificial Neural Networks (ICANN 2006) / International Conference on Engineering of Intelligent Systems (ICEIS 2006), Neurocomputing 71.13, p2937-2949, (2008).

[Ku10a] Ku, Zhi Kai, Chee Fei Ng, and Siak Wang Khor. "Shape based recognition and classification for common objects-An application in video scene analysis", IEEE 2nd International Conference on Computer Engineering and Technology (ICCET), 3, p13-16, (2010).

[Lia15a] Liang, Chung-Wei, and Chia-Feng Juang. "Moving object classification using local shape and HOG features in wavelet-transformed space with hierarchical SVM classifiers", Applied Soft Computing 28, p483-497, (2015).

[Liv05a] Livingston, Frederick. "Implementation of Breiman's random forest machine learning algorithm", ECE591Q Machine Learning Journal Paper (2005).

[Oja02a] Ojala, Timo, Matti Pietikainen, and Topi Maenpaa. "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns", IEEE Transactions on Pattern

Analysis and Machine Intelligence, 24.7, p971-987, (2002).

[Olm12a] Olmeda, Daniel, Arturo de la Escalera, and Jose Maria Armingol. "Contrast invariant features for human detection in far infrared images", IEEE Intelligent Vehicles Symposium (IV), p117-122, (2012).

[Rog15a] Rogotis, Savvas, et al. "Recognizing suspicious activities in infrared imagery using appearance-based features and the theory of hidden conditional random fields for outdoor perimeter surveillance", Journal of Electronic Imaging 24.6, p.1-10, (2015).

[Shu11a] Shu, Xin, and Xiao-Jun Wu. "A novel contour descriptor for 2D shape matching and its application to image retrieval", Image and Vision Computing 29.4, 286-294, (2011).

[Tor12a] Torabi, Atousa, Guillaume Massé, and Guillaume-Alexandre Bilodeau. "An iterative integrated framework for thermal–visible image registration, sensor fusion, and people tracking for video surveillance applications", Computer Vision and Image Understanding 116.2, 210-221, (2012).

[Tsa10a] Tsai, Grace. "Histogram of oriented gradients", University of Michigan, (2010).

[Van12a] Van Droogenbroeck, Marc, and Olivier Paquot. "Background subtraction: Experiments and improvements for ViBe", IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), p32-37, (2012).

[Wan10a]Wang, Weihong, Jian Zhang, and Chunhua Shen. "Improved human detection and classification in thermal images", IEEE 17th International Conference on Image Processing (ICIP), p2313-2316, (2010).

[Wan15a] Wang, Xin, Chen Ning, and Lizhong Xu. "Spatiotemporal saliency model for small moving object detection in infrared videos", Elsevier, Infrared Physics & Technology, 69, p.111-117, (2015).