Sebastian BUDZAN
SILESIAN UNIVERSITY OF TECHNOLOGY, INSTITUTE OF AUTOMATIC CONTROL
Akademicka 16, 44-100 Gliwice, Poland

Human Detection in Thermal Images Using Low-level Features

Abstract

In this work the human detection method in infrared has been presented. The proposed solution focuses on the use low-level features and detecting parts of the human body. Low–level processing is based on modified HOG (Histogram of Oriented Gradients) algorithm. First, the only squared cells have been used, also calculation of the gradient has been improved. Next, the model of the head from the dataset IR (Infra Red) images has been created, also the model of the human body. Finally, the probability matrix has been examined using minimal distance classifier. The novelty of the proposed solution focuses on the combination of the pixel-gradient and body parts processing, also three stage classification process (head modelling, human modelling and classifier), which has been proposed to reduce the false detection. The experiments were performed on self-created IR images database, which contains images with most of the possible difficult situations such as overlapped people, different pose, small and high resolution of the people. The performance of the proposed algorithm was evaluated using Precision and Recall quality measure.

Keywords: human detection, infrared, feature extraction, HOG.

1. Introduction

Human recognition, classification, or generally detection on thermal images is a growing part of computer vision. In practice human detection is a difficult task due to many physical parameters that can impact the final result of image processing. The detection is dependent on the human pose, clothes that it wears, type of hair, effort, the environment in which a human being is present. Different types of the cameras, angle of the camera, distance to the human, changing background. In recent years many of the solutions have been proposed, based on shape, geometric matching, skin color, classification of the body parts. There is no universal method for human detection, all existing methods have been adjusted to a specific situations,

Thermal images have several differences regarding to the visual camera images such as they include the distribution of temperatures in taken picture, what is advantage from one hand, because we get a new quantity that can be used to detect human, but the other hand thermal images include some objects with temperature similar to the human, in consequence the method of human detection should to base on thermal information, some region properties, also shape of the detected objects.

Process of the human detection can be described by a few general steps [1]. First, after the image acquiring, the preprocessing methods have been performed such as filtering, some edge detection or morphology operations. Next, the features selection process produces number of the objects in the image. Finally, used classifier decide which of the detected objects can be categorized as human. Human detection methods in many solutions have been combined with other applications, in consequence can obtain information apparently not connected with human detection [2,3] such as height, haircut, walking style, walking alone or in group of people. Nowadays, human detection algorithms find applications in many fields, for example science simulations, sport activities, automotive industry, rehabilitation, video games (MS Kinect), for intelligent cars to prevent accidents, biometrical identification systems, also give us many useful information [4,5,6,7], such as identification of human position, state, motion direction, gesture recognition, 3D modelling of the human body.

2. Related work

Most of the described in the literature human detection methods are developed to processing images taken by cameras working in visible range. Although, they may be used on the images acquired with other method such as infrared cameras and TOF (Time Of Flight) cameras, where the image to consist of time of flight of a light signal between the camera and the object.

Generally, methods of human detection can be categorized into a two main groups, such as single images processing and sequence of images. These methods are based on low-level features, shapelet features, detection of the body parts, distance transform, gait analysis, background modelling, Haar-like features.

One of the first and most popular methods of finding people on images was proposed by N. Dalal and B. Triggs [8]. Their algorithm focuses on low level features described as HOG (Histogram of Oriented Gradients) which uses the information on the intensity of the gradient for the featured nine directions (0 – 180 degrees). It is possible to analyze the share of the pixels in the certain cell on the defined gradient direction. Image for the analysis is searched by the detection window of the size 64×128 pixels, which is divided into square blocks with size 16 pixels in the next step. Finally, they are divided into 4 cells in system 2×2 with size 8×8 pixels. For each cell (64 pixels) have 9 features, which with the features calculated in the other cells create input of the SVM (Support Vector Machine) classifier. Another method which based on HOG algorithm was developed by T. Watanabe, S. Ito and K. Yokoi [9]. In this paper authors proposes a method for extracting feature descriptors consisting of co-occurrence histograms of oriented gradients (CoHOG). Method is based on the same calculation the share of the pixels like in HOG, but in this case eight directions (0 – 315 degree). Including co-occurrence with various positional offsets, the feature descriptors can express complex shapes of objects with local and global distributions of gradient orientations. The results showed that proposed method reduces miss rate by half compared with HOG, also increases accuracy of the human shape mapping. Authors of [10] proposes in the paper integration the cascade-of-rejectors approach with the Histograms of Oriented Gradients features to achieve a fast and accurate human detection system. The features used in their system are HoGs of variable-size blocks that capture salient features of humans automatically. Using AdaBoost for feature selection, they identify the appropriate set of blocks, from a large set of possible blocks. Integral image representation and a rejection cascade which significantly speed up the computation is being used with accuracy level similar to other methods. P. Sabzmeydani i G. Mori proposed in [11] algorithm of detecting some local regions of interests, known as shapelet features. First, the input image should be filtered using squared structural element, which build low-level gradient information. Using AdaBoost, these shapelet features are created as a combination of oriented gradient responses.

There exists many algorithms that work in a different kind of way, mainly there are based on detecting certain body. The main problem in this approach is that detection of individual body parts. Because many of the proposed algorithm uses combination of the detected body parts, the final results of the human classification step can be highly affected by the wrong detection of the one body part. Review of the existing literature in this area shows, that the body part which is the most loaded with wrong identification are human legs. Human detection based on a probabilistic assembly of robust part detectors has been presented in [12]. Developed method allows for human detection by searching their certain body parts such as head with shoulders, face seen from profile or 16 front, legs and creating the model of human being. First, the detector has been developed, using local features. The features are local orientations of gradient and Laplacian based filters. Furthermore, the features with the highest occurrence and co-occurrence probabilities are learnt using AdaBoost. The resulting part detector gives face detection results. Next, the human detection results are improved by computing a likelihood score for the assembly of body parts. S. Ioffe and D. Forsyth proposed in [13] algorithm based on finding people by finding candidate body segments, and then constructing assemblies of segments that are consistent with the constraints on the appearance of a person that result from kinematic properties. Algorithm does not require all body parts correctly detected, because there are many possible combinations of the human body shapes. Segmentation of the image is based on searching rectangle blocks using parallel edge detection. In consequence, in this method the head is omitted. Each detected block is described by the center, width, height, also the orientation. Next, based on the correlation between adjacent blocks the human being is calculated. Finally, we get the information about existence of human in the image. Moreover, position of the individual body parts can be determined.

An approach based on gait analysis has been presented in [14]. The proposed gait recognition system characterizes gait in terms of a gait signature computed directly from the sequence of silhouettes. The system can be seen as a generic pattern recognizer composed of a few main modules: human detection and tracking, training using Modified Independent Component Analysis (MICA) and finally, human recognition. Initially, the background modelling is done. Next, the moving objects are segmented using the background subtraction algorithm. Then, the morphological skeleton operator is used to track the moving silhouettes of a walking figure. Finally, the person's identity is determined by training and testing using MICA on the extracted feature vectors.

Human detection by infrared camera installed in a car as a part of a driver assistance system has been presented in [15] Presented pedestrian detection method is employing far infrared images. It is based mainly on localization of warm symmetrical objects with specific aspect ratio and size. Because the road infrastructures and other road participants may also have similar characteristics, the filtering process to avoid a number of false positives is performed. Finally, validation procedure based on human morphological (shape) and thermal characteristics. In [16] authors proposed algorithm based on a Shape Context Descriptor (SCD) with the Adaboost cascade classifier framework. Taking into account the fact that appearance based features may not be suitable in thermal images, because the size of the target is relatively small, pixel intensities in human bodies have considerable variations under different weather conditions, there is lack of texture information, they evaluated two different image features, i.e. sobel filter with Adaboost cascade and shape context was chosen. A real time human detection system based on far infrared vision has been described in [17]. Human being detection and tracking algorithm was created on the basis of using infrared vision in order to have reliable information on a room occupation. A foreground segmentation with a Gaussian background model has been performed in the first step. Next, tracking step based on connected components intersections is performed. A classification based on a cascade of boosted classifiers is used for the recognition. This step is based on Haar wavelets and the boosting algorithm Adaboost. An alternative night-time pedestrian detection method using far-infrared camera has been presented in [18]. Proposed method consists of two main steps, regions of interest generation and pedestrian recognition in a cascade fashion. Authors proposed pixel-gradient oriented vertical projection to estimate the input image and reduce its searching regions efficiently.

3. Proposed method

The implemented algorithm is supposed to detect human beings on visible and thermal images and the position on the thermal image. The algorithm is based on single images, in which human recognition is the most difficult form of detection, because its greatest disadvantage is the single information about the whole environment, using only one image the human being should be confirmed. Especially in thermal imagery some of the proposed methods does not work properly such as background modelling from one thermal image. Recognizing people in the images includes one of the most common problem that is the variation of the human position. With a few images, the probability that the position of the observed person will be impossible to detect decreases with increasing amounts of testing set of images. Thus, mostly the effectiveness of the algorithm depends on the set of the images used to train the algorithm.

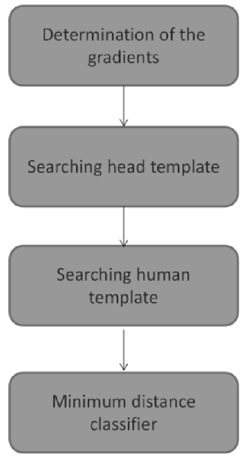

Fig. 1. Block diagram of the proposed solution

Proposed method (see Fig.1) is based on low-level features with shape of the human head modelling. Taking into account thousands of the human thermal images, head is one of the distinctive part of the human body, more effective than in visual images and more quite human position independent. First, the gradient for all of the pixels in the image is determined, what helps to find similar regions in the image – similar temperature values and direction of the temperature changes. Gradient of the one pixel can be calculated using equation:

$$\nabla G(x,y) = \left[\left(\frac{\delta f(x,y)}{\delta x}\right)^2 + \left(\frac{\delta f(x,y)}{\delta y}\right)^2\right]^{\frac{1}{2}}. \quad (1)$$

Main difference regarding to the e.g. Dalal and Triggs HOG algorithm [8] is calculation gradients in the image in the same range 0 - 180°, but with 1° step, what makes the algorithm more sensitive to the differences in the human position. In consequence, there is more possible directions distinguished, also size of the cells can be changed to the minimal squared 2×2 pixels. Next, the searching the image for model human head is performed. The head model is created using averaging set of images containing only human head. The model is recalculated from the temperature values to the gradient values, what gives opportunity to the changing of the model scale (size). In this step all of the pixels in the image are examined with above model in checked cells. At the end of this step each pixel in the image is described by the probability of head region membership. Changing the cell size, will change also the sensitivity of the algorithm.

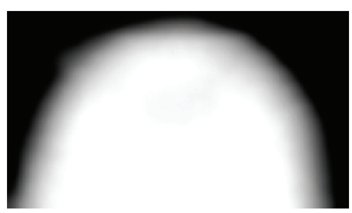

Fig. 2. Thermal image of the human head as averaging set of head thermal images

Previous step produces candidates for the human being, but it should be complemented by general shape of the human in the regions, where head was detected with high probability. This step depends on size of the "learning" set of images – increasing human poses increases effectiveness of the human body detection. Using morphological operations, there have been created n-dimensions spaces, which determine similarity to the human pattern. At the output we get matrix of the probabilities of human being, sized exactly as input thermal image. Detection of the head has been performed with high accuracy, thus, detection of the human being can be done with lower resolution – greater size of the cells.

Calculated matrix of the probabilities, which contains regions with highest probability of human being is examined with minimal distance classifier. The classifier assign the detected object to the one of the predefined classes using information about distance between two sets of variables. Proposed classifier uses only two classes: human and non-human. Each of the image in the training set has been divided into cells, then gradient has been calculated in each cell, which in basic form has the same weight. If $\boldsymbol{X}$ is a set of detected object features and $\boldsymbol{Y}$ is a set of human being model features, assignment $\boldsymbol{X}$ to the class k is determined as follows:

$$D\left(X, Y^{(k)}\right) = \min\left\{D\left(X, Y^{(i)}\right), i = 1, \dots, n\right\}, \quad (2)$$

where:

D – distance between X and Y calculated according to (3),
k – number of the predefined class,
n – number of the all defined classes (n=2).

$$D(X, Y) = \left[(X - Y)^T (X - Y)\right]^{\frac{1}{2}}. \quad (3)$$

4. Results and discussion

Thermal images for the experiments presented in this paper have been captured with a Wuhan-Guide TP8 IR camera. The camera is equipped with a 384 – 288 pixel uncooled FPA microbolometer. Its spectral range is 8 – 14 µm and thermal sensitivity (resolution) 0.08°C. About 100 examined thermal images have been acquired in outdoor and indoor conditions, also all images have been divided into four main groups, such as: humans registered from short and long distance in group of maximum three people, humans registered from short and long distance in group of minimum four people. In the experiments have been used about 100 thermal images with and without human being. The head and human being model have been determined based on this images set. Quality of the detection results has been evaluated using two quality measures, called Precision and Recall. They are defined [19] as follows:

$$Precision = \frac{TP}{TP+FP}, \quad (4)$$

$$Recall = \frac{TP}{TP+FN}, \quad (5)$$

where:

FP – False Positive,
FN – False Negative,
TP – True Positive.

Precision and *Recall* are expressed in percentage and define the measure the effectiveness of the algorithm. *Precision* give the information how many of the classified images as containing the human, really at least one human has been detected. *Recall* defined how many of the tested images contained at least one human have been classified as human class.

Gradient in the examined image is calculated in the cells, thus, size of the cell directly affects the final result, also the gradient threshold, known in the HOG literature as hit threshold also should be carefully selected. Figure 3 shows original thermal image with group of people and results of detecting with wrong selected cell size and hit threshold. The parameters should be evaluated only once for the dataset of the images, in result both parameters are used during head, human and classification step processing. Each time in proposed algorithm image has been divided into squared cells.

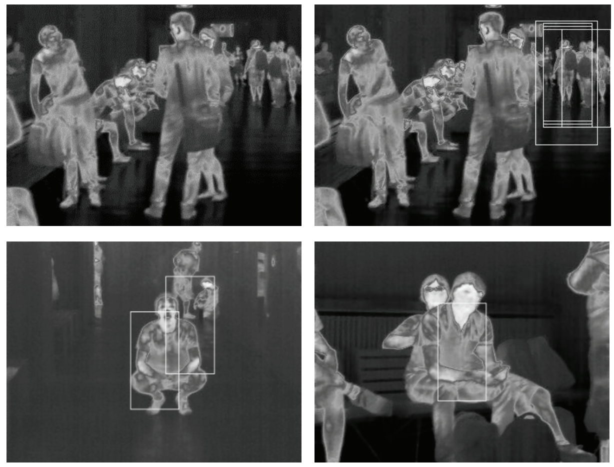

Fig. 3. Results of the sample thermal images processing with a group of people

As is presented in Figure 4 first results significantly depends on cell size, e.g. for cell size 2×2 pixels correct classification produce poor results, because it's to detailed and then many cells in the image can be categorized as human being, also other hand, greater value of the cell size will produce quite good results, especially in correct classified people, but poor, when we take into account correct rejected human candidates. The results of the classification at this level can be improved during further processing.

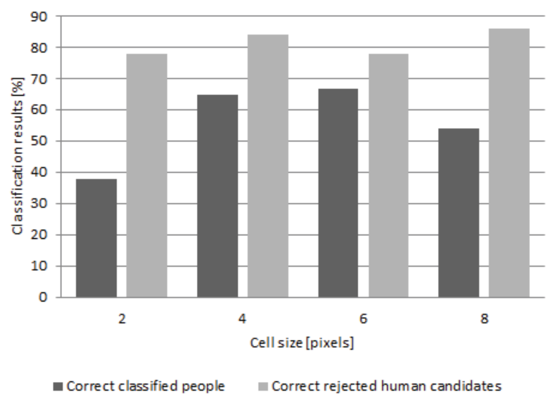

Fig. 4. Relation between quality of the detection and cell size

In the Figure 5 sample thermal images with correct detections has been presented. Good results are obtained especially for those images, where entire people are visible, even they are in back or side position. It's a consequence of proposed solution, where head detection is the first step and in the second the human being must be confirmed by other human parts existence. In this case algorithm produces good results for long and short distance. Left-bottom image shows main advantage of the algorithm, detection of the human even in the three different poses and in different distance/resolution on the image at one time. Right-bottom image shows detection results on overlapped two people, who are detected as one person.

Fig. 5. Samples of correct detection in different people pose and distance

Results presented in Figure 6 showed *Precision* and *Recall* results compared to the Benezeth Y. et al. algorithm [17]. Quite better results are obtained for short distance detection, because resolution of the people are greater. It should be emphasized, that the datasets used to build the head and human models contains images in short and long distance.

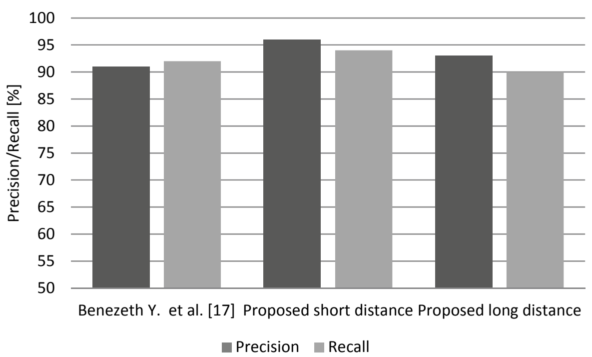

Fig. 6. *Precision* and *Recall* values for Benezeth Y. et al. and proposed algorithm

5. Conclusions

This paper presents a people detection method for infrared. The major contributions are: (1) the combination of the pixel-gradient and body parts processing, (2) three stage classification process (head modelling, human modelling and classifier) is proposed to reduce the false detection. The presented algorithm has been tested on the real thermal images taken in real environment. Also some disadvantages have been identified, such as problem with detection groups of the overlapped people, also some poses not included in the dataset.

6. References

[1] Chaquet J.M., Carmona E.J., Fernandez-Caballero A.: A survey of video datasets for human action and activity recognition. Computer Vision and Image Understanding, vol. 117, pp. 663 - 659, 2013.

[2] Yokono J.J, Poggio T.: A multiview face identification model with no geometric constraints. Intern. Conference on Automatic Face and Gesture Recognition, pp. 493 - 498, 2006.

[3] Song Y., Leung T.: Context-aided human recognition – clustering. European conference on Computer Vision, pp. 382 - 395, 2006.

[4] Ouyang Y., Zhang S., Zhang Y.: Based on cluster tree human action recognition algorithm for monocular video. Journal of Computational Information Systems, pp. 4082 - 4089, 2011.

[5] Ben-Arie J., Wang Z., Pandit P., Rajaram S.: Human activity recognition using multidimensional indexing. Transactions on Pattern Analysis and Machine Intelligence, pp. 1091 - 1104, 2002.

[6] Shimizu H.: Direction estimation of pedestrian from multiple still images. Intelligent Vehicles Symposium, pp. 596 - 600, 2004.

[7] Rani M.P., Arumugam G.: An efficient gait recognition system for human identification using modified ICA. Intern. Journal of Computer Science and Information Technology, vol. 2, pp. 55 - 67, 2010.

[8] Dalal N., Triggs B.: Histograms of oriented gradients for human detection. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 1-8, 2005.

[9] Watanabe T., Ito S., Yokoi K.: Co-occurrence histograms of oriented gradients for pedestrian detection. IPSJ Transactions on Computer Vision and Applications, vol. 2, pp. 39 – 47, 2010.

[10] Qiang Z., Yeh M.-C., Kwang-Ting C., Avidan, S.: Fast human detection using a cascade of histograms of oriented gradients. Conf. on Computer Vision and Pattern Recognition, pp. 1491 – 1498, 2006.

[11] Sabzmeydani P., Mori G.: Detecting pedestrians by learning shapelet features. Computer Vision and Pattern Recognition, pp. 1 – 8, 2007.

[12] Mikolajczyk K., Schmid C., Zisserman A.: Human detection based on a probabilistic assembly of robust part detectors. European Conference on Computer Vision, pp. 69 – 82, 2004.

[13] Ioffe S., Forsyth D.A.: Probabilistic methods for finding people. International Journal of Computer Vision, vol. 43, pp. 45 – 68, 2001.

[14] Rani M. P., Arumugam G.: An efficient human gait recognition system using modified independent component analysis (MICA), International Journal of Computer Science and Information Technology, pp.55 – 67, 2010.

[15] Bertozzi M., Broggi A., Fascioli A., Graf T., Meineckew M.: Pedestrian detection for driver assistance using multiresolution infrared vision. Transactions on Vehicular Technology, pp. 1666 – 1678, 2004.

[16] Wang W., Zhang J.. Shen C.: Improved human detection and classification in thermal images. International Conference on Image Processing, pp. 2313 – 2316, 2010.

[17] Benezeth Y., Emile B., Laurent H., Rosenberger C.: A real time human detection system based on far infrared vision. Lecture Notes in Computer Science, vol. 5099, pp 76 – 84, 2008.

[18] Liu Q., Zhuang J., Ma J.: Robust and fast pedestrian detection method for far-infrared automotive driving assistance systems. Infrared Physics & Technology, vol. 60, pp. 288 – 299, 2013.

[19] Xia L., Chen C., Aggarwal J.K.: Human detection using depth information by Kinect. Computer Society Conference on Computer Vision and Pattern Recognition Workshops, pp. 15 – 22, 2011.

Received: 18.03.2015 *Paper reviewed* *Accepted: 05.05.2015*

Ph.D. Sebastian BUDZAN

He obtained his education in the Silesian University of technology at Department of Automatic Control (M.Sc, Ph.D). He currently holds an academic position as professor assistant in the Measurements and Control Systems Division. His research interests focus around feature extraction and detection in 2D / 3D images, infrared imaging systems, machine vision and image processing mobile application.

e-mail: Sebastian.Budzan@polsl.pl