Human Silhouette Segmentation using Discrete Poisson Equation and Extended Watershed Algorithm

Paweł Gardziński, Krzyszkof Kowalak, Łukasz Kamiński, Sławomir Maćkowiak
Poznan University of Technology
Chair of Multimedia Telecommunications and Microelectronics
Poznan, Poland
{pgardzinski, kkowalak, lkaminski, smack}@multimedia.edu.pl

***Abstract*— In this paper a novel approach on human silhouette segmentation for surveillance systems was proposed. The described solution uses discrete Poisson equation and a combination of extended watershed algorithm with Region Growing algorithm. Experiments were performed on a commonly known database PETS 2006 and the results show that the proposed solution achieves high precision and accuracy.**

Keywords—discrete Poisson equation, silhouette segmentation, surveillance systems, watershed algorithm

I. Introduction

Nowadays the video surveillance has become the centerpiece of an intelligent automated surveillance systems and a demand for these systems is still increasing and is larger than some time ago. Issues such as an intrusion detection into the protected area or measuring the volume of traffic are no longer a problem and manufacturers of smart cameras offer these solutions already implemented in their devices. Despite this fact, the new recognition methods or better detection method of complex events are developed and demanded. Commonly used methods for event recognition are methods based on silhouettes [1] or object contours [5]. Other techniques exploit the Hidden Markov Model [2] or Curvature Scale Space Descriptor (CSSD for short) and Angular Radial Transform Descriptor (ARTD for short) which are normalized under the MPEG-7 standard [8,7,3,11].

Unfortunately all of these solutions have a weak point, which is a problem with occlusions between the objects. For example, if two observed objects are moving in a way that their silhouettes (or contours) connect together, the efficiency of the detection of events on these silhouettes is reduced to zero, unless a prediction method is applied. The above problem is partially solved by the use of solutions based on Histogram of Oriented Gradients [2, 12]. However, these solutions have a much lower efficiency and require prior learning. Moreover, to achieve a higher efficiency a training sets counted in tens of thousands of images are required, which still does not provide coverage of all cases.

Therefore, this article presents a method to solve above defined problem to separate the objects based on the discrete Poisson equation calculated for the joint silhouettes. Proposed solution accepts binary images of silhouettes as an input and next, the descriptors from the discrete Poisson equation [10] are calculated. Then, using extended watershed algorithm, the estimation of the number of separate objects is made and boundaries between them are determined.

This paper is organized as follows. Section 2 presents the idea of the proposed algorithm. Firstly, the descriptor of silhouette and the discrete Poisson equation are presented and then the modified watershed segmentation algorithm together with a region growing method are presented. Section 3 presents the results of the experiments. Section 4 provides a summary and a proposal for further work.

II. Human Silhouette Segmentation

Proposed solution presented in this paper is divided into two main parts. In the first part a silhouette descriptor is calculated based on a discrete Poisson equation (we will call it a Poisson Descriptor or PD for short). In the second part a watershed segmentation is performed in order to estimate local maxima on PD and then use them to perform silhouette segmentation.

A. Silhouette Descriptor and Discrete Poisson Equation

Let us consider a silhouette S that consists of a grid of points distant from each other by d and surrounded by a closed contour δS. Then in each point $(x, y) \in S$ we place a set of particles and let them move in a random walk (in each step a particle moves to one of four neighboring points choosing it with equal probability). Each of those particles continues a random walk until they hit a boundary contour δS. Then, by measuring a mean distance function $F(x, y)$ of a random walk distance from a set of particles, we get an approximation of discrete Poisson equation which is given by the following formula:

$$F(x, y) = d + 0.25(F(x + d, y) + F(x - d, y) + F(x, y + d) + F(x, y - d)) \quad (1)$$

This equation can be determined recursively in the following way: for each point $(x, y) \in \delta S$ the value of $F(x, y) = 0$ and for the rest of points $(x, y) \in S$ the value $F(x, y)$ is equal to arithmetic average of values from immediate neighbors plus a constant (expressing the distance to those direct neighbors). For simplicity we take $d = 1$, which corresponds to a distance between neighboring pixels.

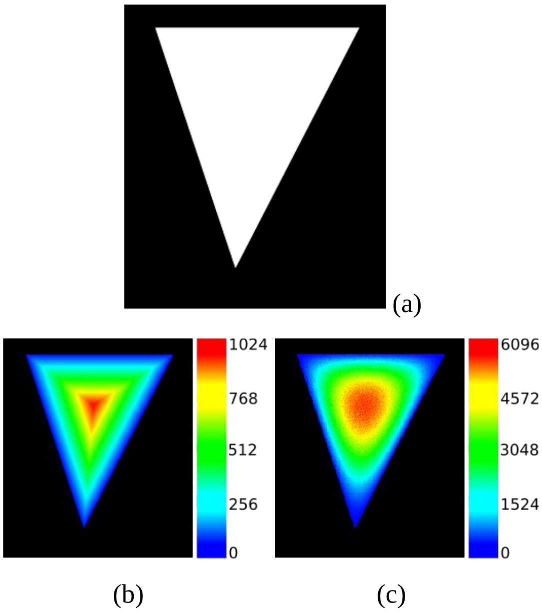

Fig. 1. Comparison of distance transform (b) with Poisson Descriptor (c) computed for input image (a).

Poisson Descriptor presented in this paper can be compared to Distance Transform (or DT for short), which is shown on Fig. 1. Both methods give a result expressing a distance from boundary contour, however DT contains only information of the closest contour point. On the other hand, a Poisson Descriptor expresses a statistical distance from many contour points (the more particles we choose the more information about whole contour we can include). In a result PD gives a more smooth descriptor and what is more, low values characterize relatively thin limbs. In video surveillance where the main focus is put on people, a PD calculated in this way gives relatively high values on their torso and low values on limbs. It allows to efficiently separate each person from a joint silhouette which has been proved in this paper.

A descriptor prepared in this way is then used to estimate local maxima on a joint silhouette that are used as seed points for region growing algorithm that has been modified so as to estimate separate silhouettes.

B. *Watershed Segmentation Algorithm*

A second part of proposed solution is the estimation of number of objects and their separation from the joint silhouette (or input image). It is a combination of two generally known segmentation techniques together with authors additions to find seeds and estimate borders.

First, we sort all points $p = (x, y)$ of input image (Poisson Descriptor calculated as described in paragraph 2.1) in ascending order of values of $F(x, y)$. Then we conduct a modified watershed algorithm that has been shown on a block diagram on Fig. 2.

Diagram uses the following notation:

- *Mask(x, y)* is a mask to mark visited points,

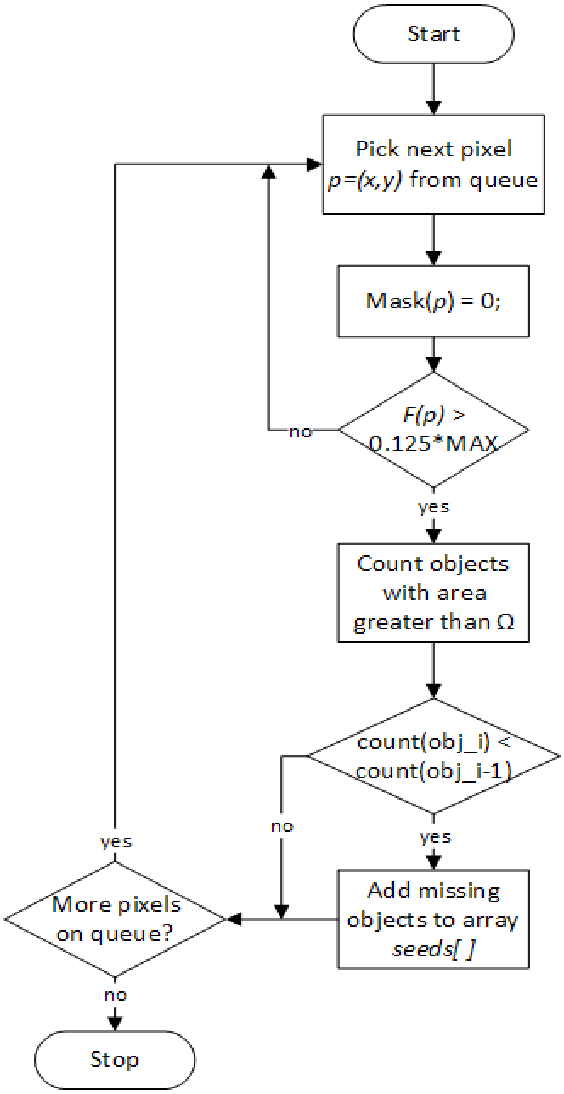

Fig. 2. Block diagram of seed finding algorithm using watershed segmentation.

- *MAX* is the highest value in PD (for example 6096 as shown on Fig. 1),
- Ω is a threshold to filter areas that are too small (artifacts or noise) which depends on input image size,
- $count(obj_i)$, $count(obj_{i-1})$ are numbers of objects in current and previous iteration steps,
- *seeds[.]* is an output array containing seed positions.

As a result of this algorithm we estimate a set of local maxima points on Poisson Descriptor. It is used as a seed array for the last part of the proposed solution.

Lastly, we conduct a region growing algorithm to estimate separate objects from a joint silhouette. An exact implementation of region growing algorithm can be found in [4, 6]. The approach in this case is straight forward. We take a set of seed points and perform this algorithm on a joint silhouette. An example of end result has been shown on Fig. 3.

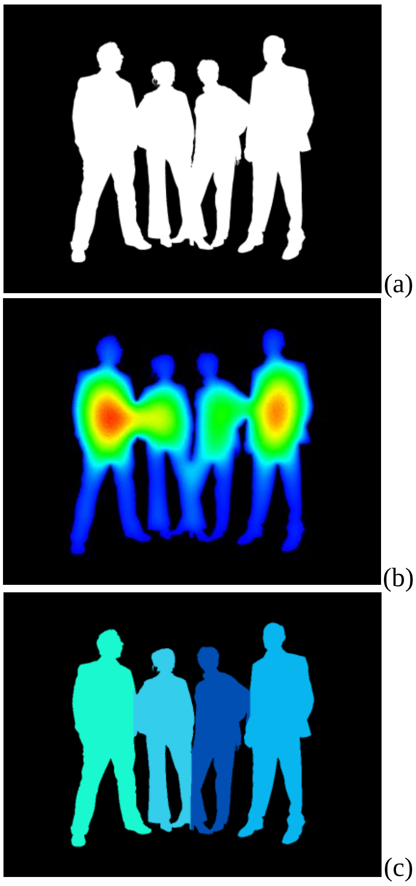

Fig. 3. Example of a result of the proposed silhouette segmentation algorithm. Input image (a), calculated Poisson Descriptor (b) and result of silhouette segmentation (c).

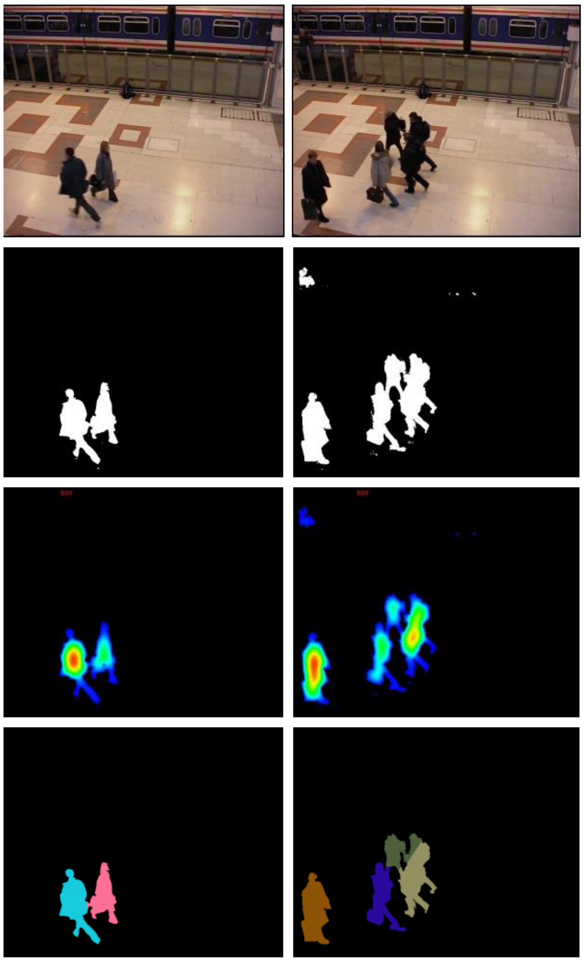

Fig. 4. Example result of proposed solution on PETS 2006 dataset.

III. Experimental Results

The effectiveness of the proposed solution has been evaluated in two stages. During first stage the algorithm has been tested and its parameters has been optimized on a varied set of test images (e.g. people shown on Fig. 3). During second stage the algorithm has been tested on the PETS 2006 dataset [9] to check the proposed solution under more realistic scenery. This dataset contains 4 views with 3020 frames. Results presented in this paper utilize view 3 from this dataset. It contains a background image and then a series of consecutive frames that were used to estimate silhouettes shown in Fig. 4. The aim of the experiment was to measure the effectiveness of the silhouette separation using estimated number of distinct objects in each frame. The efficiency has been compared with the manually marked ground truth, which enabled an effectiveness evaluation by use of two popular metrics precision and recall. They are defined as follows:

$$precision = \mathrm{TP} / (\mathrm{TP} + \mathrm{FP}), \tag{2}$$

$$recall = \mathrm{TP} / (\mathrm{TP} + \mathrm{FN}), \tag{3}$$

where: *TP* is a set of true positive results, *FP* is a set of false positive results and *FN* is a set of false negative results. An example of results on PETS 2006 dataset has been shown on Fig. 4. Then, because of PD's similarity to distance transform the proposed solution has been compared with an algorithm presented in [13]. It is a solution to object segmentation that utilizes OpenCV functions that implement distance transform and watershed algorithms to distinguish each card in a pile of cards.

The method used for comparison works in five steps. Firstly, a Laplacian filtering is done in order to acute the edges of the foreground objects. Then a distance transform is computed together with a normalization and thresholding. Next, dilation is performed in order to find peaks that are used together with cv::findContours to find markers for the last step. Finally, a watershed algorithm is performed together with post processing to give a final result.

The evaluation was conducted as follows: each frame from view 3 in dataset has been marked manually for ground truth and both methods’ detections have been compared with it. As a result for each frame a set of three metrics has been calculated for both methods. A sum of all detections, over all frames, for both distance transform with watershed algorithm and Poisson descriptor with extended watershed is shown in Table 1. A method presented in this paper is written in bold.

TABLE I. HUMAN SILHOUETTE SEGMENTATION EVALUATION

Method	Object Counting				
	Ground truth	Detections	***TP***	***FP***	***FN***
Distance Transform + Watershed	6665	5282	4089	1193	2576
Poisson Descriptor + Watershed	**6665**	**5259**	**4835**	**424**	**1830**

Therefore it is visible that the proposed method achieved a high precision of 91,9% with 72,5% recall over the other method that reached 77,4% precision with 61,3% recall.

IV. SUMMARY

Human silhouette segmentation method presented in this paper is an original solution to silhouette occlusion problem in surveillance sequences. It is a common phenomenon that reduces efficiency of silhouette or contour analysis algorithms.

In this method, processing is divided into to two stages. The first stage consists of determining silhouette descriptor from discrete Poisson equation. In the second stage there are two algorithms performed simultaneously: modified watershed segmentation and region growing. According to authors knowledge this is the first use of Poisson Descriptor to silhouette segmentation process. Presented descriptor can be successfully used in different applications in surveillance systems as it has been presented in [1,10].

Presented method can provide a basis to further processing including activity analysis and prevention in the field of video surveillance. Experimental results suggest that applications of presented algorithm allow to significantly increase the efficiency of that kind of systems.

ACKNOWLEDGMENT

The presented work has been funded by the Polish Ministry of Science and Higher Education for the status activity consisting of research and development and associated tasks supporting development of young scientists and doctoral students.

REFERENCES

[1] L. Gorelick, M. Blank, E. Shechtman, M. Irani, and R. Basri, “Actions as space-time shapes,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 29, no. 12, pp. 2247–2253, Dec. 2007

[2] Y. Huang, H. Yang, and P. Huang, “Action recognition using hog feature in different resolution video sequences,” in 2012 International Conference on Computer Distributed Control and Intelligent Environmental Monitoring (CDCIEM), March 2012, pp. 85–88.

[3] K. Mokhtarian and M. Bober, “Curvature Scale Space Representation: Theory, Applications and MPEG-7 Standardisation”, Kluwer Academic (now Springer), 2003.

[4] W. K. Pratt, “*Digital Image Processing 4th Edition”*, John Wiley & Sons, Inc., Los Altos, California, 2007

[5] S. Maćkowiak, P. Gardziński, Ł. Kamiński, K. Kowalak. „Human Activity Recognition in Multiview Video”, 11th IEEE International Conference on Advanced Video and Signal-Based Surveillance, South Korea, 2014, pp. 148-153.

[6] M. Petrou and P. Bosdogianni, *Image Processing the Fundamentals*, Wiley, UK, 2004.

[7] B.S. Manjunath, P. Salembier, and T. Sikora, “Introduction to MPEG-7: Multimedia Content Description Interface” eds. Wiley, 2002

[8] M. Bober, “MPEG-7 Visual Shape Descriptors,” IEEE Trans. Circuits and Systems for Video Technology, vol. 1, no. 6, 2001.

[9] “PETS 2006 Benchmark Dataset” http://www.cvg.reading.ac.uk/PETS2006/data.html

[10] L. Gorelick, M. Galun, E. Sharon, A. Brandt, and R. Basri, “Shape Representation and Classification Using the Poisson Equation,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 28, no. 12, Dec. 2006.

[11] “The MPEG-7 XM software repository,” http://www.lis.ei.tum.de/research/bv/topics/mmdb/e_mpeg7.html, 2005.

[12] H.Qian, J.Zhou, Y.Yuan, „Visual-based fall detection using histogram of oriented gradients of Poisson distance image” in Chinese Automation Congress (CAC), 2015, pp. 657-662

[13] “OpenCV Tutorial: Distance Transform and Watershed Segmentation” http://docs.opencv.org/3.1.0/d2/dbd/tutorial_distance_transform.html#gsc.tab=0