World Academy of Science, Engineering and Technology
International Journal of Electrical and Computer Engineering
Vol:2, No:10, 2008

Efficient Mean Shift Clustering Using Exponential Integral Kernels

S. Sutor, R. Röhr, G. Pujolle, and R. Reda

***Abstract*—**This paper presents a highly efficient algorithm for detecting and tracking humans and objects in video surveillance sequences. Mean shift clustering is applied on background-differenced image sequences. For efficiency, all calculations are performed on integral images. Novel corresponding exponential integral kernels are introduced to allow the application of non-uniform kernels for clustering, which dramatically increases robustness without giving up the efficiency of the integral data structures. Experimental results demonstrating the power of this approach are presented.

***Keywords*—**Clustering, Integral Images, Kernels, Person Detection, Person Tracking, Intelligent Video Surveillance.

I. Introduction

CLUSTERING is a widespread task in pattern recognition and image processing. Mean shift has become one of the most popular clustering algorithms over the last decades dating back to 1975 [1]. For the special case of video surveillance a very efficient approach for clustering difference images for detecting and tracking persons was proposed [2]. Here, all calculations were based on integral images [3], which accelerates the mean shift calculations dramatically. However, this approach limited the mean shift calculation to a uniform kernel, which reduced the flexibility of this algorithm. In this paper, exponential integral kernels are introduced which allow mean shift to be calculated on integral images with weighted non-uniform kernels. This brings the benefit of very efficient calculation and the advantage of weighted clustering eliminating outliers and improving overall robustness.

Section II introduces mean shift applied to background-differenced video sequences. In Section III the integral image methodology is introduced, which highly accelerates the mean shift calculation process. Section IV extends this approach by constructing non-uniform weighting functions. Finally experimental results are presented in Section V.

II. Mean Shift Clustering

In this work the focus is on applying mean shift [4] to background differenced frames of video sequences. Such background differenced images result by subtracting a statistical background model from every incoming video frame, resulting in an image like depicted in Fig. 1(b). The aim in this case is the detection and classification of foreground objects, like humans or vehicles. These background-subtracted images are then clustered to detect foreground objects. As can be seen these correspond to the brighter pixels in the image while the background remains dark.

(a) (b)

Fig. 1 A frame (a) and the corresponding background-differenced frame (b)

The mean shift clustering procedure on background-differenced images is carried out in four steps:

1) Seed points are generated around local maxima in the difference image. Around every seed point, an area of interest is generated. This area is usually chosen to be rectangular for computational complexity; it could just as well be circular or elliptical according to the chosen clustering algorithm. The size and shape of this area are important tunable parameters, usually set to the approximate size of the object to be detected and tracked. For this area a weight function is defined, to give different weights to the pixels in further calculations. This area with its weight function is referred to as „kernel“. In the case that all weights are the same (constant) this is called a uniform kernel, otherwise a non-uniform kernel. Note that in this paper the term kernel is used as a term for a weighting function that can be applied for mean shift calculation.

2) On every kernel area, a vector pointing towards the highest density point is determined. This point corresponds to

S. Sutor and R. Röhr are with KiwiSecurity Software GmbH, Austria (www.kiwi-security.com).
G. Pujolle is with Université Pierre et Marie Curie, Labortoire d'informatique de Paris 6 (www.lip6.fr).
R. Reda is with Innovation Communication Technologies, Austria / Germany (www.ictmc.com).

the brightest spot with respect to the background-differenced sequences. This vector is called the mean shift vector. In the case of a uniform kernel the mean shift vector points towards the center of gravity which is not necessarily inside an area of high density, which can be troublesome in some cases as the following sections shows.

3) The kernel is set to the point the mean shift vector pointed to, and the whole procedure starts over and over until the displacement falls below a certain threshold or a maximum number of iterations. Usually convergence is reached within a few iterations. These consecutive points, starting from the seed point to the point of termination, form the mean shift convergence path.

4) All paths converging towards the same mode are sought and grouped. Note that in practice, displacements of a few pixels may occur due to limited kernel support and rounding errors. The grouped seed points form the bounding box of an object. Due to the mean shift procedure holes in the difference image are bridged, hence objects that are approximately the size of the kernel are clustered.

The choice of kernels may differ depending on the task or scene, which the mean shift clustering is applied to. A simple uniform kernel is much faster to calculate when using integral images, while a non-uniform kernel, i.e. a Gaussian kernel, is less prone to outliers, however it is not possible to calculate the mean shift procedure on the efficient integral images. In the following section a non-uniform integral kernel will be presented which demonstrates the advantages of a weighted kernel and the efficiency of integral image calculations.

A. Calculation of Mean Shift Using a Uniform Kernel

Assuming a rectangular region of interest and a uniform kernel, the mean shift vector is calculated by first summing up all pixel values in the region. For a kernel with coordinates (x,y) as top-left corner, the sum s is calculated on the Image *I* as:

$$s = \sum_{p=x}^{x+w} \sum_{q=y}^{y+h} I(p,q). \quad (1)$$

with w and h as the width and height of this area respectively. Further, the x-weighted area sum s_x is calculated as

$$s_x = \sum_{p=x}^{x+w} \sum_{q=y}^{y+h} x \cdot I(p,q), \quad (2)$$

and the y-weighted area sum s_y is calculated as

$$s_y = \sum_{p=x}^{x+w} \sum_{q=y}^{y+h} y \cdot I(p,q). \quad (3)$$

The mean shift vector coordinates (x_{new}, y_{new}), which represent the coordinates of the next point in the mean shift convergence path are thus given by

$$x_{new} = \frac{s_x}{s}$$
$$y_{new} = \frac{s_y}{s}.$$

III. Integral Images

Usually, the computationally expensive summations must be calculated for every single seed point in the image and subsequently need to be iterated several times. This computationally expensive process can be severely accelerated. A speed boost of up to a factor of 30 has been measured [2].

Three images are pre-calculated on the incoming difference image. First, the integral image which is calculated as

$$\text{int}(x,y) = \sum_{p=0}^{x} \sum_{q=0}^{y} I(p,q). \quad (5)$$

The corresponding value at (x,y) is the sum of all gray values in the image area [(0,0);(x,y)], which is calculated as

$$s = sum(ax,ay,bx,by) = \text{int}(bx,by) - \text{int}(ax,by) - \text{int}(bx,ay) + \text{int}(ax,ay). \quad (6)$$

Here the advantage is that the sum of all pixels in the desired area [(ax,ay)–(bx,by)] is simply calculated by four additions and subtractions. Fig. 2 exemplifies this.

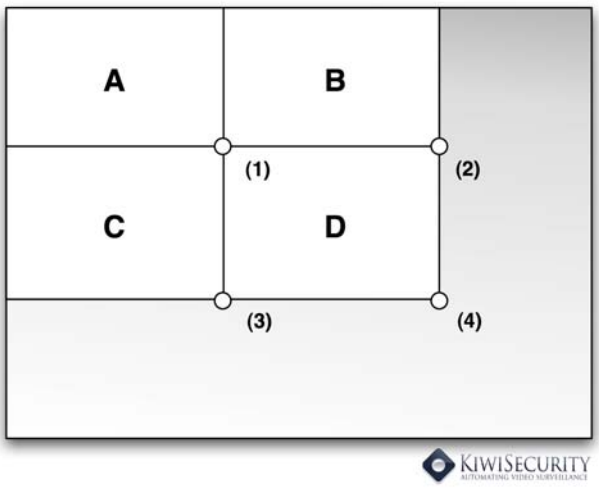


Fig. 2 The area sum of D can be computed by only 4 additions and subtractions: (4) + (1) – (2) – (3)

The x- and y-weighted integral images are calculated the same way, for the x-weighted area sum int_x:

$$\mathrm{int}_x(x,y) = \sum_{p=0}^{x}\sum_{q=0}^{y} p \cdot I(p,q), \tag{7}$$

similarly for the y-weighted area sum int_y:

$$\mathrm{int}_y(x,y) = \sum_{p=0}^{x}\sum_{q=0}^{y} q \cdot I(p,q). \tag{8}$$

IV. Non-Uniform Integral Kernels

The aim of a non-uniform kernel is to weight pixel values differently depending on their location, i.e. by giving pixels closer to the kernel center a higher weight than pixels further away. Nevertheless it is still desirable to use the efficient integral images as data structure.

Accordingly, the following problem arises: When considering a uniform kernel, pixel values are not weighted (or just weighted by a constant factor); if a higher weight closer to the kernel center is now desired it would be possible to basically split the kernel in half and weight each side linearly as depicted in Fig. 3. (Note that this will only be discussed for the vertical case, the horizontal are calculated analogous.)

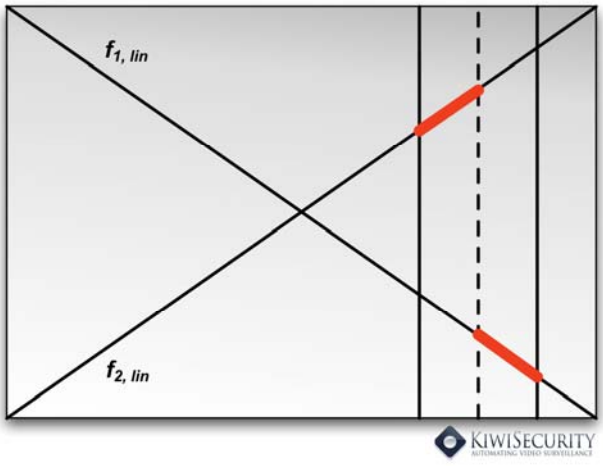


Fig. 3 Constructing a linear kernel resulting in asymmetrical weights

Inside the kernel, the pixel values on the left side are weighted with a linear monotonous growing function $f_{2,lin}$; whereas the pixel values on the right side are weighted with a linear monotonous falling function $f_{1,lin}$. Constructing the linear weight function W(p) from these two functions, weighting becomes asymmetrical. The following equations further show this problem for W(p) as a linear function:

$$x_{new} = \frac{s_x}{s} = \frac{\sum_{p=x}^{x+w}\sum_{q=y}^{y+h} y \cdot I_{weighted}(p,q)}{\sum_{p=x}^{x+w}\sum_{q=y}^{y+h} I_{weighted}(p,q)} =$$

$$\frac{\sum_{p=x}^{x+w/2}\sum_{q=y}^{y+h/2} y \cdot I_{weighted\_left}(p,q)}{\sum_{p=x}^{x+w/2}\sum_{q=y}^{y+h/2} I_{weighted\_left}(p,q)} + \tag{9}$$

$$\frac{\sum_{p=x+w/2}^{x+w}\sum_{q=y+h/2}^{y+h} y \cdot I_{weighted\_right}(p,q)}{\sum_{p=x+w/2}^{x+b}\sum_{q=y+h/2}^{y+h} I_{weighted\_right}(p,q)}$$

$$\begin{aligned} I_{weighted\_left}(p,q) &= I(p,q) \cdot W_{left}(p) \\ I_{weighted\_right}(p,q) &= I(p,q) \cdot W_{right}(p) \\ W_{left}(p) &= p \\ W_{right}(p) &= B - p \end{aligned} \tag{10}$$

When solving these equations, it can be seen the weights are not symmetrical. Hence, linear weighting functions cannot be applied.

A. The Exponential Integral Kernel

A self-similar kernel function needs to be found to avoid the symmetry problem that was shown when using linear functions. This is a function that fulfills

$$W(p) = C \cdot W(p+a), \tag{11}$$

where C is a constant factor that will be averaged out and a is constant shift in x-direction.

The group of functions that fulfills this condition is the class of exponential functions as the following proves:

$$\begin{aligned} W(p) &= q^x \\ q^x &= C \cdot q^{x+a} \\ q^x &= C \cdot q^x \cdot q^a \\ C &= \frac{1}{q^a} \end{aligned} \tag{12}$$

Hence, every exponential function, as exemplified in Fig. 4 is a suitable weighting function for constructing a weighting function that can be applied to integral images to calculate the mean shift vector; an integral kernel.

The weighting function is constructed of two exponential functions $f_{1,exp}$ and $f_{2,exp}$, which have the property of being self-similar, and hence guaranteeing a symmetrical weighting relative to the kernel center.

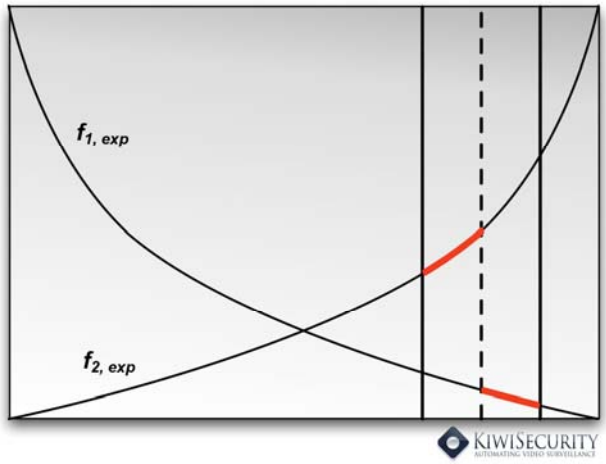


Fig. 4 A weighting function constructed from two exponential functions

V. Experimental Results

The proposed method was implemented and applied to various test sequences. In Fig. 5 a snapshot of a single frame is illustrated comparing the presented mean shift clustering on background-differenced images using a uniform kernel to using the proposed constructed exponential kernel. The mean shift convergence paths are shown and bounding boxes are superimposed for each separately detected object, hence for each cluster center.

The applied kernels - no matter if weighted or not- need to be rectangular to make use of the integral image data structure approximating the human contour outliers. These are bound to overlap when humans are in close proximity. This often causes mean shift to converge towards the "wrong" object. This phenomenon can be seen in Fig. 5 (a). The two persons on the left are merged into one cluster, while they are separated in Fig. 5 (b) where a weighting towards the kernel center was applied.

In some situation, a slower convergence towards the detected cluster centers could be observed, however, further calculations and tuning need to be done to fully explore the power of this approach. Finally due to the constructed kernel, the calculation speed of the weighted kernel in comparison to the uniform kernel is insignificantly higher, but the memory requirement is increased by a factor of four.

VI. Conclusion

The task of automatic detection and tracking of objects still remains one of the most challenging tasks in intelligent video surveillance. Current algorithms still produce too many errors and are computationally too inefficient for many real-world applications. Hence, the industry is a great driver for research in this area.

In this paper a highly efficient algorithm for detecting and tracking objects from static video surveillance cameras was presented. Integral images as data structure were exploited to achieve this efficiency for clustering background-differenced image sequences. This results in clusters for each object in the scene. Due to the introduction of the exponential integral kernels the robustness and flexibility of this algorithm could be improved dramatically.

Acknowledgments

This work was supported by the Austrian Federal Ministry for Transport, Innovation and Technology, www.bmvit.gv.at and the Austrian Research Promotion Agency "Österreichische Forschungsförderungsgesellschaft", www.ffg.at. This work will be partially included in a PhD thesis at the Pierre & Marie Curie University, Paris 6.

References

[1] K. Fukunaga and L. Hostetler, "The Estimation of the Gradient of a Density Function, with Applications in Pattern Recognition," *IEEE Transactions on Information Theory*, vol. 21, pp. 32-40, 1975.

[2] C. Beleznai, B. Frühstück, H. Bischof, and W. Kropatsch, "Detecting Humans in Groups Using a Fast Mean Shift Procedure", In Proceedings of the 28th Workshop of the Austrian Association for Pattern Recognition, vol. 179, pp. 71-78, 2004.

[3] P. Viola, M. Jones, "Rapid Object Detection using a Boosted Cascade of Simple Features", IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'01) - Volume 1, pp. 511, 2001

[4] Y. Cheng, "Mean Shift, Mode Seeking and Clustering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 17, pp. 790-799, 1995.

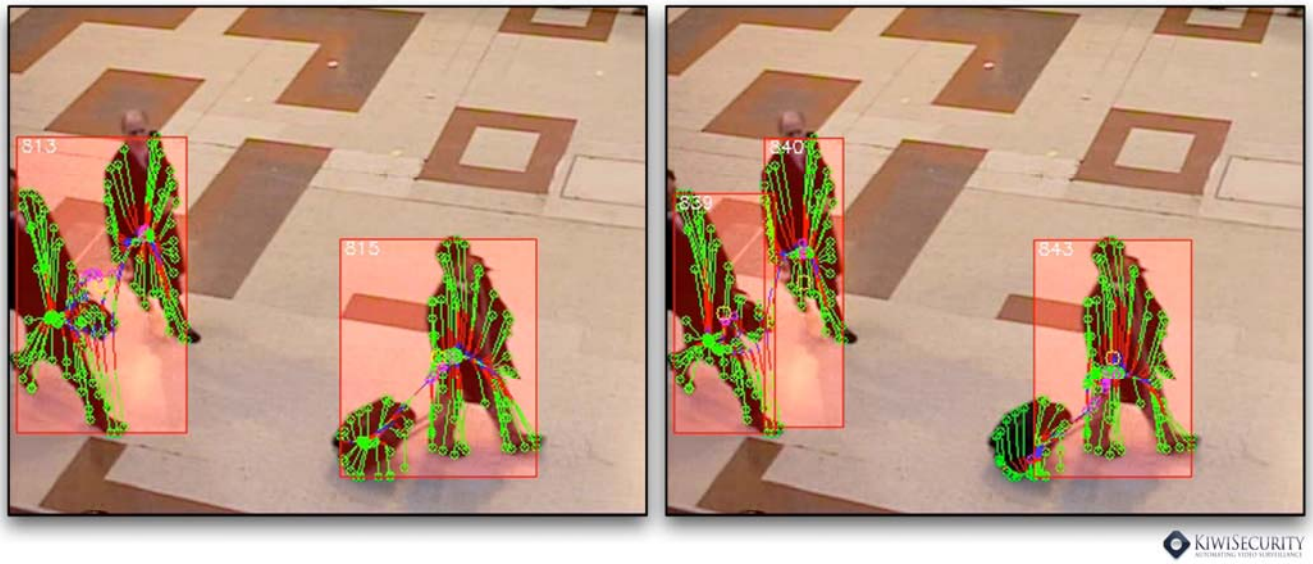
[5] F. Matusek, S. Sutor, K.Kruse, K. Kraus and R. Reda, Large-Scale Video Surveillance Systems: New Performance Parameters and Metrics, to be published at in Press, The Third International Conference on Internet Monitoring and Protection, Bucharest June 29 - July 5, 2008

[6] A. Yilmaz, K. Shafique, and M. Shah, "Target Tracking in Airborne Foward Looking Infrared Imagery," *Image and Vision Computing Journal*, vol. 21, pp. 623-635, 2003.

[7] D. Comaniciu, V. Ramesch, and P. Meer, "Kernel-Based Object Tracking," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, pp. 564-577, 2003.

[8] S. Sutor, "A Mean Shift Based Approach towards Automated Person Tracking", Master Thesis, Vienna University of Technology, 2007

Stephan R. Sutor graduated at the Vienna University of Technology in computer science (BSc and MSc). During his studies he co-founded the company KiwiSecurity, http://www.kiwi-security.com, which is developing an automated video surveillance and sensor network system, KiwiVision. KiwiSecurity has won a number of awards and was elected among the most innovative and promising start-ups in Austria. KiwiSecurity is providing products and services for large infrastructure operators and public institutions with high security requirements. Stephan is currently acting as Managing Director for Research & Development at KiwiSecurity. He further co-founded the European Security and Trust Experts Alliance "ESTE Alliance", www.estealliance.com, which provides security expertise by some of the leading security experts of Europe. Stephan is acting as TPC in a number of security related conferences and is currently working on his PhD thesis at the Pierre & Marie Curie University, Paris 6 in the area of automated video surveillance.



(a) (b)

Fig. 5 Experimental results: (a) shows mean shift applied using a uniform kernel, while in (b) the proposed constructed exponential integral kernel was used