heatmap construction using background subtraction

Nicolas A. Maduro
Laboratory for Interdisciplinary Research on Multimedia Information
CEFET-MG
Belo Horizonte, Brazil

Thales B. Nascimento
CEFET-MG
Belo Horizonte, Brazil

***Abstract*—**

***Keywords*-Heatmap; Background subtraction;**

I. Introduction

As anlises de comportamento de clientes em lojas possuem grande valor para varejistas, empresas e organizaes, j que podem ser utilizadas para aperfeioar suas estratgias de marketing, e ajudar os clientes nas tomadas de decises. Algumas das anlises realizadas por tcnicas computacionais atualmente so feitas por meio de tecnologias que impe algumas limitaes, exigindo dispositivos acoplados ao cliente ou at mesmo identificaes previamente definidas nos produtos de uma loja. Do mesmo modo, as ferramentas tecnolgicas existentes ainda so pouco acessveis, sendo compostas de solues comerciais que exigem um investimento alto e consequentemente pouco exploradas.

Este trabalho implementa um sistema baseado em viso computacional capaz de detectar clientes em lojas a fim de criar um mapa de calor para apoiar as anlises de comportamento de clientes. In order to build the heatmap, we segmented the image using background subtraction, and used morphological transformations to remove noise and correct small failures on the objects.

The remainder of this work is as follows: The second section presents the related work. The third section shows definitions of important concepts used in this work. The fourth section shows the proposed approach to solve the problem. The next section exposes the experimental results and the last section finishes with concluding remarks.

A. Contribution

The contribution of this work is to present a simple method to detect people in an store environment and provide a heatmap to help understanding the client behavior in real time.

II. Related work

O trabalho de Padua [1] desenvolve um sistema baseado em viso computacional para apoiar as anlises ttica e fsica no futsal. Em seu sistema foram utizadas as tcnicas de subtrao de fundo baseado em misturas gaussianas descritas em [2] e operaes morfolgicas sobre imagens como descrito em [3]. The work of Padua [1] design a computer vision based system

Liciotti et al [4] presents an integrated system consisted of a RGB-D camera and software able to monitor shoppers behavior and their interactions with products in shelves. Their system univocally identifies people within a store, and by using depth information it determines the kind of action the shopper is taking on the products, either touching, picking up or putting back.

Popa et al. [5] approaches the possibility of automatic understanding of customerss shopping behavior. From the video recordings, they extract features related to the spatio-temporal behavior of customers, like dynamics and time spent in each region of interest (ROI) and customer-products interaction patterns. Then they analyze the shopping sequences using a Hidden Markov Model (HMM). They got accurately classify trajectories (93%), discriminate between different shopping related actions (91.6%), and recognize shopping behavioral types by means of the proposed reasoning model in 95% of the cases.

Breitenstein [6] proposes an approach for multi-person tracking in a particle filtering framework. It combines online trained, instance-specific classifiers with generic object category knowledge, resulting in a robust multi-person tracking. The algorithm detects and tracks a large number of dynamically moving persons in complex scenes with occlusions, without relying in background models or camera calibration, using only information from the past, hence imposing very few restrictions and is suitable for online applications.

Haritaoglu and Myron [7] create a monocular real-time computer vision system that identifies shopping groups by detecting and tracking multiple people as they wait in a checkout line or service counter. The system segments each frame into foreground regions which contains multiple people. Those are further segmented into individuals using a temporal segmentation and motion cues. Once a person is detected, an appearance model based on color, edge density and mean-shift tracker is used to recover the persons trajectory. People are grouped together as a shopping group by analyzing interbody distances. The system also monitors cashiers activities to determine when shopping transactions start and end.

III. Fundamentals

Definition 1 (Image binarization)**.** *Binarization is the conversion of a gray scale image to a two values image. There are many binarization formulas, we used the following:*

$$output(x,y) = \begin{cases} G_{max} & \textit{if } input(x,y) > \textit{threshold} \\ 0 & \textit{otherwise} \end{cases}.$$

Definition 2 (Gaussian filter). *The Gaussian Filter is a 2D convolution with a kernel defined by samples of the 2D Gauss function. This function is defined as follows:*

$$G_{\sigma,\mu_x,\mu_y}(x,y) = \frac{1}{2\pi\sigma^2} e^{\frac{(x-\mu_x)^2}{2\sigma^2}} e^{\frac{(y-\mu_y)^2}{2\sigma^2}}.$$

Definition 3 (Morphological Dilation). *Dilation is the morphological transformation which combines two sets using vector addition of set elements. Let A and B be subsets of image carrier Ω. The dilation is defined as:*

$$A \oplus B = \{c \in \Omega | c = a + b \textit{ for some } a \in A \textit{ and } b \in B\}.$$

Definition 4 (Morphological Erosion). *Erosion is the morphological dual to dilation.Let A and B be subsets of image carrier Ω. The erosion is defined as:*

$$A \ominus B = \{x \in \Omega | x + b \in A \textit{ for every } b \in B.$$

Definition 5 (Morphological Opening). *The opening of image B by structuring element K is denoted by $B \circ K$ and is defined as:*

$$B \circ K = (B \ominus K) \oplus K.$$

Definition 6 (Morphological Closing). *The closing of image B by structuring element K is denoted by $B \circ K$ and is defined as:*

$$B \circ K = (B \oplus K) \ominus K.$$

Definition 7 (Background subtraction). *Background subtraction (BS) is a technique used for detecting moving objects in videos from static cameras. It calculates the foreground performing a subtraction between the current frame and a background model, which contains everything that can be considered as background.*

IV. Proposed Approach

The dataset used for the experiments was CAVIAR[1]. In order to detect people in the scene we follow scheme described in Figure 1.

[1]http://homepages.inf.ed.ac.uk/rbf/CAVIARDATA1/

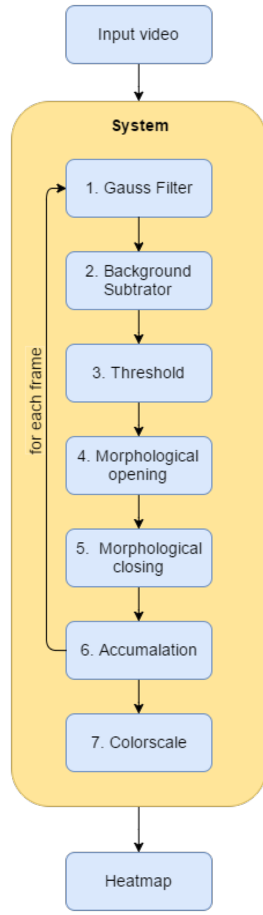

Fig. 1. Block diagram of the system.

A. Gaussian filter

For each frame of the input video we use a Gaussain filter to reduce noise. This filter produces an output image blurrier than the original image, as exemplified on Figure 2, but as we are not interested on small details, this effect has no relevance.

Fig. 2. Example of an application of a Gaussian filter on the left image

B. Background subtraction

With each filtered frame, the system perform a background subtraction in order to segment the image and detect people in the scene. While there are many BS implementations, our application demands an adaptive technique that is able to update its background model when the scene changes permanently, for example when a customer removes a product from the shelve. For that reason, we used the BS technique proposed by Zivkovic [2], which is implemented in OpenCV [8]. Figure 3 shows an example of the technique.

Fig. 3. example of a background subtraction. The the algorithm was applied on the left image to produce the right one

C. Threshold

The BS implementation we used deliver an image with the shadow pixels marked in gray, while the foreground is marked in white. The shadows are of no interest to our application, so they are removed with a threshold operation. The results of this step can be seen in Figure 4.

Fig. 4. Comparation of two images after(right) and before(left) morphological transformation

D. Morphological opening and closing

In order to improve the detection capability, we apply morphological opening and closing[3] in the output of the threshold step. The first is used to remove small noise from the image, while the second is used to remove small holes on the detected objects. A comparation of a foreground before and after the application is show in Figure 5;

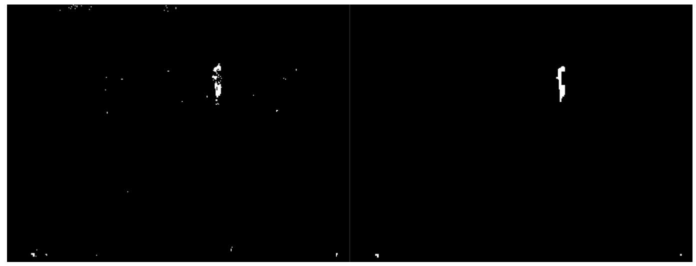

Fig. 5. Comparation of two images after(right) and before(left) morphological transformation

E. Accumulation

In this step we add the result of previous operation in an accumulator, for further generating a heatmap.

F. Normalization and display

If all frames were computed, the accumulator matrix is normalized in order to be in a displayable format *i.e.* between 0 and 1, and then it is colorized for better viewing with one of the colorscales show in Figure 6. One example of heatmap is showed in Figure 7.

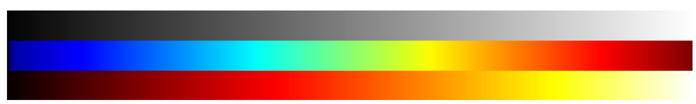

Fig. 6. colorscales jet (middle) and hot (bottom)

Fig. 7. Example of a heatmap produced by the algorithm. Both images represent the same heatmap, left in grayscale and right in jet colorscale

G. Proof of concept

In order to verify the correctness of the algorithm, it was designed an manual heatmap builder. For each frame, the user has to mark the persons on the frame. At the end the application generates a heatmap based on the user marks.

V. Experimental results

VI. Concluding Remarks

Acknowledgment

The authors would like to thank Flavio Cardeal for his course and support on this work.

References

[1] F. L. C. Pádua and P. H. C. de Pádua, "Sistema de apoio às análises tática e física no futsal baseado em visão computacional," in *Seminário de Discentes dos Programas de Pós-Graduação Stricto Sensu*, 2014.

[2] Z. Zivkovic, "Improved adaptive gaussian mixture model for background subtraction," in *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, vol. 2. IEEE, 2004, pp. 28–31.

[3] R. M. Haralick, S. R. Sternberg, and X. Zhuang, "Image analysis using mathematical morphology," *IEEE transactions on pattern analysis and machine intelligence*, no. 4, pp. 532–550, 1987.

[4] D. Liciotti, M. Contigiani, E. Frontoni, A. Mancini, P. Zingaretti, and V. Placidi, "Shopper analytics: a customer activity recognition system using a distributed rgb-d camera network," in *International Workshop on Video Analytics for Audience Measurement in Retail and Digital Signage*. Springer, 2014, pp. 146–157.

[5] M. C. Popa, L. J. Rothkrantz, C. Shan, T. Gritti, and P. Wiggers, "Semantic assessment of shopping behavior using trajectories, shopping related actions, and context information," *Pattern Recognition Letters*, vol. 34, no. 7, pp. 809–819, 2013.

[6] M. D. Breitenstein, F. Reichlin, B. Leibe, E. Koller-Meier, and L. Van Gool, "Online multiperson tracking-by-detection from a single, uncalibrated camera," *IEEE transactions on pattern analysis and machine intelligence*, vol. 33, no. 9, pp. 1820–1833, 2011.

[7] I. Haritaoglu and M. Flickner, "Detection and tracking of shopping groups in stores," in *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, vol. 1. IEEE, 2001, pp. I–431.

[8] G. Bradski, "The OpenCV Library," *Dr. Dobb's Journal of Software Tools*, 2000.