ARTICLE IN PRESS

Available online at www.sciencedirect.com

SCIENCE DIRECT®

Pattern Recognition Letters xxx (2004) xxx–xxx

Pattern Recognition Letters

www.elsevier.com/locate/patrec

Motion characterization from co-occurrence vector descriptor

Zeev Zalevsky *, Ehud Rivlin, Michael Rudzsky *

Department of Computer Science, Technion Israel Institute of Technology, Technion City, 32000 Haifa, Israel

Received 2 August 2004

Abstract

A novel operator called the co-occurrence vector to be used for motion characterization is presented. The vector has the advantage of requiring low computation complexity. It allows estimating the relative position, the scale and the rotation of objects in the scene image. This estimation may be useful for obtaining time to impact and relative distance between objects for navigation and collision avoidance. Experiments confirm the value of co-occurrence vector for estimation of motion parameters and as a tool for characterization of scale, rotation and relative shift of objects in the scene.

Keywords: Co-occurrence matrix; Time to impact; Navigation; Collision avoidance

1. Introduction

Several researches have demonstrated various approaches of deriving partial information about motion characterizations (Fablet and Bouthemy, 2003). Examples include relative distances of objects (Aloimonos and Duric, 1994), time to collision (Ancona and Poggio, 1993; Fermuller and Aloimonos, 1995) and qualitative information about an object's shape (Cipolla and Zisserman, 1992). The main problem of those approaches is that they are based on the "optical flow" equation and thus require heavy computation resources.

The co-occurrence matrix is a statistical tool for extracting second order texture information from images (Haralick et al., 1973; Haralick, 1979; Chen and Pavlidis, 1979; Conners et al., 1984; Parkkinen et al., 1990). It is very useful for detection of texture periodicity and for texture classification and segmentation. It can be used also for calculation of optical flow as described by Boyce et al. (1992). Recently it was used for object detection and recognition by Kovalev and Petrou (1996).

In this paper we define the term co-occurrence vector and show how to use this tool for estimation

* Corresponding authors. Tel.: +972 4 8294961; fax: +972 4 8293900 (M. Rudzsky).

E-mail addresses: zzeev@eng.tau.ac.il (Z. Zalevsky), ehudr@cs.technion.ac.il (E. Rivlin), rudzsky@cs.technion.ac.il (M. Rudzsky).

0167-8655/$ - see front matter © 2004 Published by Elsevier B.V.
doi:10.1016/j.patrec.2004.09.015

of the motion parameters of an image. The co-occurrence vector has low computation complexity and it may characterize scale, rotation and relative translation of objects in the observed scene.

The co-occurrence matrix is a matrix extracting second order texture information from the image. An element a_{nm} of the co-occurrence matrix corresponds to the number of times that the gray level $n-1$ and the gray level $m-1$ appeared in the input image while a pre-determined translation vector $\vec{V} = \{\delta x, \delta y\}$ separates them. This matrix has a histogram related insight. Generally speaking, this matrix is four-dimensional:

$$P(n, m, \delta x, \delta y) = T_{\text{co}}\{I_{\text{in}}\} \tag{1}$$

where δx and δy are the separation distances in the x and the y directions respectively, T_{co} is the co-occurrence operator and I_{in} is the input image. Usually a distance of one pixel describes backgrounds fairly good ($\delta x = \delta y = 1$), and thus the co-occurrence matrix P becomes two-dimensional. The element $P(n, n, 0, 0)$ equals to the histogram of the image $h(n)$: $h(n) = P(n, n, 0, 0)$. If the smoothing operator T_{sm} (this operator involves a convolution with a smoothing kernel as detailed in Section 3) and the binarization operator T_{bin} are applied over the input image, the resulted co-occurrence matrix becomes a matrix with reduced dimensions since m and n may accept only the values of 1 and 2. In this case the element corresponding to the value of $m = n = 2$ describes the outlines of the details appearing within the image. Note that the binarization operation T_{bin} is executed by comparing the gray levels of the image to a given threshold. If the gray level exceeds the threshold it is chosen to be one otherwise it is set to be 0. In order to examine the details appearing in the binarized image in a certain direction θ we will choose: $(\delta x, \delta y) = (\mu \cdot \cos(\theta), \mu \cdot \sin(\theta))$. Thus, for a pre-determined direction angle θ the $4-D$ co-occurrence matrix P from Eq. (1) is reduced to a $1-D$ vector which we shall call the co-occurrence vector:

$$V^{\theta}_{\text{co}}(\mu) = (T_{\text{co}} T_{\text{bin}} T_{\text{sm}} I_{\text{in}})(2, 2, \mu \cdot \cos(\theta), \mu \cdot \sin(\theta)) \tag{2}$$

This co-occurrence vector has most significant features that may characterize the movement of objects within the image. Note that since the co-occurrence vector is a $1-D$ vector its computation involves small numerical complexity based upon the FFT algorithm as we describe in Section 2. One of the most important features to be extracted from the co-occurrence vector is the time to impact τ. Since the size of the details seen in an image is a function of the range between the scene and the camera, τ may easily be extracted out of the co-occurrence vector descriptor. For instance, if the followed object is getting closer to the observer its size is increased. This increasing will scale the features of the co-occurrence vector. Thus, knowing the size of the object and by calculating the scaling rate between the different frames one may compute the time to impact parameter. In order to be able to identify rotation of the object between two frames at least two co-occurrence vectors are required. For instance, assuming that one computes two co-occurrence vectors for $\theta = 0$ and $\theta = \pi/2$. A rotation of the object will cause scaling of the projection of the object's features on those two orthogonal axes. An increasing scale and position of features will occur in one axis and a decreasing scale will occur in the other. Due to the ratio between the scaling obtained in the two co-occurrence vectors the parameters of the rotation may be extracted. Note that one vector is not sufficient since then no distinction can be obtained. For instance, in case when the object is getting closer to the observer, in each axis scale of features occurs but the ratio between the axes remain constant. Note also that except of using the co-occurrence vector as a tool to characterize image motion, it may also be used as a recognition tool (Kovalev and Petrou, 1996). The co-occurrence vector of an input image and a reference image may be correlated in order to estimate their mutual matching.

In Section 2 we will present the mathematical derivation of the co-occurrence descriptor and demonstrate how to estimate it using the FFT algorithm. Using the same approach backwards allows estimating objects positions and dimensions within the reference scene. In Section 3 we demonstrate several experimental validations including extraction of image rotation and time to impact parameters. The paper is concluded in Section 4.

2. Mathematical derivation

Now let us derive the mathematical model describing the co-occurrence descriptor. After applying the smoothing and the binarization we may model an image of a scene with N objects as:

$$I(x) = \sum_{i=1}^{N} \mathrm{rect}\left(\frac{x - X_i}{\Delta X_i}\right) \tag{3}$$

where ΔX_i is the width of the object i and X_i is the position of its center,

$$\mathrm{rect}(x) = \begin{cases} 1 & \text{if } -1/2 \leqslant x \leqslant 1/2 \\ 0 & \text{if } |x| > 1/2 \end{cases} \tag{4}$$

See Fig. 1 where $I(x)$ is depicted. Applying the co-occurrence operator T_{co} from Eq. (2) over $I(x)$ from Eq. (3) for $\theta = 0$ and following the basic definition of the co-occurrence as described just in front of Eq. (1) we obtain co-occurrence function $f(\mu)$

$$f(\mu) = \sum_{i=1}^{N} \left\{ (\mathrm{rect} \otimes \mathrm{rect})\left(\frac{\mu}{\Delta X_i}\right) \right\} + \sum_{j=1}^{N} \sum_{i>j}^{N} \left\{ (\mathrm{rect} \otimes \mathrm{rect})\left(\frac{\mu - X_{ij}}{\Delta X_i}\right) \right\} \tag{5}$$

Here symbol $\otimes$ denotes a convolution operator. Performing summation in Eq. (5) we obtain:

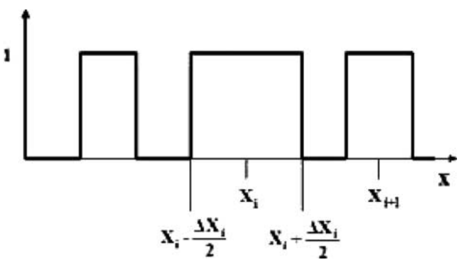

Fig. 1. 1-D image function.

$$f(\mu) = \sum_{i=1}^{N} \mathrm{trian}\left(\frac{\mu}{\Delta X_i}\right) + \sum_{j=1}^{N} \sum_{i>j}^{N} \left\{ \mathrm{trian}\left(\frac{\mu - X_{ij} - \Delta X_{ij}}{\min(\Delta X_i, \Delta X_j)}\right) + \mathrm{trian}\left(\frac{-\mu - X_{ij} + \Delta X_{ij}}{\min(\Delta X_i, \Delta X_j)}\right) + \mathrm{rect}\left(\frac{\mu - X_{ij}}{2\Delta X_{ij}}\right) \right\} \tag{6}$$

where

$$\Delta X_{ij} = \frac{\max(\Delta X_i, \Delta X_j) - \min(\Delta X_i, \Delta X_j)}{2},$$
$$X_{ij} = X_i - X_j$$
$$\mathrm{trian}\left(\frac{\mu}{\Delta X}\right) = (-\mu + \Delta X) \times [u(\mu) - u(\mu - \Delta X)] \tag{7}$$

Here $u(x)$ is the unit step function

$$u(x) = \begin{cases} 1 & \text{if } x \geqslant 0 \\ 0 & \text{if } x < 0 \end{cases} \tag{8}$$

The Fourier transform of Eq. (5) (spectrum of the co-occurrence descriptor) is:

$$F(\nu) = \sum_{i=1}^{N} \Delta X_i^2 \mathrm{sinc}^2(\nu \Delta X_i) + \sum_{i=1}^{N} \sum_{j>i}^{N} \Delta X_i \Delta X_j \mathrm{sinc}(\nu \Delta X_i) \times \mathrm{sinc}(\nu \Delta X_j) \cos(2\pi \nu \Delta X_{ij}) \tag{9}$$

This expression could be computed using the FFT algorithm. One way of estimating the parameters of the objects from the co-occurrence descriptor $f(\mu)$ could be by observing the positions and the width of the peaks (see Eq. (5)). The position of the peaks or plateaus corresponds to X_{ij}. Note that the various equations relating the position of the peaks to X_{ij} are dependent equations according to:

$$X_{i_0 j_0} = \sum_{k=0}^{j_0 - 1} X_{i_0+k, i_0+k+1} \tag{10}$$

The width of the peak is $\Delta X_i + \Delta X_j$. The value of $f(\mu = 0)$ equals to $\sum_{i=1}^{N} \Delta X_i$. Using those approximations one may obtain initial estimation for the desired parameters. Fig. 2 presents an example of

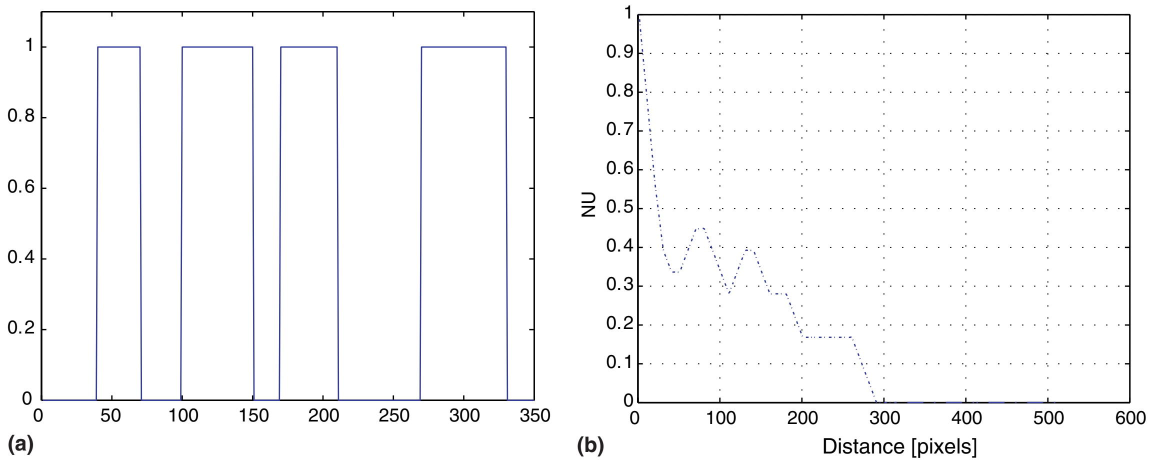

Fig. 2. Example of co-occurrence descriptors. $\Delta X_1 = 30$, $\Delta X_2 = 50$, $\Delta X_3 = 40$, $\Delta X_4 = 60$, $X_{12} = 70$, $X_{13} = 135$, $X_{14} = 245$, $X_{23} = 65$, $X_{24} = 175$, $X_{34} = 110$.

co-occurrence descriptor (normalized to 1). Fig. 2a is thresholded picture of several 1-D objects. Following the presented logic one obtains: in Fig. 2a and b there are four peaks/plateaus and thus one has four objects with three peaks that overlap with the existing four. The extraction of the ΔX_i and the X_{ij} parameters yields correspondence of up to five pixels of estimation deviation.

3. Experimental validation

In the following experiments we used the $1 - D$ co-occurrence descriptor to extract the size and the position of various objects in the difference image. To do that we have subtracted two adjacent frames and then applied a smoothing operator that involved a convolution with a 4×4 Gaussian kernel. Then, we binirized the result and computed its co-occurrence descriptor. Out of the co-occurrence descriptor we tried to extract the size and the relative position of objects in the scene:

$$T_{\text{smooth}}\{M\} = \sum_k \sum_l M(k,l)K(m-k, n-l) \tag{11}$$

where K is the Gaussian 4×4 smoothing kernel.

The first example is depicted in Fig. 3a and b. The difference image after smoothing and thresholding is seen in Fig. 3c. Fig. 3d presents in solid line the co-occurrence descriptor of Fig. 3c. The dashed line of Fig. 3d corresponds to the co-occurrence descriptor of a scene in which there are three bodies having dimensions of 28, 70 and 77 pixels respectively and relative locations of 161 and 108 pixels between them. As one may see there is good correspondence between the solid and the dashed line of Fig. 3d yielding a good estimation of the positions and dimensions of the relevant bodies within the scene. Fig. 4 presents additional example in which the computations were done along the lines at $\theta = 42°$. Figs. 4a and b present two frames from an outdoor scene. Fig. 4c is the background image. Fig. 4d is the difference after background subtraction and thresholding. Fig. 4e presents the co-occurrence descriptor obtained following the definition appearing in front of Eq. (1) and the descriptor obtained after estimation of objects dimensions and positions (computed also following the basic definition of the co-occurrence as described in front of Eq. (1)). One may observe the correspondence between the two curves of Fig. 4e. One interesting feature of the co-occurrence descriptor is possibility to estimate the rotation angle of the main movement axis of the targets. For instance if one takes the image scene depicted in Fig. 4a and b and applies the extraction of the maximal number of separate objects in different angles

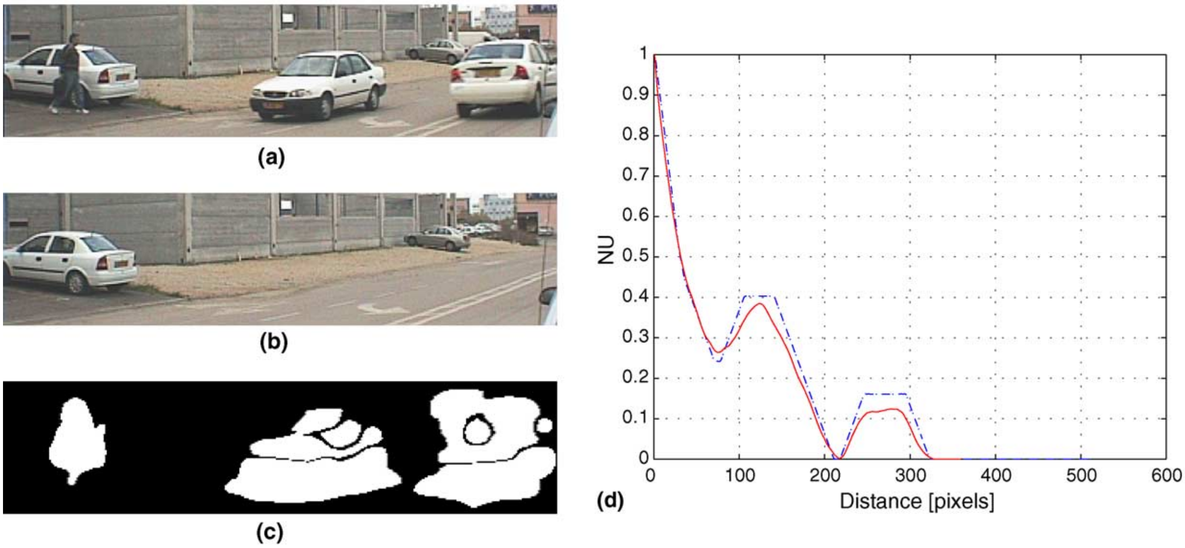

Fig. 3. (a)–(b) Two adjacent video frames. (c) The result obtained after subtraction and thresholding. (d) The co-occurrence descriptor of Fig. 2c obtained using Eq. (6) (solid line) and of the co-occurrence descriptor obtained using the definition appearing in front of Eq. (1) after estimating the position and the widths of the objects in the thresholded scene as described in the previous section (dashed).

of computation of the co-occurrence descriptor the following result is obtained: Fig. 4f was computed for several cross-sections at angles of 0°, 22.5°, 45°, 67.5° and 90°. From the maximum obtained at the angle $\theta = 45°$ one may conclude that most of the targets are oriented around this angular direction of movement. Repeating this computation for several sequential frames may assist in finding consistency in the relative angle between the camera and the axis of object's motion. The change between the positions of the peaks of the co-occurrence descriptor corresponds to change in the relative distance between the objects in the scene. If this change is due to rotation then the rotation angle may be extracted. Assuming that prior to computing the co-occurrence descriptor we perform a projections of the 2-D image on the horizontal and the vertical axes. We assume that two objects had a separation distance of R and that prior to rotation, the line connecting them creates an angle of α with the horizontal axis. Then a rotation of angle β was performed. Thus, after the rotation the separation distance is still R but the line connecting both objects creates an angle of $\alpha + \beta$ in comparison to the horizontal axis. Thus, the projections, on the horizontal axis are:

$$\begin{aligned} X_{ij}^{(1)} &= R\cos(\alpha+\beta) \\ X_{ij}^{(0)} &= R\cos\alpha \end{aligned} \tag{12}$$

where (0) and (1) designate prior and after performing the rotation, respectively. The projections on the vertical axis equal:

$$\begin{aligned} Y_{ij}^{(1)} &= R\sin(\alpha+\beta) \\ Y_{ij}^{(0)} &= R\sin\alpha \end{aligned} \tag{13}$$

Using basic trigonometric relations one obtains:

$$\begin{aligned} \frac{X_{ij}^{(1)}}{X_{ij}^{(0)}} + \frac{Y_{ij}^{(1)}}{Y_{ij}^{(0)}} &= \sin\beta\left(\frac{1}{\tan\alpha} + \tan\alpha\right) \\ \frac{Y_{ij}^{(1)}}{X_{ij}^{(0)}} - \frac{X_{ij}^{(1)}}{Y_{ij}^{(0)}} &= \cos\beta\left(\frac{1}{\tan\alpha} + \tan\alpha\right) \end{aligned} \tag{14}$$

Thus, the estimation for the rotation angle β becomes:

$$\tan\beta = \frac{\dfrac{X_{ij}^{(1)}}{X_{ij}^{(0)}} + \dfrac{Y_{ij}^{(1)}}{Y_{ij}^{(0)}}}{\dfrac{Y_{ij}^{(1)}}{X_{ij}^{(0)}} - \dfrac{X_{ij}^{(1)}}{Y_{ij}^{(0)}}} \tag{15}$$

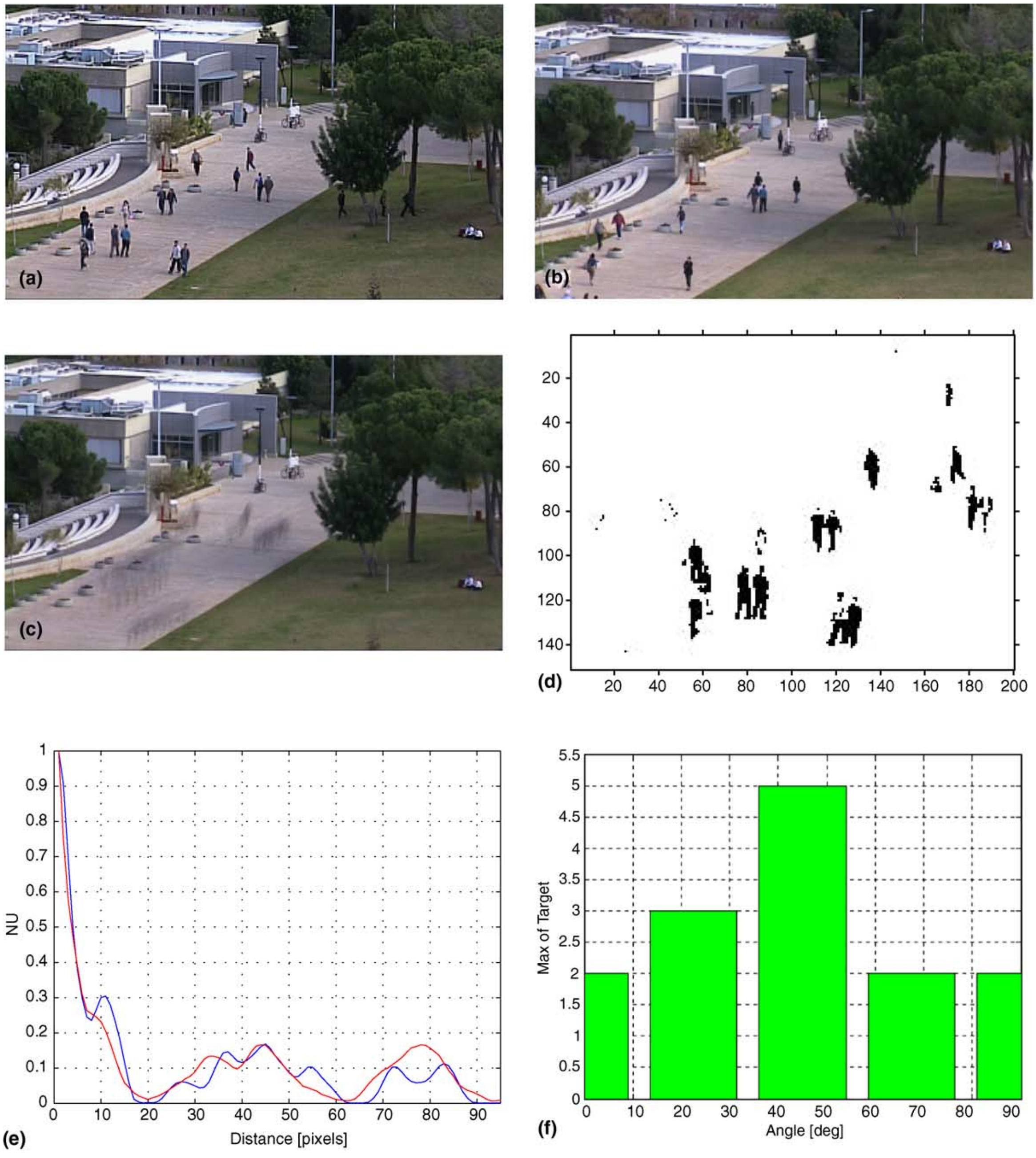

Fig. 4. (a)–(b) Two adjacent video frames. (c) Background. (d) The result obtained after background subtraction and threshold. (e) The co-occurrence descriptor of Fig. 2c and of the estimated scene. (f) Co-occurrence vector for different angles.

Temporal variations of this extracted angle will lead to obtaining the relative angular motion in time generated between the photographing platform and the targets. Note that the fact that the co-occurrence descriptor is not a linear operation (a co-occurrence descriptor of sum of vectors is

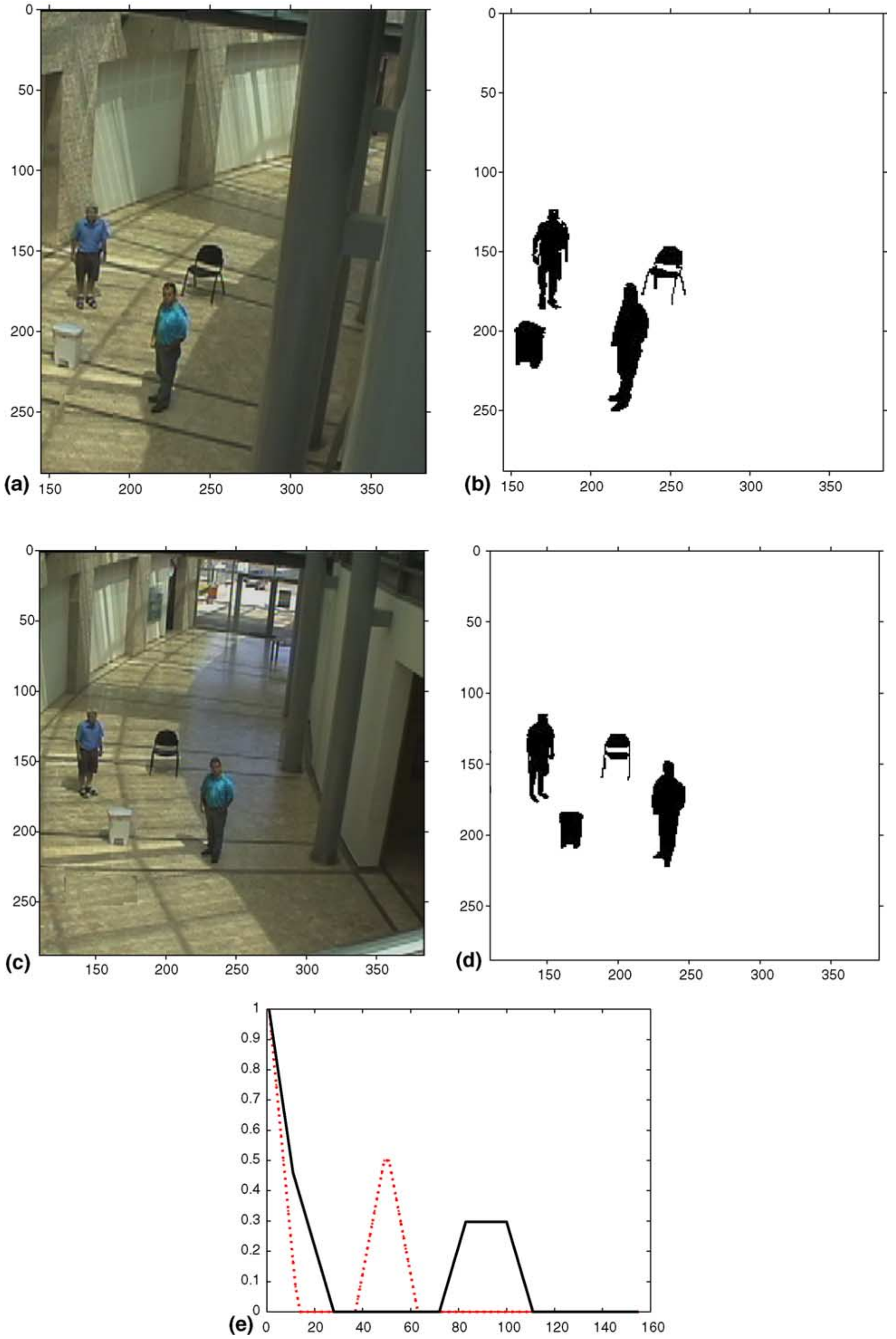

Fig. 5. (a) Original image. (b) 5(a) after background subtraction and thresholding. (c) Image rotated by 50°. (d) 5(c) after background subtraction and thresholding. (e) The co-occurrence of 5(b) performed over row number 183 (dashed line) and of 5(d) calculated over row 172 (solid line).

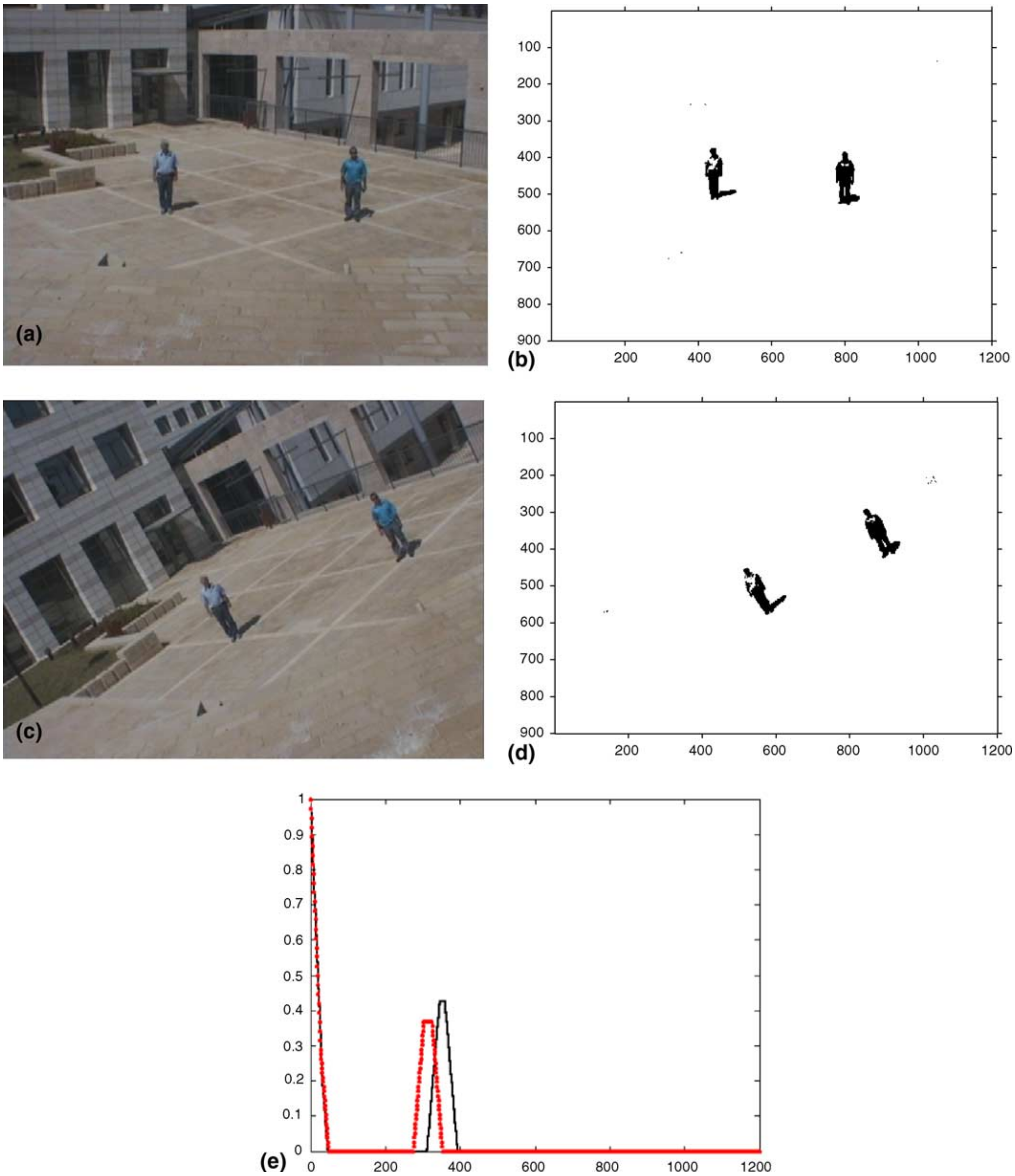

Fig. 6. (a) Original image. (b) 6(a) after background subtraction. (c) Image rotated by 30°. (d) 6(c) after background subtraction. (e) The co-occurrence of 6(b) (dashed line) and of 6(d) (solid line) after projecting and thresholding.

not the sum of the descriptors) aids in improving the detection discrimination capability. This is depicted in Figs. 5 and 6. Fig. 5a presents an image containing two objects. The difference between a background and an input scene is seen in Fig. 5b. Fig. 5c presents the image rotated around the vertical axis (pan) by an angle of approximately 50°. The difference between the background and the rotated input scene is presented in Fig. 5d. Fig. 5e presents the co-occurrence of Fig. 5a performed over row number 183 (red dashed line) [1]

[1] For interpretation of color in Figs. 5 and 6, the reader is referred to the web version of this article.

and of Fig. 5c calculated over row 172 (black line). From the position of the peaks in Fig. 5e one may extract the rotation angle between the two scenes (the peaks of Fig. 5e appear at pixel 50 and 90). Using those parameters and following notations of Eq. (15) one obtains $\theta = 56.25°$. Fig. 6a presents a reference input image. The difference between a background and an input scene is seen in Fig. 6b. Fig. 6c presents the image rotated around the optical axis (roll) by an angle of approximately 30°. The difference between the background and the rotated input scene is presented in Fig. 6d. Fig. 6e depicts the co-occurrence of Fig. 6b (red dashed line) and Fig. 6d (black line) after they

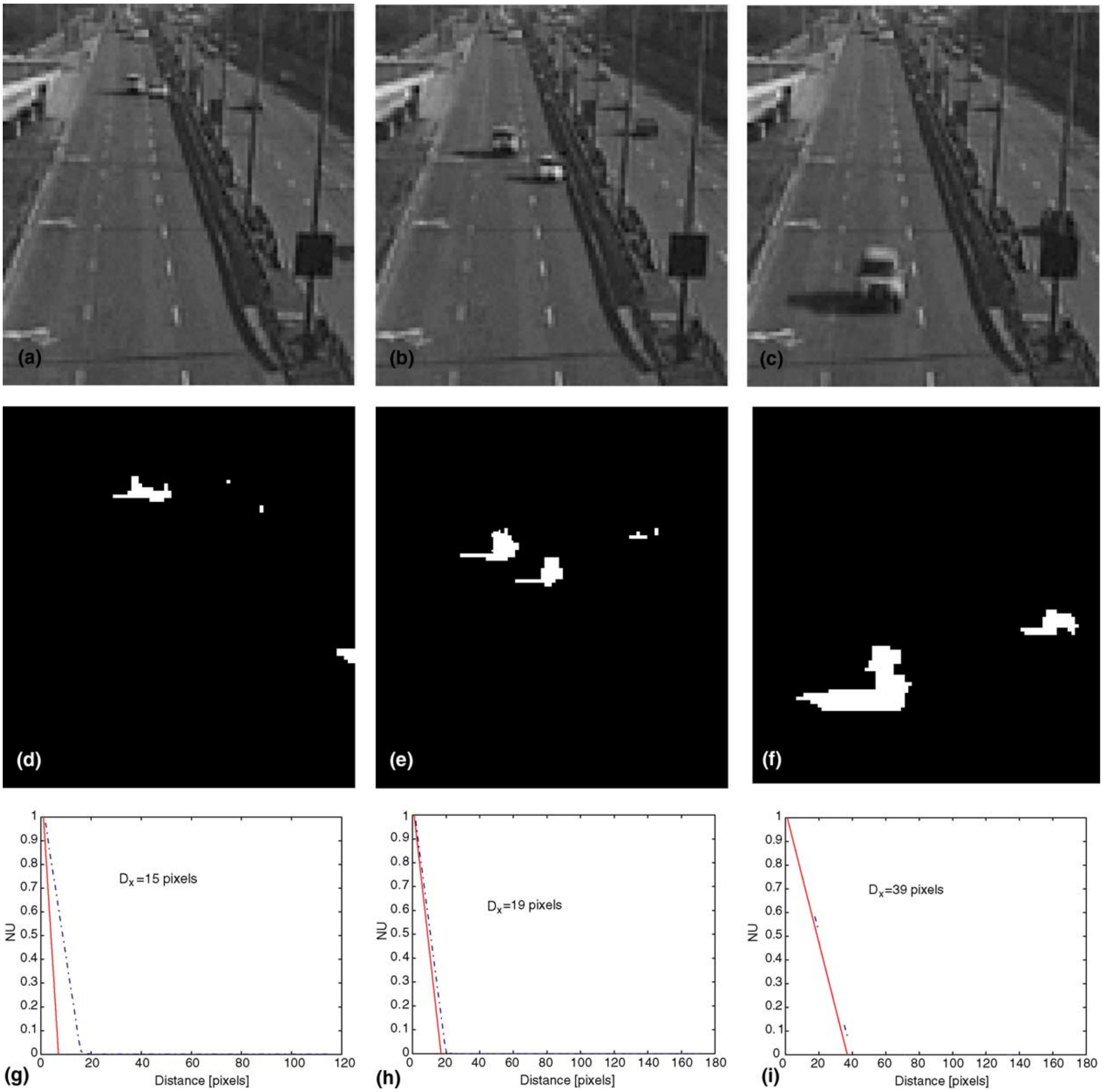

Fig. 7. (a)–(c) Three video frames. (d)–(f) The result obtained after background subtraction and filtering. (g)–(i) The real (solid line) and the estimated (dashed line) co-occurrence descriptors.

were projected over the horizontal axis and a threshold of 40% was applied. From the position of the peaks in Fig. 6e one may extract the rotation between the two scenes since the peaks position corresponds to the distance between the two objects in the images. The peaks appear at pixels 310 and 350 respectively and thus using the notations of Eq. (15) one may extract the rotation angle between scenes of Fig. 6a and c which results with: $\theta = 27.66°$.

Let us now use the co-occurrence descriptor for extracting the time to impact parameter. The equation that describes it may be formulated as:

$$\tau = \frac{y}{\dot{y}} = \frac{Z}{\dot{Z}} \tag{16}$$

where y is a y coordinate of the moving object (in pixels) which can be calculated with a co-occurrence vector. Z is the distance to the object (which is unknown). Fig. 7a–c depicts three video frames taken at 5.48 s, 8 s and 10.52 s respectively. After background subtraction and filtering one obtained the results presented in Fig. 7d–f. After computing the co-occurrence descriptor estimation using the described approach one obtains the plots presented in Fig. 7g–i. Those figures present the co-occurrence descriptor obtained from the images (solid line) and the one obtained after estimating object dimensions (dashed line). The best fit from the co-occurrence descriptor was obtained for objects of: 15, 19 and 39 pixels respectively for objects from Fig. 7d–f. In this case the true co-occurrence descriptor and the estimated one had the best coinciding. Using this information for time to contact estimation yields:

$$\begin{aligned} \tau_1 &= \frac{y_1}{\dot{y}_1} = \frac{y_1}{\frac{y_1 - y_0}{t_1 - t_0}} = \frac{19}{\frac{19-15}{8-5.48}} = 11.97\,[\mathrm{s}] \\ \tau_2 &= \frac{y_2}{\dot{y}_2} = \frac{y_2}{\frac{y_2 - y_1}{t_2 - t_1}} = \frac{39}{\frac{39-19}{10.52-8}} = 4.91\,[\mathrm{s}] \end{aligned} \tag{17}$$

4. Conclusions

In this paper we have defined the co-occurrence descriptor vector as a tool for obtaining motion characteristics. The new tool is a non-linear processing operation having the advantage of requiring low computation complexity. In addition it may be a good descriptor for scaling, rotation and shifting obtained in the scene. Using a single descriptor one may extract the positions and the dimensions of the various objects in the scene. The rotation cause to different scales and shifts in the different co-occurrence vectors. From the ratio of those changes one may obtain the rotation parameters. The scaling of the object causes to the scale of its features and a corresponding scale of the co-occurrence vector. This scale may help to extract the time to impact parameter. The rotation of the object may be obtained by using several co-occurrence vectors (at least two) computed in different directions in the image plane. Experimental validation demonstrated the potential of the new descriptor.

References

Aloimonos, Y., Duric, Z., 1994. Estimating the heading direction using normal flow. Internat. J. Comput. Vision 13, 33–56.

Ancona, N., Poggio, T., 1993. Optical flow from 1-d correlation: Application to a simple time-to-crash detector. In: Proc. DARPA Image Understanding Workshop, pp. 673–682.

Boyce, J.F., Protheroe, S.R., Haddon, J.F., 1992. A relaxation computation of optic flow from spatial and temporal co-occurrence matrices. In: Internat. Conf. on Pattern Recognition, pp. 594–597.

Chen, P.C., Pavlidis, T., 1979. Segmentation by texture using a co-occurrence matrix and a split-and-merge algorithm. Comput. Graphics Image Process. 10, 172–182.

Cipolla, R., Zisserman, A., 1992Qualitative Surface Shape From Deformation of Image Curves, vol. 8, pp. 53–69.

Conners, R.W., Trivedi, M.M., Harlow, C.A., 1984. Segmentation of high resolution urban scene using texture operators. In: Computer Vision, Graphics and Image Processing CVGIP(25), vol. 25, pp. 273–310.

Fablet, R., Bouthemy, P., 2003. Motion recognition using non parametric image motion models estimated from temporal and multiscale cooccurrence statistics. IEEE Trans. Pattern Anal. Machine Intell. 25 (12), 1619–1624.

Fermuller, C., Aloimonos, Y., 1995. Direct perception of three-dimensional motion from patterns of visual motion. Science 270, 1973–1976.

Haralick, R.M., 1979. Statistical and structural approaches to texture. Proc. IEEE 67 (5), 786–804.

Haralick, R.M., Shanmugam, K., Dinstein, I., 1973. Textural features for image classification. IEEE Trans. Systems Man Cybernet., 610–621.

Kovalev, V., Petrou, M., 1996. Multidimensional co-occurrence matrices for object recognition and matching. Graph. Models Image Process. 58, 187–197.

Parkkinen, J., Selkainaho, K., Oja, E., 1990. Detecting texture periodicity from the co-occurrence matrix. Pattern Recognition Lett. 11, 43–50.