AN IMAGE PROCESSING ALGORITHM
Saving valuable time in a sequence of frames analysis

E. Karvelas, D. Doussis and K. Hrissagis
Zenon S.A., Kanari 5, 15354 Glyka Nera, Greece

Keywords: image processing, motion detection.

Abstract: This paper describes a new algorithm to detect moving objects in a dynamic scene based on statistical analysis of the greyscale variations on a sequence of frames which have been taken in a time period. The main goal of the algorithm is to identify changes (e.g. motion) while coping with variations on environmental changing conditions without being necessary to perform a prior training procedure. In this way, we use a pixel level comparison of subsequent frames in order to deal with temporal stability and fast changes. In addition, this method computes the temporal changes in the video sequence by incorporating statistical results and it is less sensitive to noise. The algorithm's goal is not to detect motion but rather to filter out similar frames in a sequence of frames, thus making it a valuable tool for those who would like to evaluate and analyze visual information obtained from a captured video frames. Finally, experimental results and a performance measure establishing the confidence of the method are presented.

1 DESCRIPTION OF THE ALGORITHM

The developed algorithm identifies those grayscale frames with different content form the immediate previous ones in a sequence of frames. The algorithm has been tested with video frames with rate 1 frame/ second.

The algorithm marks every frame as hidden or shown. After M frames there is a number of M^{shown} frames marked as shown and a number of M^{hidden} frames marked as hidden, where $M = M^{shown} + M^{hidden}$. The M^{shown} frames include the M^{shown}_{True} as the correctly marked frames and the M^{shown}_{False} as mistakenly marked ones. Thus, we have the following equation $M^{shown} = M^{shown}_{True} + M^{shown}_{False}$. The same apply to the hidden frames, i.e. $M^{hidden} = M^{hidden}_{True} + M^{hidden}_{False}$.

The algorithm assigns different significance to the M^{hidden}_{False}, M^{hidden}_{False} frames. In particular, it is allowable for the algorithm to show frames which should be hidden while it is not acceptable to hide frames which should be shown. This algorithm is useful in projects in which it is desirable to eliminate the M^{hidden}_{False} frames while it is acceptable to keep the M^{shown}_{False} frames. The M^{shown}_{False} frames although are not critically important, they affect the efficiency of the image processing.

In the implementation of the algorithm we used videos with the frame rates of 1 frame/ second. In order to test the algorithm we used the following different sets of captured frames:

– Set A: It included frames, which exhibit no motion and no difference in the light condition.
– Set B: It included frames, which exhibit differences in less than 10% of their contents and no difference in the light condition.
– Set C: It included frames, which exhibit differences in more than 10% of their contents and no difference in the light condition.
– Set D: It included frames, which exhibit no motion and difference due to illumination variance.

The developed algorithm has two stages of image analysis. During the first stage the algorithm searches for differences between the two frames on a pixel level and marks blocks of 9×9 pixels that have significant changes.

Karvelas E., Doussis D. and Hrissagis K. (2005).
AN IMAGE PROCESSING ALGORITHM - Saving valuable time in a sequence of frames analysis.
In *Proceedings of the Second International Conference on Informatics in Control, Automation and Robotics - Robotics and Automation*, pages 232-236
DOI: 10.5220/0001170502320236

During the second stage the algorithm uses the patterns of differences, which were found in the first stage in order to decide if it will show or hide the frame.

1.1 First Stage Analysis

In order to eliminate the noise the algorithm converts the frame from the initial array of $h \times w$ $pixel_{1\times1}$ in an array of $\left(h/3\right)\times\left(w/3\right)$ $pixel_{3\times3}$. The grayscale value of the $pixel_{3\times3}$ of the i, j position (where i is the row and j is the column) is given by:

$pixel^{ij}_{3\times3} = \sum_{n=i-1}^{i+1}\sum_{m=j-1}^{j+1} pixel^{nm}_{1\times1}$, where $pixel^{nm}_{1\times1}$ is the grayscale value of the original pixel at n, m position.

Thus, during the first stage of the image analysis the algorithm divides the two compared frames to form a grid $\left(h/9\right)\times\left(w/9\right)$ which is consisted of areas of 3x3 $pixel_{3\times3}$.

The corresponding areas between the two consecutive frames are compared as shown in the following equations so that the differences between them, which are not attributed to illumination variations, can be quantized. For each set of compared frames we have the following variables:

$$D^{ij}_{normal} = \left[\frac{pixel^{ij}_{3\times3}}{\sum_{n=i-1}^{i+1}\sum_{m=j-1}^{j+1} pixel^{ij}_{3\times3}}\right]_{FirstFrame} - \left[\frac{pixel^{ij}_{3\times3}}{\sum_{n=i-1}^{i+1}\sum_{m=j-1}^{j+1} pixel^{ij}_{3\times3}}\right]_{SecondFrame}$$

$$D^{ij}_{negative} = \left[\frac{(S_{\max} - pixel^{ij}_{3\times3})}{\sum_{n=i-1}^{i+1}\sum_{m=j-1}^{j+1}(S_{\max} - pixel^{ij}_{3\times3})}\right]_{FirstFrame} - \left[\frac{(S_{\max} - pixel^{ij}_{3\times3})}{\sum_{n=i-1}^{i+1}\sum_{m=j-1}^{j+1}(S_{\max} - pixel^{ij}_{3\times3})}\right]_{SecondFrame}$$

The $S_{\max}$ value is equal to the maximum value which the variable $pixel_{3\times3}$ can take. Taking into that an 8-bit grayscale video is used, this value is equal to: $S_{\max} = 3\times3\times255$.

The values of the variable D^{ij}_{normal} as well as the $D^{ij}_{negative}$ versus the value of a $\overline{pixel}^{ij}_{3\times3} \equiv \left(\left[pixel^{ij}_{3\times3}\right]_{FirstFrame} + \left[pixel^{ij}_{3\times3}\right]_{SecondFrame}\right)/2$

are shown in Figure 1 for set A frames and in Figure 2 for set B frames. The variable D^{ij} is defined as

$$D^{ij} \equiv \begin{cases} D^{ij}_{normal} & \overline{pixel}^{ij}_{3\times3} \geq S_{\max}/2 \\ D^{ij}_{negative} & \overline{pixel}^{ij}_{3\times3} < S_{\max}/2 \end{cases}$$

Figure 1 and Figure 2 reveal that we can differentiate between the set A and set B frames when we compare the “bright” areas ($\overline{pixel}^{ij}_{3\times3} \geq S_{\max}/2$) and the negatives of the dark” areas ($\overline{pixel}^{ij}_{3\times3} < S_{\max}/2$).

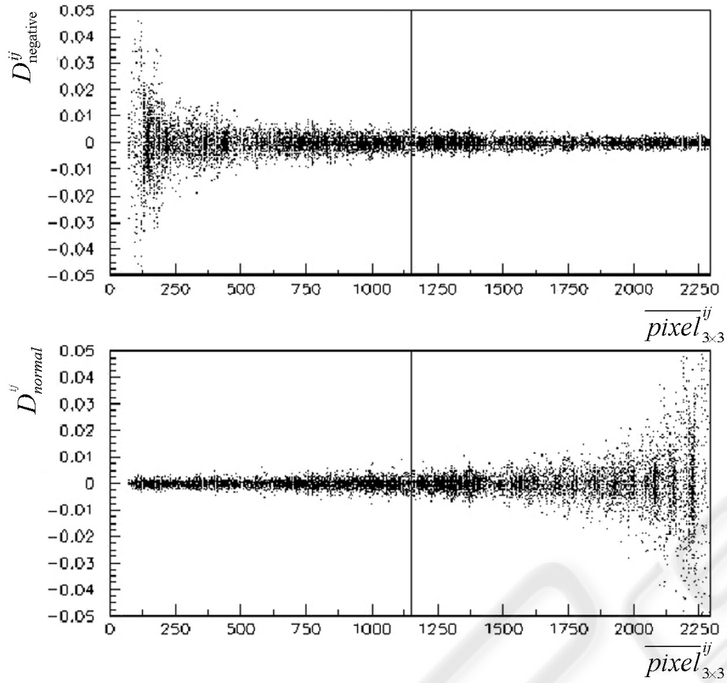

Figure 1: The values of the variable D^{ij}_{normal} as well as the $D^{ij}_{negative}$ versus the value of a $\overline{pixel}^{ij}_{3\times3}$ for the set A frames

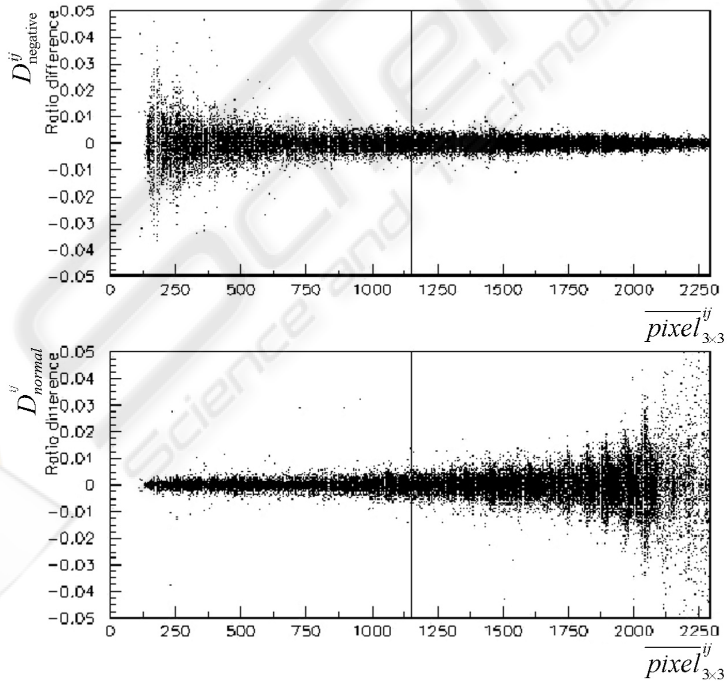

Figure 2: The values of the variable D^{ij}_{normal} as well as the $D^{ij}_{negative}$ versus the value of a $\overline{pixel}^{ij}_{3\times3}$ for the set B frames

1.2 Second Stage Analysis

Next a further analysis of the identified as “different” areas is necessary in order to eliminate the noise factor and identify human motion. In particular every “different” area is analyzed by taking into consideration the behaviour of its surrounding areas. Thus, the notion of the “clusters” is introduced. The concept of the clusters derives from the fact that human motion presents a relatively sizable motion. During the second stage of the visual analysis not only the differences between the compared frames is examined but also the relation of these differences with the “neighbour” surrounding differences. In particular we statistically analyze the clusters of differences which are formed. The size of these clusters varies along with their distribution for the four cases which we consider in this paper. Analysis of the distribution results in the graph shown in Figure 3.

As it is shown in Figure 3 we can distinguish between the three cases of motion and ambient light. These different distributions help us to establish additional criteria for identifying human “motion” in the sequence of frames which was our main target in the study case analysis of the developed method.

1.3 Results

In order to test the algorithm we have used two different sets of data. In the first one the data was captured from 6 different cameras which were placed inside a collapsing building. In the second set the data was captured during building evacuation. The results are shown in the Table 1.

2 CONCLUSION

Image processing has typically focused on only accuracy or only speed. This algorithm represents a good compromise between speed and accuracy. The method is also very robust in presence of noise. It yielded reasonable results for fairly low signal to noise levels. In addition it does not require any training procedure. The developed algorithm is a valuable tool to the hands of those who want to process a vast number of frames which have been captured with a time difference of one second or more and would like to focus only on those frames which provide useful information.

ACKNOWLEDGMENT

This study was partially funded by the European Commission under the contract No IST-2000-29401-Project LOCCATEC.

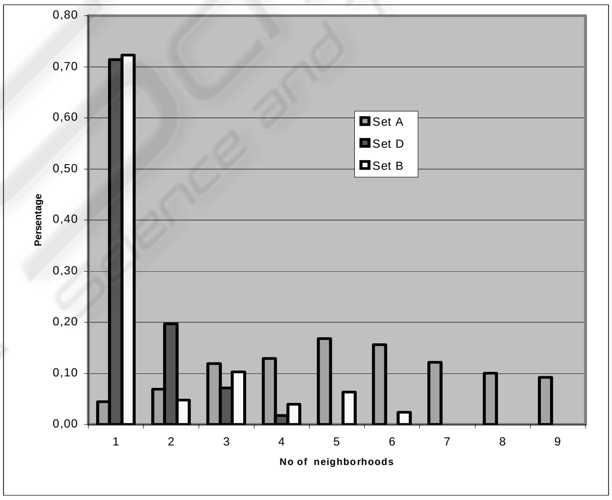

Figure 3: The distribution of the "motion" detection clusters

Table 1: The results of the implementation of the developed algorithm

LCD set	M^{hidden}_{False}	M^{hidden}_{True}	M^{shown}_{False}	M^{shown}_{True}	M
Collapsed building	0	509	3	120	632
Evacuated building	0	67	2	261	330

REFERENCES

LOCCATEC Low Cost Catastrophic Event Capturing, , Contract No IST-2000-29401

Y. Amit and A. Kong. Graphical templates for model registration. IEEE Transactions on Pattern Analysis and Machine Intelligence, 18:225–236, 1996.

A. Blake and M. Isard. 3d position, attitude and shape input using video tracking of hands and lips. In Proc. ACM Siggraph, pages 185–192, 1994.

C. Bregler and J. Malik. Tracking people with twists and exponential maps. In Proc. IEEE CVPR, pages 8–15, 1998.

C. Chow and C. Liu. Approximating discrete probability distributions with dependence trees. IEEE Transactions on Information Theory, 14:462–467, 1968.

T. Cover and J. Thomas. Elements of Information Theory. John Wiley and Sons, 1991.

N. Friedman and M. Goldszmidt. Learning bayesian networks from data. Technical report, AAAI 1998 Tutorial, http://robotics.stanford.edu/people/nir/tutorial/, 1998.

D. Gavrila. The visual analysis of human movement: A survey. Computer Vision and Image Understanding, 73:82–98, 1999.

L. Goncalves, E. D. Bernardo, E. Ursella, and P. Perona. Monocular tracking of the human arm in 3d. In Proc. 5th Int. Conf. Computer Vision, pages 764–770, Cambridge, Mass, June 1995.

I. Haritaoglu, D. Harwood, and L. Davis. Who, when, where, what: A real time system for detecting and tracking people. In Proceedings of the Third Face and Gesture Recognition Conference, pages 222–227, 1998.

S. Ioffe and D. Forsyth. Human tracking with mixtures of trees. In International Conference on Computer Vision, pages 690–695, July 2001.

M. Jordan, editor. Learning in Graphical Models. MIT Press, 1999.

M. Meila and M. Jordan. Learning with mixtures of trees. Journal of Machine Learning Rearch, 1:1–48, 2000.

R. Polana and R. Nelson. Detecting activities. In DARPA93, pages 569–574, 1993.

J. Rehg and T. Kanade. Digiteyes: Vision-based hand tracking for human-computer interaction. In Proceedings of the workshop on Motion of Non-Rigid and Articulated Bodies, pages 16–24, November 1994.

K. Rohr. Incremental recognition of pedestrians from image sequences. In Proc. IEEE Conf. Computer Vision and Pattern Recognition, pages 8–13, New York City, June, 1993.

Y. Song, X. Feng, and P. Perona. Towards detection of human motion. In Proc. IEEE CVPR 2000, volume 1, pages 810–817, June 2000.

Y. Song, L. Goncalves, E. D. Bernardo, and P. Perona. Monocular perception of biological motion in johansson displays. Computer Vision and Image Understanding, 81:303–327, 2001.

C. Tomasi and T. Kanade. Detection and tracking of point features. Tech. Rep. CMU-CS-91-132,Carnegie Mellon University,1991.

S.Wachter and H.-H. Nagel. Tracking persons in monocular image sequences. Computer Vision and Image Understanding, 74:174–192, 1999.

M. Weber, M. Welling, and P. Perona. Unsupervised learning of models for recognition. In Proc. ECCV, volume 1, pages 18–32, June/July 2000.

Y. Yacoob and M. Black. Parameterized modeling and recognition of activites. Computer Vision and Image Understanding, 73:232–247, 1999.

Bock, P., The Emergence of Artificial Cognition: an Introduction to Collective Learning, Singapore: World Scientific, 1993

Bock, P., Klinnert, R., Kober, R., Rovner, R. and Schmidt, H. “Gray Scale ALIAS”, IEEE Special Trans. Knowledge and Data Eng., vol 4, no 2, Apr 1992

Haynes, S.M., and Jain, R. “Detection of Moving Edges,” Computer Vision, Graphics and Image Processing, vol 21, no 3, Mar 1982

Horn, B.K.P., and Schunck, B.G. “Determining Optical Flow,” Artificial Intelligence, vol 17, no 1-3, Aug 1981

Howard, C.G. and Kober, R. “Anomaly Detection in Video Images”, Proceedings of the Fifth Neuro-Nimes Conference: Neural Networks and their Applications, Nimes, France, Nov 1992

Hubshman, J. and Achikian, M. “Detection of Targets in Terrain Images with ALIAS”, Proc. Twenty-Third Annual Pittsburgh Conf. on Modeling and Simulation, Apr 1992

Kober, R., Bock, P., Howard, C., Klinnert R., and Schmidt, H. (1992). “A Parallel Approach to Signal Analysis”, Neural Network World, vol 2, no 6, Dec 1992

Nagel, H. “Displacement Vectors Derived from Second-Order Intensity Variations in Image Sequences,” Computer Vision, Graphics, and Image Processing, vol 21, no 1, Jan 1983

Schalkoff, R.J., and McVey, E.S. “A Model and Tracking Algorithm for a Class of Video Targets,” IEEE Trans. on Pattern Analysis and Machine Intelligence, vol 4, no 1, pps 2-10, Jan 1982

Schmidt, H., and Bock, P. “Traffic Jam Detection Using ALISA”, Proc. of the IX Int. Symposium on Artificial Intelligence, Nov 1996