Detecting Motion in Single Images Using ALISA

From: Proceedings of the Eleventh International FLAIRS Conference. Copyright © 1998, AAAI (www.aaai.org). All rights reserved.

Taras P. Riopka and Peter Bock

Department of Electrical Engineering and Computer Science
The George Washington University, Washington DC USA
riopka@seas.gwu.edu pbock@seas.gwu.edu

Abstract

Motion analysis often relies on differencing operations that inherently amplify noise and are hindered by the spatial correspondence problem. An alternative approach is proposed using ALISA (Adaptive Learning Image and Signal Analysis) to detect differences in types of motion by classifying the imaging effects of the motion in single frames. With an appropriate feature set, the ALISA engine accumulates a multi-dimensional histogram that estimates the probability density function of a feature space and uses the result as a basis for classification. As a function of image sampling rate and the scale of image structures, the ALISA engine was able to discriminate between a slow moving and fast moving object with a confidence greater than 99%.

Introduction

Motion analysis typically requires processing of temporally sequential frames of image data and is generally dependent on the order of frame processing. Many standard approaches [Haynes and Jain 1982][Horn and Schunck 1981][Nagel 1983][Schalkoff and McVey 1982] rely on analysis of sequential frames, often using a differencing operation that inherently amplifies noise. Further, point correspondence of the same, but moving object between frames is often problematic. In real scenes, an object can often be differentiated by the characteristics of its motion alone, for example, a slow moving object vs. a fast moving object, circular motion vs. linear motion, an approaching object vs. a receding object, *etc.* If the effects of a type of motion can manifest itself in the texture of an image, then single frames might be sufficient for identifying the type and amount of motion, given sufficient statistical evidence.

With this in mind, an alternative approach is proposed using **ALISA** (Adaptive Learning Image and Signal Analysis), an adaptive image classification engine based on collective learning systems theory [Bock 1993]. Using an appropriate set of features, the ALISA engine accumulates a multi-dimensional histogram that estimates the probability density function of a feature space and uses the result as a basis for classification. Feature extraction in the ALISA engine is based on an **analysis token**, a small window from which feature values are computed, scaled, and quantized. The results are concatenated into a feature vector and used to index into a multi-dimensional **histogram**. Because the analysis token is applied to each image in a fully overlapped manner, a single input image yields a very large number of feature vectors. During training, the weight for each bin indexed by the feature vector is incremented, yielding a relative frequency of occurrence. During testing, the weight in each bin is normalized into a feature-vector conditional probability that represents the **normality** of the feature vector. After processing an entire image, the normalities for each pixel are quantized and assembled into a **normality map**, which is spatially isomorphic with the original image.

This paper addresses the research question: **Can the defining characteristics or type of motion be effectively and efficiently encoded in the multi-dimensional histogram accumulated by the ALISA engine from single frames of an object in motion?** Only features computed within a single frame, not across successive frames, are used to configure the ALISA engine, under the assumption that objects moving at one speed will exhibit feature value distributions that are significantly different from those exhibited by objects moving at a different speed. Given the nature of the collective learning paradigm, this condition requires that the ALISA engine be trained on a sufficient number of examples spanning the entire range of object motion to be learned. Clearly, training on all possible images depicting a particular type of motion, especially if multiple objects are to be considered, is not a feasible approach. However, if only a subset of images is to be used, two subordinate research questions are: **What parameters determine a minimum subset of images**; and **What are the optimal configuration parameters for the ALISA engine to learn to discriminate between two types of motion?**

Clearly, image sampling rate and the speed of object motion determine the minimum number of images of the moving object that must be captured for processing and subsequent training. In addition, because only a small subset of all the possible images of object motion are used for training, to enable the ALISA engine to detect object motion on which it was not trained, a mechanism for **image interpolation** was postulated. Thus, the scale of image structures was an important parameter affecting the ability of the ALISA engine to interpolate between frames.

The objective of this research was, therefore, to measure the effect of image sampling rate and the scale of image structures on the ability of the ALISA engine to discriminate between slow moving and fast moving objects moving at constant speed under approximately the same lighting and background conditions.

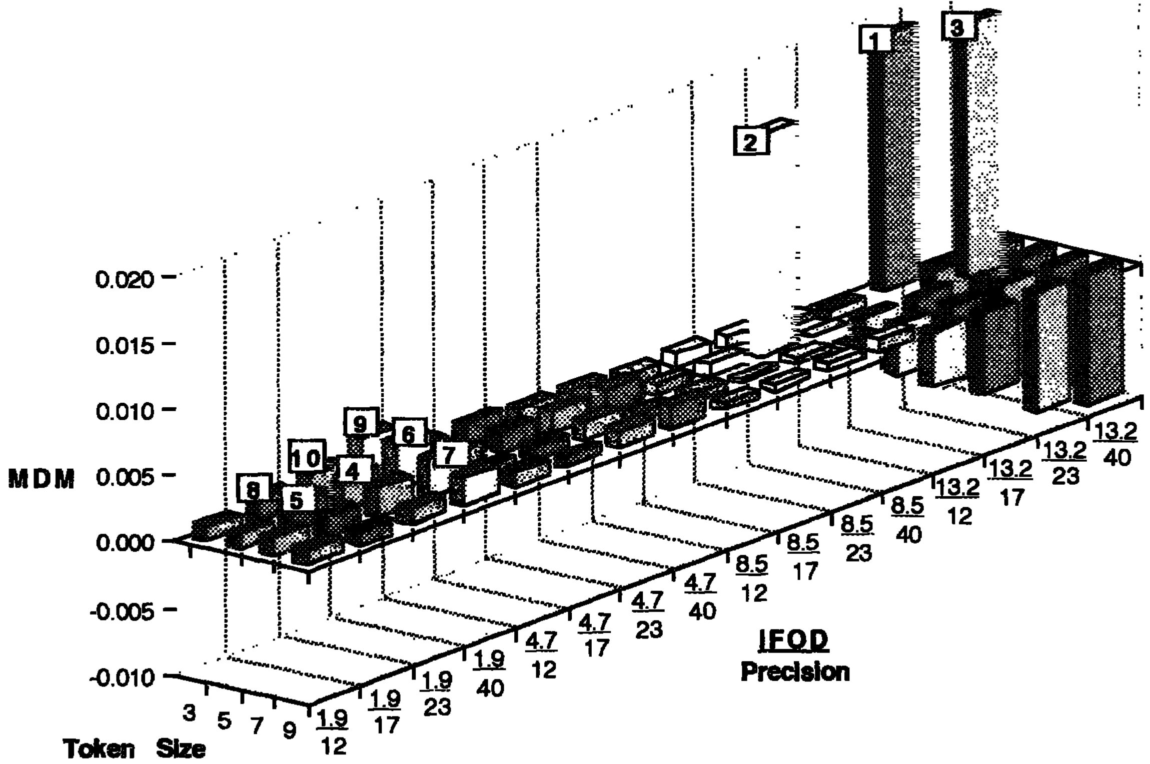


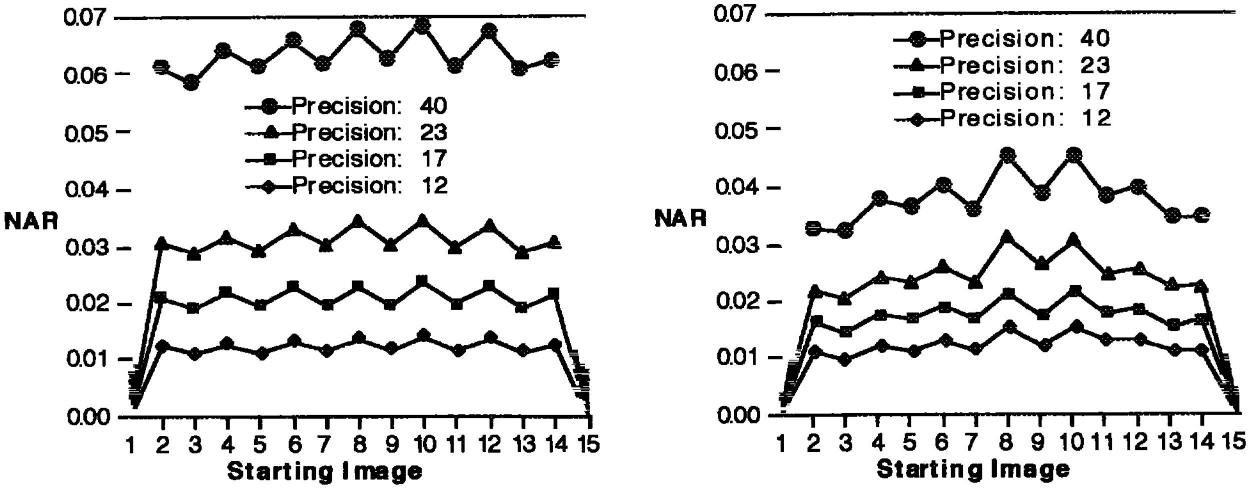
Figure 2: Motion discrimination by the ALISA engine as a function of the experiment factors. The numbered peaks show which combinations of factors enable the ALISA engine to discriminate between walking and running motion with very high confidence.

The three most significant operating points (labeled 1, 2, and 3 in Figure 2) are given by the 3-tuples (IFOD, position-precision, analysis-token-size): (13.2, 17, 3), (8.5, 23, 5) and (13.2, 23, 5). These operating points resulted in NAR values for the running person which exceeded their corresponding NAT values with a confidence greater than 99%, clearly identifying the motion as non-normal. Thus, the null hypothesis can be rejected with 99% confidence. **This strongly suggests that under the given conditions, there is an optimal set of factors that enables the ALISA engine to discriminate between walking and running motion.** Seven other operating points that exhibit a confidence for rejecting the null hypothesis of greater than 80% are labeled 4 through 10 in Figure 2 in order of decreasing confidence.

Figure 3 shows four examples of graphs of the average NAR for sub-sampled-control-sequences plotted against the corresponding starting image (phase) in the base sequence. Recall that the sub-sampled-control-sequences correspond to images of object motion out of phase with the images used for training. Sub-sampled-control-sequences close to either the left or right side of each graph correspond to images almost in phase with the training images, while those in the middle correspond to images very out of phase with the training images.

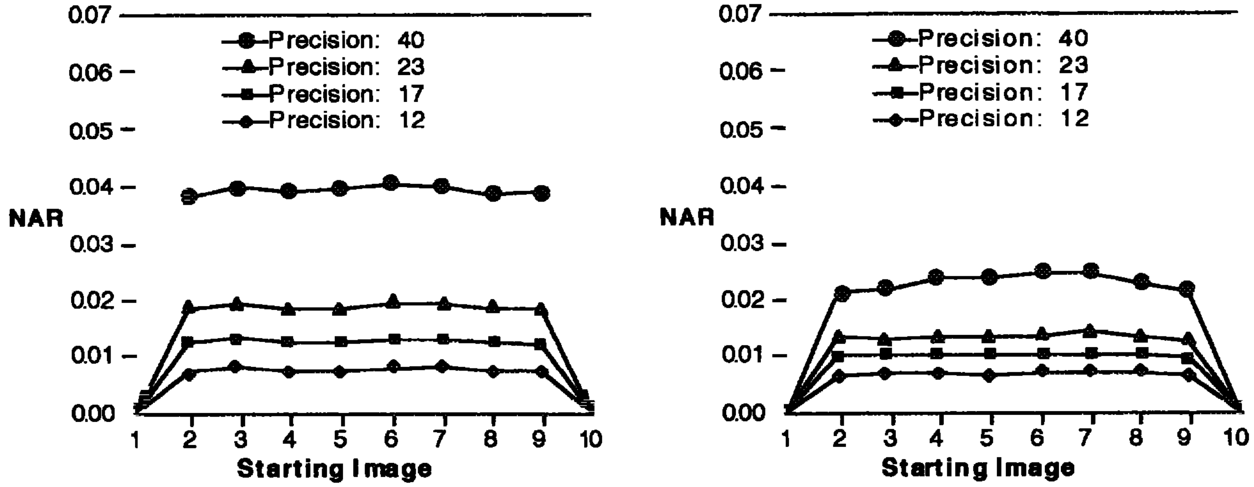
In all cases, the average non-normality area ratio (NAR) increases as the precision used for the position features increases. Average NAR also increases as the analysis-token-size decreases. Both of these observations are consistent and reasonable, since a higher feature resolution necessarily results in a larger histogram, decreasing the proportion of the feature space that can be effectively explored during training with a fixed number of images. In other words, smoothing (using lower precision and larger analysis-token-size) results in a more thorough exploration of the corresponding partition of the feature space, enabling the ALISA engine to *interpolate* between frames more accurately.

In the sub-sampled-control-sequences, although lower position-precisions and larger analysis-token-sizes worked together to smooth images, resulting in lower average NAR values, additional informal experiments revealed that this was not the case when clearly anomalous objects were present. In such test images (not shown) lower precision for the position features tended to reduce noise in the normality maps, while larger analysis-token-size tended to increase noise in the normality maps. Low position-precision, therefore, consistently reduced noise and resulted in superior image interpolation, allowing a lower image sampling rate and hence fewer images for object motion training.



(a) IFOD = 13.2; Analysis-Token-Size = 3

(b) IFOD = 13.2; Analysis-Token-Size = 9



(c) IFOD = 8.5; Analysis-Token-Size = 3

(d) IFOD = 8.5; Analysis-Token-Size = 9

Figure 3: Examples of average NAR for sub-sampled-control-sequences as a function of phase

Although a larger analysis-token-size improved image interpolation and reduced the NAT obtained from sub-sampled-control-sequences, counterintuitively it also increased the noise in the associated normality maps. This may be due to the possibility that smoothing of anomalous regions only increases the area over which the points are anomalous, which works in opposition to the benefits obtained from lower position-precision. Except for the increased computational requirements, this can be advantageous, because it suggests that with larger analysis-token-size, conjunctive benefits result, *i.e.*, anomalies are emphasized and image interpolation is enhanced.

It is interesting to note that the variation of average NAR with image phase was contrary to expectation. The average NAR for images nearly in phase with the training images was almost the same as the average NAR of images farthest away in phase from the training images. This is true for most combinations of research factors, except for combinations of large IFOD and large analysis-token-size (see Figure 3b and 3d). The effect appears to increase with position-precision and becomes significant with a confidence greater than 75% when IFOD is greater than twice the resolution of the position coordinates. Provided this condition is true, this implies that the ALISA engine is capable of interpolating equally well (or equally poorly) *regardless of the phase of the missing images.*

Finally, the peculiar periodic fluctuations in average NAR for IFOD=13.2 (see Figure 3a and 3b) arise from aliasing, because this IFOD is close to the period of the person's walk.

To eliminate aliasing and reduce ambiguities in the results, it may be useful to repeat this work with objects that are not symmetrical and do not exhibit periodicities in motion. However, because motion analysis often involves real scenes with all kinds of motion, this may not always be possible. Unfortunately, it is unclear if the optimum operating points obtained for IFOD=13.2 are due to an interesting combination of factors, or merely coincidences arising from aliasing. The optimum operating point for IFOD=8.5, however, may be quite real. If so, this begs the question of why the subtle distinctions between running motion and walking motion would be more apparent to the ALISA engine with a more sparsely populated histogram (IFOD=8.5 and IFOD=13.2) than to the ALISA engine with a more densely populated histogram (IFOD=1.9 and IFOD=4.9). The results of this work clearly emphasize the need for as low a value of IFOD as possible to reduce the possibility of aliasing, but warrant further investigation.

References

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