# Exploration of The Vector Fusion Method for Basic Behavior Unit Segmentation from Visual Data

Xinwei Xue and Thomas C. Henderson

Abstract-It becomes an increasingly important research area to automatically analyze object behaviors from visually captured data (e.g., motion) or video recordings. Among this research, the automatic basic behavior unit (BBU) discovery is very important. In this paper, we explore the applicability of the vector fusion (SBP) method, a multi-variate vector visualization technique, in BBU segmentation. This technique is also inherently a data dimension reduction technique: it reduces the multiple dimensional data into two dimensional (SBP)space, and the spatial and temporal analysis in SBP space can help discover the underlying data groups. We present results on a physical system and a synthetic mouse-in-a-cage scenario. The vector fusion method provides a good distinction and interpretation for the bouncing ball example and the analytical data from the synthetic video simulation upon certain selected features. Our experiments show that several factors influence the effectiveness of the vector fusion method in BBU segmentation. The temporal analysis in SBP space seems to be very effective to detect periodic BBUs. Overall, this method is simple and effective for grouping BBUs with periodic motion.

## I. INTRODUCTION

It has become an increasingly important research area to automatically analyze object behaviors from visually (e.g. motion) captured data or video recordings. Humans and vehicles have been mostly the focus of the visual tracking and behavior understanding research [1], [2], [3], [4]. Due to the increasing need to study animal behaviors in areas of biology, pharmacology, toxicology, entomology and animal welfare, and the popular use of video recordings, the automatic animal behavior analysis from visual data is drawing more and more attention [5], [6]. In the area of robot control, it is also desired for robots to automatically learn behaviors from motion capture data [7], [8], [9].

In an automated behavior analysis system, the basic behavior unit (BBU) classification (or segmentation) is one important task [10]. Usually the sequences of visual data from images need first to be grouped into BBUs [11], or primitive (atomic) behaviors [7], and then complex behaviors are analyzed based upon the relationship between the BBUs and contexts.

Prior to the BBU segmentation step, spatiotemporal features are usually extracted. In the literature, interest points, shape properties of the detected object blobs, contours, or features derived thereby are used to perform BBU classification. Feature extraction itself is an important task. In this paper, for simplicity without loss of generality, we assume the features are readily available, and we focus on applying a new approach, the vector fusion method to BBU segmentation. We use synthetically generated visual data for our experiments.

The rest of the paper is organized as follows: Section II briefly describes the related work; Section III presents the new method for BBU segmentation; Section IV describes our experiments and shows the results; conclusions are drawn in Section V.

## II. RELATED WORK

In the visual surveillance literature, most of the existing techniques extract basic behaviors (or actions) directly based upon one or more features extracted (trajectory, motion, posture, etc.) from the detection and tracking results. Pattern recognition techniques (template matching, clustering analysis) are used to classify the video sequence into actions or behavior units, as discussed in the survey papers [1], [3], [4], [2]. These methods are effective in their specific applications. The idea is to utilize all the available distinguishing features to perform classification.

Recently, new approaches based on data (or feature) variance or similarity analysis have been developed for discovering BBUs: PCA-related techniques [7], [9], and affinity graph-based techniques [12], [13], [11]. The former captures the variance in a dataset in terms of principle components, and the latter utilize the degree of similarity between the data elements. The commonalities of these two approaches lies are that, first a covariance matrix (for PCA) or affinity matrix (for affinity method) is constructed, then Singular Value Decomposition (SVD) is performed to derive eigenvalues and eigenvectors. Segmentation are performed upon the eigenvector corresponding to the largest eigenvalue.

In most of the BBU segmentation methods, the feature data usually has a large dimension, which usually makes the algorithms computationally expensive. Hence feature dimension reduction is often applied before applying BBU segmentation algorithms. The algorithm we are going to present here, the vector fusion method [14], [15], is inherently one such technique: it reduces however large dimension to a two dimensional space, which can help discovering the underlying structure of the data. This method is originally proposed as a method for visualizing the structure of multiple dimensional data, and has also been applied in characterizing and measuring data. Here we propose to explore its applicability in grouping behavioral data.

Xinwei Xue is a Ph.D. candidate at School of Computing, University of Utah, Salt Lake City, Utah 84112, USA xwxue@cs.utah.edu

Thomas C. Henderson is with the Faculty of School of Computing, University of Utah, Salt Lake City, Utah 84112, USA tch@cs.utah.edu

## III. THE VECTOR FUSION ALGORITHM FOR BBU SEGMENTATION

In this section, we describe Johnson's vector fusion method (denoted as **SBP** – **S**ingle-point **B**roken-line **P**arallel-coordinate in [14], [15]) and how we apply it in BBU discovery.

The vector fusion method is a vectorized generalization of the parallel coordinates [16] method for visualizing multidimensional datasets, which allows one to see any number of dimensions concurrently by arranging the coordinates parallel to each other. The vector fusion method maps a multi-variate vector into a 2D vector, by adding each element of the row (the multi-variate vector) rotated by some angle to the prior one, and summing the whole row to a singleend-point resultant, as expressed in equation 1.

$$\mathbf{w} = w_1 e^{i\theta_1} + w_2 e^{i\theta_2} + \ldots + w_d e^{i\theta_d}$$
$$= \sum_d w_i cos(\theta_i) + i \sum_d w_i sin(\theta_i)$$
$$= (w_{sumx}, w_{sumy})$$
$$= (SBPx, SBPy)$$
(1)

where

 $\theta_i = (i-1)180^\circ/d$ 

d is the dimension of the multi-variate vector  $w_i$  is the value of the *i*th dimension.

This concept is further demonstrated in Figure 1, which shows how the 4 dimensional vector is 'fused' to form a two-dimensional vector (coordinate). By fusing each element vector of the data, and plotting the final coordinate sequence, this method is able to reveal some underlying structure within the data. The advantage of this method is its simplicity in representing the multiple-dimension vectors. However many dimensions the data element may have, it reduces it to a two dimensional coordinate in SBP space. Johnson has demonstrated its effectiveness in several applications, such as spectral signature identification, medical data analysis, etc. [14], [15].



Fig. 1. Vector Fusion Demonstration (vector of 4 dimensions,  $\alpha = 45^{\circ}$  )

We are interested in BBU segmentation of visually captured data. The data we have are multiple dimensional sequential feature points, either extracted from video sequence, or calculated analytically. By applying the vector fusion method, the multiple dimensional data is reduced to twodimensional points in SBP space. We analyze the 2D SBP points in two ways: one is to directly find the spatial structure of the sequence in the SBP space, i.e., identifying clusters of SBP points; the other is to analyze the temporal properties in the SBP space, and discover motion patterns for different BBUs. This can be considered the training process. Then we can group BBUs based upon the spatial and temporal properties of the SBP points.

In the next section, we are going to present and discuss the results of applying this approach to different datasets, which are based on simulations of a physical system, and an artificial mouse that mimicked the behaviors of a real mouse in a cage scenario.

## **IV. EXPERIMENTAL RESULTS**

We have experimented with the vector fusion method with data derived from two cases: 1) a bouncing ball, and 2) an artificial mouse.

## A. Bouncing Ball

**Data**: In this case, a ball falls down and bounces backs, assuming no friction. A temporal sequence of the ball position and speed is generated by simulation, as shown in Figure 2. The BBUs to be distinguished are 'falling down,' 'bounce,' and 'rising up.' We use the position and velocity of the ball as input feature data (2D), with the length of 100.



Fig. 2. Bouncing Ball Example (position and speed)

**Result**: The result of applying vector fusion method to the bouncing ball is shown in Figure 3. Note that, in this figure, as well as in the Figures 6, 7, 8, 9, 10, 11 and 12, the horizontal axis is the  $SBP_x$  coordinate, and the vertical axis is the  $SBP_y$  coordinate. In the bouncing ball example, the point where the ball reaches its highest position corresponds to the rightmost point (denoted as P1) in Figure 3, the point where the ball has the lowest position corresponds to the upper-left-most point(denoted as P2) in Figure 3, and the point immediately after the lowest position corresponds to the bottom-left-most point (bouncing point, denoted as P3) in Figure 3. The 'falling down' BBU corresponds to the section curve between the P1 and P2, 'bounce' corresponds

to the transition from P2 to P3, and 'rising up' corresponds to the curve from P3 to P1.



Fig. 3. Vector Fusion Result for Bouncing Ball (position and speed). Horizonal  $Axis:SBP_x$ ; Vertical  $Axis:SBP_y$ 

#### B. Artificial Mouse Video Data

We synthesized several clips of mouse-in-cage scenario, where the artificial mouse is constructed with ellipsoids. There are four behaviors simulated in this video, shown in Figure 4:

- **Resting**. No movement. The body and limbs do not move.
- **Exploring**. The body moves in random direction, while the limbs move in such a fashion: the front right and back left leg moves in same pace (same rotating angle), and the front left and back right leg moves in same pace.
- Eat. Reaching up to the 'food' above (represented as a little sphere), and getting down, and repeat up and down.
- **Grooming** Standing on tail with two front legs brushing the head with slight body motion.

This 2000-frame synthetic video sequence consists of 8 rest segments, 4 segments of reaching up, 2 grooming segments, and the rest are exploring segments, as shown in Figure 5.

**Data.** The feature data are obtained in the following two ways:

- Extraction from Synthetic Video Data: First, the artificial mouse blob is tracked and extracted from each frame by simple background subtraction method. Then we calculate the following features: the speed (x,y), aspect ratio, filling ratio, and orientation of the extracted bounding box of the synthetic mouse blob, and the orientation of the mouse. Each feature element is a 5-D vector.
- Direct Analytical Data from Simulation: We use a selection of the following features that are calculated analytically or recorded during simulation: position (x, y, z), speed $(v_x, v_y, v_z)$ , orientation  $(\theta_x, \theta_y, \theta_z)$ , and orientation change $(d\theta_x, d\theta_y, d\theta_z)$  of the body and four limbs of the artificial mouse simulation. Both position



(u) 01001

Fig. 4. Synthetic Mouse-in-Cage Scenario Video Clips



Fig. 5. Behaviors in the synthetic video sequence. Rest= 0, Explore = 1, Eat = 2, Groom = 3

and orientation are derived analytically from the simulation. Each feature element is a 60-dimensional vector.

**Results.** For the feature data extracted from the synthetic video, the results are shown in Figures 6 and 7. The four BBUs are not clearly separated.

For the analytical artificial mouse data, if we use all 60-dimensional feature data, the vector fusion result does not distinguish the behaviors either. Figure 8 uses absolute position of the mouse body and limbs, the orientation is in radians ( $0 \sim 2\pi$ ). Figure 9 shows the result using relative position of the limbs (relative to the mouse body), and the orientation is in radians. The result of using relative position using radians start showing some kind of pattern for different BBUs, comparing to using absolute positions. This is reasonable, since the relative motion of the limbs best distinguishes the four BBUs. Also, we found that proper normalization is needed for each dimension of the feature



Fig. 6. Vector Fusion Result for Mouse BBU



Fig. 7. Vector Fusion Result for Mouse BBU Zoom In. 'o'-Rest, Green 'x'-Explore, '+'-eat, Red 'x'-Groom

data. Otherwise, the result would not be meaningful.

Based upon the previous results and the motion pattern (Explore, Eat, and Groom also exhibits some periodic limb motion) for each BBU for the analytical data, we changed to use only artificial mouse limb orientation (rotation angles relative to the mouse body– local motion). Each dimension of the feature data is normalized to the range of  $0 \sim 1$ . This time we get much better result, as shown in Figure 10. The result using artificial mouse limb orientation (four limbs)  $(\theta 1, \theta 2, \theta 3)$  and the body speed( $d\mathbf{x}$ ) is shown in Figure 11, comparable to Figure 10. Now we can easily distinguish the BBUs, by fitting lines or ellipses to the data.

Figure 12 shows the vector fusion result for each BBU, where the SBPx and SBPy coordinates of each BBU sequence are plotted (in the vertical axis) against the time step (in the horizontal axis). The SBPx coordinate of each BBU sequence is plotted in the top figure, and the SBPycoordinate of each BBU sequence is plotted in the bottom figure. We can see that the SBP coordinate sequence for each BBU exhibit either stationary or periodic pattern. By making movies of how the SBPx, SBPy coordinates (or the SBP point in the SBP space) change over time for each BBU, we can observe more clearly the temporal patterns of each BBU (see http://www.cs.utah.edu/ ~xwxue /vectorFusion/ for the movies): The rest BBU is basically a stationary point, the



Fig. 8. Vector Fusion Result for Mouse BBU – Absolute Limb Position, Total 60 Dimensions. 'o'-Rest, '\*'-Explore, '+'-eat, '•'-Groom



Fig. 9. Vector Fusion Result for Mouse BBU – Relative Limb Position, Total 57 Dimensions. 'o'-Rest, '\*'-Explore, '+'-eat, '•'-Groom

explore, eat, and groom BBUs show obvious periodic motion along different lines. Hence we can easily distinguish each BBU in the sequence.

### V. CONCLUSIONS

We have explored the vector fusion method for its application in object basic behavior unit segmentation in a temporal sequence, and presented results on a physical system and a synthetic mouse-in-a-cage scenario. The vector fusion method reduces multiple dimensional data into the 2D SBP space, and the spatial and temporal analysis in SBP space provides a good distinction and interpretation for the bouncing ball example and the analytical data from the synthetic video simulation upon certain selected features.

Our experiments show that several factors influences the effectiveness of the vector fusion method in BBU segmentation. First, proper features with enough BBU distinguishing power needs to be selected, just as in other BBU segmentation methods. Second, the weights of each feature element in the multiple-dimensional feature space plays an important role, hence, each feature element needs to be properly normalized to account for the different value range (hence different weight) for each feature element, and the distinguishing power of the features.

The result of the temporal analysis in SBP space suggests



Fig. 10. Vector Fusion Result for Mouse BBU – Normalized Limb Angles Only Total 8 Dimensions. '\*'-Rest, '+'-Explore, '•'-eat, 'x'-Groom



Fig. 11. Vector Fusion Result for Mouse BBU – Normalized Limb Angle and Body Speed, Total 9 Dimensions. '\*'-Rest, '+'-Explore, '•'-eat, 'x'-Groom

it can be very powerful for BBUs consists of periodic motion [17], and may be potentially a good approach for motion capture data analysis (where joint angles can be easily calculated). Its great simplicity (reducing multi-dimensional feature space to the 2D SBP space) is a great advantage over the more complex methods.

The potential future work includes the optimal selection of features for real video, and further exploration of this method on other BBUs of periodic motion from motion capture data and real video data.

## VI. ACKNOWLEDGMENTS

The authors thank Bob Johnson for helpful discussions on the vector fusion method.

#### REFERENCES

- J. Aggarval and Q. Cai, "Human motion analysis: A review," Computer Vision and Image Understanding, vol. 73, no. 3, 1999.
- [2] L. Wang, W. Hu, and T. Tan, "Recent developments in human motion analysis," *Chinese Journal of Computers*, vol. 25, no. 3, pp. 225–237, 2002.
- [3] W. Hu, T. Tan, L. Wang, and S. Maybank, "A survey on visual surveillance of object motion and behaviors," *IEEE Trans. on Systems, Man, and Cybernetics*, vol. 34, no. 3, pp. 334–351, 2004.
- [4] T. Moeslund and E. Granum, "A survey of computer vision-based human motion capture," *Computer Vision and Image Understanding*, vol. 81, no. 3, pp. 231–268, 2001.



Fig. 12. Vector Fusion Result for Each Artificial Mouse BBU Using Normalized Limb Orientation and Body Speed: a)Rest, b)Explore, c)Eat d) Groom. For each BBU, the top figure plots the temporal  $SBP_x$  coordinate, and the bottom figure plots the temporal  $SBP_y$  coordinate.

- [5] P. van Lochem, M. Buma, J. Rousseau, and L. Noldus, "Automatic recognition of behavioral patterns of rats using video imaging and statistical classification," in *Measuring Behavior*, Groningen, The Netherlands, 1998.
- [6] L. Noldus, A. J. Spink, and R. A. Tegelenbosch, "Ethovision: A versatile video tracking system for automation of behavioral experiments," *Behavior Research Methods, Instruments, & Computers*, vol. 33, no. 3, pp. 398–414, 2001.
- [7] O. C. Jenkins and M. J. Mataric, "Deriving action and behavior primitives from human motion data," in *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems(IROS)*, Lausanne, Switzerland, 2002, pp. 2551–2556.
- [8] A. Fod, M. J. Mataric, and O. C. Jenkins, "Automated derivation of primitives for movement classification," *Autonomous Robots*, vol. 12, no. 1, pp. 39–54, 2002.
- [9] J. Barbic, A. Safonova, J.-Y. Pan, C. Faloutsos, J. K. Hodgins, and N. S. Polland, "Segmenting motion capture data into distinct behaviors," in *Proceedings of Graphics Interface 2004 (GI'04)*, London, Ontario, Canada, May 2004.
- [10] X. Xue and T. C. Henderson, "Video-based animal behavior analysis," University of Utah, TechReport UUCS-06-006, June 2006.
- [11] T. C. Henderson and X. Xue, "Construct complex behaviors: A simulation study," in ISCA 18th Intl. Conf. on Computer Applications in Industry and Engineering (CAINE), Hawaii, Nov. 2005.
- [12] L. Zelnik-Manor and M. Irani, "Event-based analysis of video," in *Proceedings of IEEE CVPR*, Hawaii, 2001.
- [13] F. Porikli and T. Haga, "Event detection by eigenvector decomposition using object and frame features," in *Workshop on Event Mining, IEEE CVPR*, Washington DC, 2004.
- [14] R. Johnson, "Visualization of multi-dimensional data with vector fusion," in *IEEE Proceedings of Visualization*, 2000, pp. 297–302.
- [15] —, "Relational data analysis: Characterizing and measuring data to discover relationships in that data," http://www.n-dv.com, Research Papers, 2006.
- [16] A. Inselberg and B. Dimsdale, Parallel coordinates, a tool for visualizing multivariate relations. Human-Machine Interactive Systems. New York: Plenum Press Publishing, 1991.
- [17] R. Cutler and L. Davis, "Robust real-time periodic motion detection, analysis, and applications," *IEEE Transactions on PAMI*, vol. 22, no. 8, pp. 781–796, 2000.