Pair-Activity Classification by Bi-Trajectories Analysis

Yue Zhou[1], Shuicheng Yan[2], and Thomas S. Huang[1]

[1] Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, USA

[2] Department of Electrical and Computer Engineering, National University of Singapore, Singapore

Abstract

In this paper, we address the pair-activity classification problem, which explores the relationship between two active objects based on their motion information. Our contributions are three-fold. First, we design a set of features, e.g., causality ratio and feedback ratio based on the Granger Causality Test (GCT), for describing the pair-activities encoded as trajectory pairs. These features along with conventional velocity and position features are essentially of multi-modalities, and may be greatly different in scale and importance. To make full use of them, we then present a novel feature normalization procedure to learn the coefficients for weighting these features by maximizing the discriminating power measured by weighted correlation. Finally, we collected a pair-activity database of five categories, each of which consists of about 170 *instances. The extensive experiments on this database validate the effectiveness of the designed features for pair-activity representation, and also demonstrate that the proposed feature normalization procedure greatly boosts the pair-activity classification accuracy.*

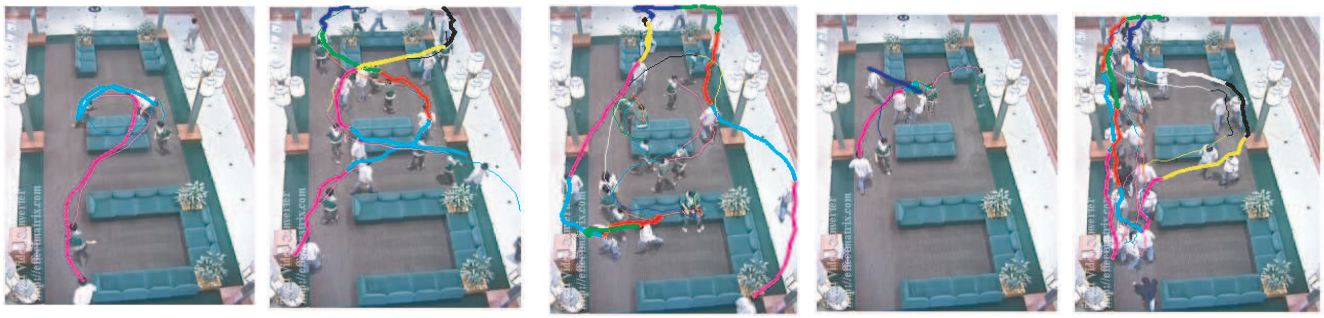
1. Introduction

Computer vision research has experienced increasing passions on activity analysis in the past few years. Previous research devoted to the activity analysis problem can be roughly divided into two categories: parametric approaches and non-parametric ones. For parametric approaches [8], the activity models are explicitly built based on the visual features, such as position, velocity, and appearance, extracted by certain object detection and tracking algorithms [7][16][17]. These models are either rule-based or learnt using supervised learning techniques, *e.g.*, probabilistic graphical models [10][3][11]. Non-parametric approaches, on the other hand, do not explicitly define the activity/event models, instead they learn the activity patterns from the statistical properties of the observed data. In [14], joint co-occurrence statistics of the object trajectories over a codebook are accumulated, and the hierarchical classification method is applied to identify the activities. Zhong *et al.* [19] proposed an unsupervised technique for detecting unusual activity in a large video set using many simple features. Zhou *et al.* [20] used motion trajectory similarity to detect anomalous events. Boiman and Irani [2] proposed to determine the video regularity by its probability to be composed from reasonably large chunks of spatial-temporal data.

Most previous research on activity analysis stems from the low-level visual information. Those harnessing the middle-level visual information emphasized on the scenarios with only a single active object and focused on the single-role activities. Recently, Wu *et al.* [18] proposed an algorithm to classify the activity of one object with the interactions with the circumambient appliances or articles, *e.g.*, microwave, coffee maker, and cups. Besides the single-role activities, many activities existing in real life involve two or even more active objects, *e.g.*, chasing and working together, and these activities are much more complicated than single-role activities. The analysis of these activities is critical for practical video surveillance systems. In this work, we focus on the pair-activities, which describe the relationship between two active objects, but all the techniques discussed in this work can be easily extended for group-activity analysis with multiple active objects.

The study of the pair-activities is new for the computer vision research community although a large portion of the activities existing in real life belong to this type. In this paper, we contribute to this problem from the following aspects. First, we design a set of features to characterize these pair-activities. More specifically, we first encode the each pair-activity example as two trajectories, referred to as bi-trajectory in this work, extracted from the mean-shift tracking algorithm [5][4]. Then a set of features, *e.g.*, causality ratio and feedback ratio, are derived by the Granger Causality Test (GCT) [6] based on the bi-trajectories, and finally the length-variable pair-activity instance is represented by these features along with the velocity and distance features conventionally used for single-trajectory analysis. It is observed that the extracted features are of multi-modalities, and often greatly different in scales and importance. How

978-1-4244-2243-2/08/$25.00 ©2008 IEEE



(a) Chasing (b) Following (c) Independent (d) Meeting (e) Together

Figure 1. Examples of pair-activities. From left to right: (a) chasing, (b) following, (c) independent, (d) meeting, and (e) together. Note that for each example, the trajectories of two objects are obtained from the object tracking algorithm, and the multiple instances of one object in the image are obtained by background substraction and object superposition. The two trajectory segments (one thick and another thin) with the same color correspond to the motion trajectories of two objects within the same time period.

to weight and normalize these features is crucial for the subsequent classification step, since different weights will result in different margins and the separation hyperplane will also be changed. Thus we then present a novel feature normalization/weighting procedure to learn the coefficients for weighting different features such that the normalized data will be good at discriminating power measured by the correlation similarity. Finally, we collected a par-activity database with five categories of pair-activities, namely chasing, following, independent, meeting and together, each of which consists of averagely 170 instances. The extensive experiments on this database validate the effectiveness of the designed features as well as the proposed feature normalization procedure.

The rest of the paper is organized as follows. Section 2 introduces the details of features extraction from the Granger Causality Test for pair-activity representation. The procedure to learn the feature weighting coefficients for enhancing the feature discriminating power is described in Section 3. Section 4 provides the details on database collection and the comparison experiments on pair-activity classification. The concluding remarks are given in Section 5.

2. Representation by Bi-trajectory Analysis

The success of a solution to a particular classification task generally relies on two aspects: 1) how to represent the samples, and 2) how to measure the similarity between the sample pair. In this work, we discuss these two problems for the concerned pair-activity classification problem. The first problem is discussed in this section, and we discuss the second problem in the next section.

2.1. Pair-Activity Encoded as Bi-Trajectory

For a video with two active objects, there exist multiple levels of information for characterizing the involved activity. The low-level information, *e.g.*, color and local motion information, is useful for describing single active object, yet insufficient to reveal the contextual information between two active objects. The high-level information, *e.g.*, gender and identity information, also generally characterizes single object. Hence, in this work, we focus on the middle-level information, namely bi-trajectories, for characterizing the relationship between two active objects.

Motion trajectory has proved effective in sing-role activity analysis [15]. Generally, the trajectory of a moving object is obtained by certain object tracking algorithm. Figure 1 depicts five examples corresponding to five categories of pair-activities, and the motion bi-trajectories are also imposed. From these bi-trajectories, we can intuitively observe that the shapes of the bi-trajectories of different categories are different, and the different is even greater if the temporal information is also considered. Pair-activity belongs to high-level concept, and reflects both spatial and temporal information of two active objects. Conventionally, each point on a motion trajectory is characterized by its position, velocity and curvature features. These features are useful for describing the activity of one object, but there still exists a large gap between these middle-level information and the high-level concept of pair-activity, which is much more complicated than single-trajectory based activity. We introduce in the next subsection how to utilize the Granger Causality Test for extracting semi-high-level features to bridge this gap.

2.2. Features via Granger Causality Test

The Granger Causality Test (GCT) was originally proposed for computing the relationship between certain economical factors involving causality and feedback. GCT was first proposed by Granger in [6] and then further popularized by Sims in [13]. In the case with two variables/objects, the GCT breaks the feedback mechanism into two causal relations, and each is closely connected with one of the causations.

Before we formally introduce how to extract semi-high-level features for describing the pair-activities, we first define some terminologies as follows.

1. Let $A_t = (a_0, a_1, a_2, , a_t)$ and $B_t = (b_0, b_1, b_2, , b_t)$ be two motion trajectories of two objects, where a_i and b_i are tuples $[x, y]$ of the object coordinates in 2D image plane at time i. We assume that the interaction between two trajectories is a stationary process, *i.e.* the prediction functions $P(a_t|B_{t-l_a})$ and $P(b_t|A_{t-l_b})$ do not change within a short period of time.

2. Let $P(a_t|B_{t-l}(k))$ be the optimal predictor of a_t using the information of B_{t-l} where B_{t-l} is the l-delayed information of trajectory B_t and l is a nonnegative number. The prediction function is estimated by using k samples in B_t, namely, $(b_{t-l}, b_{t-l-1}, \cdots, b_{t-l-k+1})$.

3. The prediction error is assumed to of Gaussian noise with standard deviation as $\delta(a_t|B_{t-l}(k))$.

4. Let U_{t-l} be all the information accumulated from both A_t and B_t before and at the time $t-l$, and $U_{t-l}-B_{t-l}$ be all the information except for B_{t-l}.

To model $P(a_t|B_{t-l}(k))$, we use the linear predictor owing to its simplicity and efficiency in parameter estimation. The optimal parameters of the prediction function can be computed by Least Square Error (LSE) approach. The GCT does not specify any particular prediction function for computing the standard deviation, and hence more complex prediction functions, *e.g.*, polynomial function and logistic function, can also be used for modeling $P(a_t|B_{t-l}(k))$.

From GCT, we can obtain a set of features for measuring certain properties within the bi-trajectory as follows:

1. *Causality* - if the prediction error $\delta(a_t|U_{t-l}(k)) < \delta(a_t|U_{t-l} - B_{t-l}(k))$, we say B is Granger causal for A.

2. *Feedback* - if B_t is Granger causal for A_t and A_t is Granger causal for B_t, we say A_t and B_t have feedback. However, they may have different causality time lags.

3. *Causality ratio* - $\delta(a_t|U_{t-l}-B_{t-l}(k))/\delta(a_t|U_{t-l}(k))$, which measures the relative strength of the causality, and is more stable for measuring causality compared the absolute causality value.

4. *Feedback ratio* - $\delta(b_t|U_{t-l} - A_{t-l}(k))/\delta(b_t|U_{t-l}(k))$, which measures the relative strength of the feedback.

Intuitively, causality is an effective criteria in distinguishing different categories of pair-activities. For example, for chasing, the causality exists for both active objects, while for following, the following object will only have very limited causality for the object who is being followed, and hence the difference between feedback ratio and causality ratio should be larger than that for the chasing case. In the next subsection, we introduce the five categories of pair-activities we concern in this work, and intuitively justify how these pair-activities can be characterized by the above features from GCT.

2.3. Pair-Activities vs. GCT Features

In this work, we are interested in five categories of pair-activities as follows:

1. *Chasing* - One object tries to minimize the relative distance while the other object tries to maximize the relative distance. Hence they affect to each other, and intuitively feedback ratio is high while the causality ratio may also be high.

2. *Following* - One object tries to minimize the relative distance while another object may not realize and ignores the follower. In this case, the feedback ratio is high, yet the casuality ratio is low.

3. *Independent* - Neither object is the causation of the other. In this case, the causality ratio and feedback ratio are both low.

4. *Meeting* - Both objects first try to minimize the relative distance, stay together for a while, and then depart. The causality ratio and feedback ratio are both high in the first stage, but are low in the third stage. In this case, the relative position and velocity information of the objects may also play important role in characterizing the activity.

5. *Together* - The distance between two objects are small or nearly constant at all the time. In this case, the causality ratio and the feedback ratio are both low, and the relative distance is useful for describing this activity.

As discussed above, the designed features from GCT are effective in distinguishing different categories of pair-activities, but not enough for differentiating all the pair-activities concerned in this work, *e.g.*, the pair-activities of meeting and together heavily depend on the relative positions of two objects. Hence, the pair-activities are represented by the features from GCT along with the conventional features, *e.g.*, the velocity and relative distance features.

Here, let $\{x_i | x_i \in \mathbb{R}^m\}_{i=1}^N$ be the pair-activity sample set, where x_i is the feature vector which combines the features from GCT and conventional velocity and position features, and their corresponding class labels are $\{c_i | c_i \in \{1, ..., N_c\}\}_{i=1}^N$, where N_c is the number of classes.

These m-dimensional vectors involve the features of different sources, and their scales may be dramatically different, and also their contributions to the classification problem may also be different. The scales of these features directly determine the between-class margins, and hence are important for the consequent pair-activity classification. As above-mentioned, these features are proposed with different purposes and of multi-modalities, and hence it is desirable to have a procedure which can automatically determine the weights for different features such that the final discriminating power is the best. Our procedure in the next section for feature normalization based on weighted correlation is proposed for such a purpose.

3. Classification by Feature Normalization

In this section, we study the underlying general machine learning problem on how to learn the coefficients for weighting the features from GCT and conventional trajectory representation so as to boost the discriminating power of these features based on the weighted correlation.

3.1. Formulation for Feature Weighting

Correlation similarity (cosine distance) was reported to be generally more robust and effective for classification tasks compared with the conventional metrics such as L^2 and L^1 distances [9]. But most traditional algorithms for feature extraction or feature selection are based on Euclidean distances. In this subsection, we present a novel supervised feature weighting algorithm which characterizes discriminant power based on correlation similarity directly. The target is to search for a weighting vector $p \in \mathbb{R}^m$ such that the weighted representation $p \odot x = (p_1 x_1, p_2 x_2, \cdots, p_m x_m)^T$ possesses boosted discriminating power.

To measure the discriminating power of the weighted feature space, intra-class compactness and inter-class separability are two most used criteria. Here, we define them based on the correlation similarity.

Intra-class Similarity: For a representation with good discriminating power, it is desirable that the samples from the same class are be close to each other as much as possible. On the other hand, it is often the case that there may exist diverse variations within individual class, and hence it is unnecessary that all sample pairs within certain class are close to each other. A trade-off way is to ensure that the neighboring samples of the same class are close to each other. The neighborhood relationship can be defined from the $\hat{k}$-nearest neighbors or ϵ-ball criteria in the original feature space. In this work, we use the $\hat{k}$-nearest neighbors criteria for measuring the intra-class compactness as

$$\begin{aligned} S_c &= \sum_{i=1}^{N} \sum_{j \in N_{\hat{k}}^{+}(x_i)} \frac{< p \odot x_i, p \odot x_j >}{\|p \odot x_i\| \, \|p \odot x_j\|} \\ &= \sum_{i=1}^{N} \sum_{j \in N_{\hat{k}}^{+}(x_i)} \frac{\sum_{k=1}^{m} p_k^2 x_i^k x_j^k}{\sqrt{\sum_{k=1}^{m} p_k^2 x_i^k x_i^k} \sqrt{\sum_{k=1}^{m} p_k^2 x_j^k x_j^k}}, \end{aligned} \quad (1)$$

where $N_{\hat{k}}^{+}(x_i)$ is the index set for all the $\hat{k}$-nearest neighbors of the sample x_i and from the same class.

Inter-class Similarity: For good discriminating power, the sample pair from different classes should be far away to each other. Generally, the number of this kind of pairs is huge even for a moderate size database. To alleviate the computational cost, this kind of inhomogeneous pairs are considered mainly for marginal pairs. In this paper, we also use the $\hat{k}$-nearest neighbors criteria for measuring the inter-class separability as

$$\begin{aligned} S_p &= \sum_{i=1}^{N} \sum_{j \in N_{\hat{k}}^{-}(x_i)} \frac{< p \bigodot x_i, p \bigodot x_j >}{\|p \bigodot x_i\| \, \|p \bigodot x_j\|} \\ &= \sum_{i=1}^{N} \sum_{j \in N_{\hat{k}}^{-}(x_i)} \frac{\sum_{k=1}^{m} p_k^2 x_i^k x_j^k}{\sqrt{\sum_{k=1}^{m} p_k^2 x_i^k x_i^k} \sqrt{\sum_{k=1}^{m} p_k^2 x_j^k x_j^k}}, \end{aligned} \quad (2)$$

where $N_{\hat{k}}^{-}(x_i)$ is the index set for all the $\hat{k}$-nearest neighbors of sample x_i and from difference classes. In this work, the $\hat{k}$ is set as 10 for all the experiments.

The correlation similarity takes the value within $[-1, 1]$, and hence the relative larger (or smaller) term will not dominate the value of S_c or S_p, which is the main superiority of correlation similarity over the L^2 and L^1 distances for measuring intra-class compactness and inter-class separability. Aiming at good classification capability, we maximize the intra-class similarity and at the same time minimize the inter-class similarity, and consequently we have the following objective function

$$\max_{p} \{F(p) = S_c - S_p\}. \quad (3)$$

From (1) and (2), we can see that the S_c and S_p are invariant to the scaling of the weighting vector p. Hence, we add the constraint $\|p\| = 1$ to the objective function and this constraint can be easily satisfied by normalizing p after each iteration in the learning process.

3.2. Parameter Optimization

The objective function of (3) is nonlinear, and hence we use the iterative gradient descent approach for optimization.

Table 1. Details of the collected pair-activity database.

	Chasing	Following	Together	Meeting	Independent	In Total
Session Number	6	5	3	7	3	24
Total Clip/Sample Number	153	198	182	131	203	867

Table 2. Feature details for pair-activity representation.

Type	Causality/Feedback	Relative Velocity	Relative Distance	Absolute Velocity	Absolute Position
Number	2	4	4	8	8

The derivative of the objective function with respect to p is

$$\frac{\partial F(p)}{\partial p} = \sum_{i=1}^{N}\{ \sum_{j \in N_k^+(x_i)} \frac{\partial f_{ij}(p)}{\partial p} - \sum_{j \in N_k^-(x_i)} \frac{\partial f_{ij}(p)}{\partial p} \}, \quad (4)$$

where

$$f_{ij}(p) = \frac{\sum_{k=1}^{m} p_k^2 x_i^k x_j^k}{\sqrt{\sum_{k=1}^{m} p_k^2 x_i^k x_i^k}\sqrt{\sum_{k=1}^{m} p_k^2 x_j^k x_j^k}}. \quad (5)$$

The derivative of $f_{ij}(p)$ with respect to p_k is calculated as

$$\frac{\partial f_{ij}(p)}{\partial p_k} = \frac{2x_{ij}^k b_{ii}^{\frac{1}{2}} b_{jj}^{\frac{1}{2}} + b_{ij}(b_{jj}^{\frac{1}{2}} x_{ii}^k / b_{ii}^{\frac{1}{2}} + b_{ii}^{\frac{1}{2}} x_{jj}^k / b_{jj}^{\frac{1}{2}})}{b_{ii} b_{jj}}, \quad (6)$$

where

$$x_{ij}^k = x_i^k x_j^k, \quad (7)$$

$$b_{ij} = \sum_{k=1}^{m} p_k^2 x_i^k x_j^k. \quad (8)$$

Based on the derivative, we can optimize the objective function in (3) with gradient descend approaches, *e.g.*, levenberg-marquart which is implemented in Matlab as *lsnonlin* function.

3.3. Classification with Weighted Features

Based on the derived vector p in the above subsection, each datum x is reweighted as $p \odot x$, and then we can directly classify the pair-activities with the Nearest Neighbor approach based on this rescaled feature space. We can also use other more complex multi-class learning algorithm to further improve the algorithmic performance. In this paper, we choose the Linear Discriminant Analysis [1] to further reduce the feature dimension with the projection matrix $P \in \mathbb{R}^{m \times d}$ (usually $d \ll m$). Then, when a new datum comes, its class label is predicted as that of the sample with index

$$\arg\max_i \frac{< P^T \times (p \odot x), P^T \times (p \odot x_i) >}{\|P^T \times (p \odot x)\| \, \|P^T \times (p \odot x_i)\|}. \quad (9)$$

Also we evaluate the performance of Support Vector Machine (SVM) [12] in pair-activity classification based on weighted features, owing to its popularity in general classification problems. In this work, we use the one versus all approach to extend the basic two-class SVM for handling the multi-class problem.

4. Experiments

In this section, we first introduce the details of the pair-activity database we collected, and then demonstrate the effectiveness of the designed features from Granger Causality Test for pair-activity representation, and finally we evaluate the effectiveness of the proposed feature weighting/normalization procedure.

4.1. Pair-Activity Database Construction

The database we used for the experiments was collected at the indoor environment with a still camera. The camera (Sony 908E, 30f/s) was installed in the second floor of a building, and captures the hall area in the first floor from the top-view. The whole database consists of five categories of most common pair-activities: chasing, following, together, meeting, and independent. For each category of pair-activity, multiple sessions with different character pairs are captured. Within each session, two characters continuously staged the same pair-activity and change the style frequently, and then the long video is divided into a series of clips with overlapping between successive clips. Finally, each category consists of 131 to 203 video clips of 8 seconds or longer, and altogether we have 867 video clips for all the five categories. The detailed information of session number and clip number for each category of the pair-activities is listed in Table 1, and some example clips are displayed in Figure 1.

To estimate the trajectories of the two moving objects in the videos, we use the mean-shift tracking algorithm [5][4], and the tracker is automatically initialized from the Gaussian Mixture Models (GMM) based background subtraction algorithm.

4.2. Feature Effectiveness

The features we used for the experiments include: causality ratio, feedback ratio, relative velocity and distance, and absolute velocities and positions for two objects. For the conventional features of single trajectory representation such as velocity, position and etc., we compute both

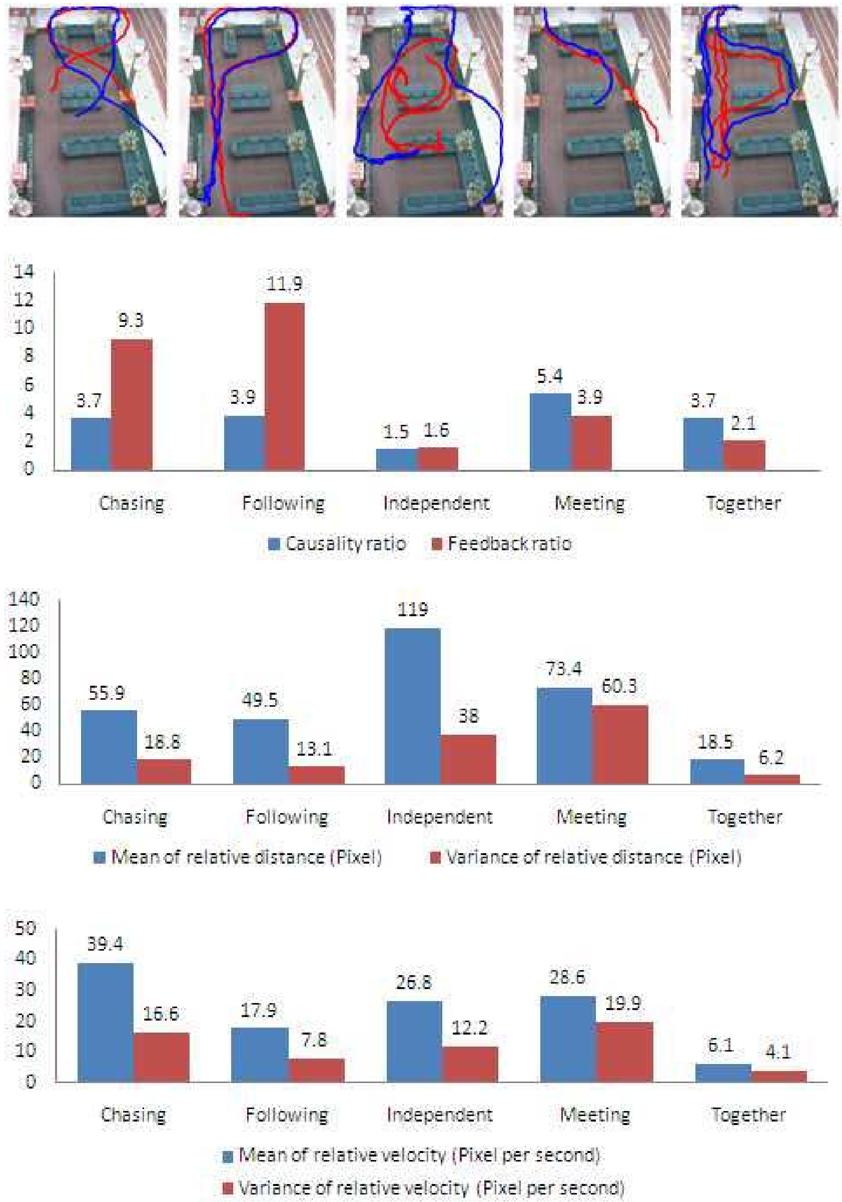


Figure 2. Comparison of feature values of causality ratio, feedback ratio, relative distance and relative velocity of examples from different categories.

mean and variance of their magnitudes and orientations, so finally we have 26 features in total as shown in Table 2. The causality ratio and feedback ratio are estimated using 10th order regression model, namely $k = 10$ for computing the casuality ratio value $\delta(a_t|U_{t-l} - B_{t-l}(k))/\delta(a_t|U_{t-l}(k))$, where l is set as 15 due to the high frame rate of the video clips.

In Figure 2, we list the values of the causality ratio, feedback ratio, and relative velocity features for five example pair-activity instances from five different categories. From these values, we can observe that: 1) the causality and feedback ratios are very useful in discriminating different categories of pair-activities; and 2) the relative velocity feature is not so important as the causality ratio feature, but also contributes greatly, especially for the pair-activities independent and chasing.

4.3. Pair-Activity Classification

For pair-activity classification, the 26 features weighted by the derived vector p are fed into the Linear Discrimi-

Table 3. Comparison confusion tables of pair-activity classification accuracy (%): Original features plus LDA (OF-LDA) vs. weighted features plus LDA (WF-LDA).

OF-LDA	Chasing	Following	Together	Meeting	Independent
Chasing	89.1%	7.3%	1.9%	0%	1.7%
Following	8.4%	91.2%	0.1%	0.2%	0%
Together	0.3%	7.5%	74.6%	8.1%	9.5%
Meeting	6.1%	5.6%	9.8%	64.6%	13.9
Independent	8.5%	7.0%	6.9%	11.2%	66.4%
WF-LDA	Chasing	Following	Together	Meeting	Independent
Chasing	89.6%	7.3%	1.7%	0.2%	1.2%
Following	6.3%	93.4%	0.1%	0.2%	0%
Together	0.3%	4.1%	85.3%	5.5%	4.8%
Meeting	5.2%	5.7%	10.4%	62.4%	16.3
Independent	5.7%	6.1%	5.3%	11.9%	71.0%

Table 4. Comparison confusion tables of pair-activity classification accuracy (%): Original features plus SVM (OF-SVM) vs. weighted features plus SVM (WF-SVM).

OF-SVM	Chasing	Following	Together	Meeting	Independent
Chasing	89.5%	8.1%	1.1%	0%	1.3%
Following	3.3%	95.2%	0%	1.5%	0%
Together	3.2%	8.3%	80.1%	0%	7.9%
Meeting	6.9%	5.8%	6.0%	69.7%	11.5
Independent	8.4%	7.3%	7.8%	15.2%	61.3%
WF-SVM	Chasing	Following	Together	Meeting	Independent
Chasing	92.1%	7.0%	0.9%	0%	0%
Following	2.7%	97.0%	0%	0.3%	0%
Together	1.2%	3.1%	89.1%	3.5%	2.7%
Meeting	6.5%	2.2%	8.1%	70.1%	13.1
Independent	3.2%	5.1%	2.5%	10.5%	78.7%

native Analysis (LDA) algorithm, and then the final classification is conducted based on the dimensionality reduced feature space with the Nearest Neighbor approach. Also these features can be used as input for SVM based classification. The experimental protocol is leave-one-session-out, namely each time, the samples within one session are used as testing data and all the others as training data.

We compare the experimental results from LDA based on original features with those from LDA and those from LDA based on weighted features. The detailed results are listed in Table 3, from which we can observe that the classification accuracies are improved for most categories by the feature normalization procedure, especially for the together category. We also compare the experimental results from Gaussian-kernel based SVM with original features and those from SVM with weighted features, and the results listed in Table 4 validate the effectiveness of the feature normalization process as well. Also the results show that SVM is more powerful for classification task as compared with LDA.

In addition, we evaluate the effectiveness of individual feature for pair-activity classification, and here we focus on two most representative ones, namely causality ratio and relative velocity, one of which from Granger Causality Test and the other from the conventional single-trajectory representation. The experiments are conducted based on LDA along with the feature normalization procedure. Table 5 lists the detailed results where the causality ratio or relative velocity feature is removed respectively for pair-activity representation. The observation can be made that: 1) when causality ratio feature is removed, the classification performance degrades dramatically, especially for the categories of chasing, following and together, and hence the causality information is significant in identifying the pair-activities; and 2) when the relative velocity feature is removed, the performance degrades, but not so great as for causality ratio feature, which means that the relatively velocity is important but less important than the causality ratio feature for pair-activity classification.

5. Conclusions and Future Works

This work was devoted to the pair-activity classification problem from three aspects: database construction, feature design, and feature normalization for classification. To the best of our knowledge, the pair-activity database introduced

Table 5. Comparison confusion tables of pair-activity classification accuracy (%) based on weighted features plus LDA: without causality ratio feature vs. without relative velocity features (mean and variance).

Without Causality Ratio	Chasing	Following	Together	Meeting	Independent
Chasing	47.0%	25.3%	21.8%	2.2%	2.9%
Following	17.9%	61.4%	12.4%	4.2%	4.1%
Together	8.8%	15.1%	65.9%	4.6%	5.6%
Meeting	8.2%	10.1%	10.4%	53.1%	18.2
Independent	6.2%	17.5%	16.1%	18.5%	41.7%
Without Relative Velocity	Chasing	Following	Together	Meeting	Independent
Chasing	84.6%	6.7%	5.4%	1.2%	2.1%
Following	9.1%	81.5%	5.8%	2.9%	0.7%
Together	3.1%	10.9%	73.4%	3.8%	8.8%
Meeting	7.0%	3.2%	7.3%	63.1%	19.4
Independent	13.1%	9.2%	4.5%	11.1%	62.1%

in this work is the first one available for studying the pair-activities, and also it is the first work to study the pair-activity classification problem. We are planning to further exploit this topic in three aspects: 1) to combine the middle-level visual information (trajectories) with the low-level (patch appearance) and high-level information (object categories) for better understanding the pair-activities; 2) to extend our proposed techniques for analyzing group-activities involving multiple active objects; and 3) to extend the current work to a more general framework for automatic pair-activity detection, segmentation, and classification within a long video.

Acknowledgment

This work was supported in part by US Government VACE Program, and in part by AcRF Tier-1 Grant of R-263-000-464-112/133, Singapore.

References

[1] P. Belhumeur, J. Hespanha, and D. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, pp. 711–720, 1997.

[2] O. Boiman and M. Irani. Detecting irregularities in images and in video. *IEEE International Conference on Computer Vision*, pp. 462–469, 2005.

[3] M. Brand, N. Oliver, and A. Pentland. Coupled hidden markov models for complex action recognition. in *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 994–999, 1997.

[4] R. Collins. Mean-shift Blob Tracking through Scale Space, *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 234–240, 2003.

[5] D. Comaniciu, V. Ramesh, and P. Meer. Kernel-Based Object Tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 5, pp. 564–574 2003.

[6] C. Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, vol. 37, no. 3, pp. 424–38, 1969.

[7] I. Haritaoglu, D. Harwood, and L. David. W4: Real-time surveillance of people and their activities. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 809–830, 2000.

[8] Y. Ivanov and A. Bobick. Recognition of visual activities and interactions by stochastic parsing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, pp. 852–872, 2000.

[9] D. Lin, S. Yan, and X. Tang. Comparative study: face recognition on unspecific persons using linear subspace methods. *IEEE International Conference on Image Processing*, vol. 3, pp. 764-767, 2005

[10] G. Medioni, I. Cohen, F. Bremond, S. Hongeng, and R. Nevatia. Event detection and analysis from video streams. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 8, pp. 873–889, 2001.

[11] D. Moore, I. Essa, and M. Haye. Exploiting human actions and object context for recognition tasks. in *IEEE International Conference on Computer Vision*, pp. 80–86, 1999.

[12] E. Osuna, R. Freund, and F. Girosi. Training support vector machine: An application to face detection. *IEEE Conference on Computer Vision and Pattern Recognition* pp. 130–136, 1997.

[13] C. Sims. Money, income, and causality. *American Economic Review*, pp. 540–52, 1972.

[14] C. Stauffer and W. Grimson, Learning patterns of activity using real-time tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 747–757, 2000.

[15] P. Turaga, A. Veeraraghavan, and R. Chellappa. From Videos to Verbs: Mining Videos for Activities using a Cascade of Dynamical Systems. *IEEE Conference on Computer Vision and Pattern Recognition*, 2007.

[16] P. Viola and M. Jones. Robust real-time face detection. *International Journal on Computer Vision*, pp. 137–154, 2004.

[17] C. Wren, A. Azarbayejani, T. Darrell, and A. Pentland. Pfinder: Real-time tracking of the human body. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 780–785, 1997.

[18] J. Wu, A. Osuntogun, T. Choudhury, M. Philipose, and J. Rehg. A Scalable Approach to Activity Recognition Based on Object Use. *International Conference on Computer Vision*, 2007.

[19] H. Zhong, J. Shi, and M. Visontai. Detecting unusual activity in video. *IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 819–826, 2004.

[20] Y. Zhou, S. Yan, and T. Huang. Detecting Anomaly in Videos from Trajectory Similarity Analysis *International Conference on Multimedia and Expo*, 2007.