A New Combinatorial Approach to Supervised Learning: Application to Gait Recognition

Rong Zhang[1], Akshay Vashist[1], Ilya Muchnik[1,2],
Casimir Kulikowski[1], and Dimitris Metaxas[1]

[1] Dept. of Computer Science
{roni, vashisht, kulikows, dnm}@cs.rutgers.edu
[2] DIMACS Rutgers, The State University of New Jersey,
Piscataway, NJ 08854, USA
muchnik@dimacs.rutgers.edu

Abstract. In many supervised learning problems, objects are represented as a sequence of observations. To classify such data, existing methods build classifiers either based on their dynamics, or the statistics of the observations. However, similar observations shared by most objects are uninformative for identification. In this paper, we present a new approach that identifies similar observations across objects and use only informative data for classification. To do this, we construct a weighted multipartite graph from the training data, with weights representing the similarities between observations from different objects. Identification of uninformative observations is modeled as clustering on this multipartite graph using a combinatorial optimization formulation. Two-level hierarchical classifiers are, then, built using the clustering results. The first layer of the classifiers associates the test observations with a certain cluster, whereas the second level identifies the object within the cluster. Data associated with uninformative clusters are screened out. Final identification for the group of observations is obtained using the majority voting rule only from the informative observations.

We apply our algorithm to the gait recognition problem. The hierarchical classifiers are built in four different feature spaces for silhouette images. Final classification is determined by aggregating results from these four feature spaces. The experimental results show that our method results in improved recognition rates in most cases compared with other previously reported methods.

1 Introduction

In many supervised learning problems, the object of interest is represented as a sequence of observations. For instance, in cytological research, living tissue under a microscope is observed as an ensemble of cell images [1]; and in human motion analysis, an individual's movement is captured by cameras as a sequence of image frames [2]. To distinguish tissue cells, or to identify a person from image sequences, two groups of approaches have been applied to classify these observations. In the first group, the dynamics of all observations from an object, e.g., training a hidden Markov model [3], are employed to capture the characteristics of the subject. In the second group of approaches, static features, such as averages and standard deviations in quadratic discriminant function [4], are used to characterize the subject. These "dynamic" and "static"

W. Zhao, S. Gong, and X. Tang (Eds.): AMFG 2005, LNCS 3723, pp. 55–69, 2005.

features, used to classify the subjects, are obtained through "intra-class" analysis of each individual subject, i.e., without "inter-class" comparing the similarities and differences among observations of different subjects. In many cases, some observations from different subjects are very similar, which do not provide any characteristic information useful for classification. More significantly, including these uninformative frames in the classification process tends to smear the differences among subjects, can lead to a weaker classifier due to the potential confusions. Hence, it is necessary to exclude these uninformative frames to enhance the classification accuracy and robustness.

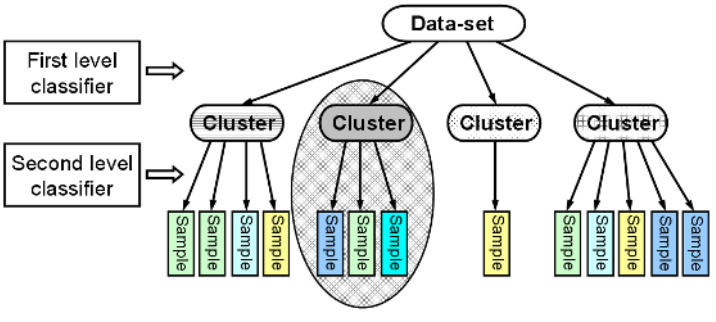

Fig. 1. Outline of the proposed classification framework. The uninformative cluster and the corresponding samples are shown in shadow. Samples from different subjects are color-coded.

A common approach for removing redundant information is feature selection using principle component analysis, independent component analysis etc. [5]. These approaches remove irrelevant features from a given feature set, and thus will modify the feature space for every sample across different objects. However, there has been no systematic research on removing uninformative samples, keeping the feature space representation unchanged.

In this paper, we present a novel combinatorial optimization approach to identify similar observations across subjects, ignoring similarities between frames from the same subject. We construct a multipartite graph structure from observations in the data, where a partite set corresponds to a subject and vertices in a partite set correspond to the samples from an object. Such a graph only specifies similarities between observations from different objects. Then, the problem of identifying similar samples across objects is formulated as a clustering problem on the multipartite graph. We give an efficient algorithm for finding the clusters of uninformative samples.

Subsequently, we build supervised hierarchical classifiers, as illustrated in Fig. 1. The first level classifier performs an assignment of a testing sample to clusters, whereas the second level carries out the actual subject identification.

The above classification framework could be applied to many problems with ensemble data. In this paper, methodological details of this work is demonstrated through the gait recognition problem where objects are the human subjects and the samples correspond to frames/images from the those subjects.

Gait Recognition

To identify a walking subject from an image (frame) sequence, current gait recognition approaches can be classified as either model-based [9,10,11,12,13] or model-free [2,14,15,16,17,18,19,20] based on whether a specific humanoid model is employed. Joint angle trajectories [9,10,11,12,13], body shape parameters [16,19], or other features [2,14,15,17,18,20] extracted from images, were used to represent gait characteristics. When a probe frame sequence is available, the same features are extracted and compared to those in the training data.

Both of these approaches calculate the similarity between the probe sequence and a sequence from any subject in the data, and label the probe sequence as the one with the highest similarity value. However, for subjects represented by silhouette images, or binary images with foreground pixels labeled as 1 and background as 0, very often similar silhouette images occurs among different subjects. Accurate and robust identification can be achieved by training classifiers on frames that are signatures of individual subjects, and excluding the ubiquitously present frames. There is no previous work, to our knowledge, on studying the similarity across frames from different subjects and pinpoint the uninformative ones.

We constructed four feature spaces for silhouette images to formulate of similarity measure between samples. Hierarchical classifiers are built in each feature space. Following classification in each space separately, results in the four feature spaces for all images within the test sequence are aggregated to provide the final classification results.

The organization of this paper is as follows. Section 2 presents a description of the four feature spaces for silhouette images. In Section 3, we present our algorithm for identifying uninformative clusters of similar image frames across subjects. The hierarchical classification process and the aggregation method are described in Section 4. In Section 5, we present the experimental results and discussions, followed by conclusions in Section 6.

2 Feature Spaces for Silhouette Images

The usual inputs in the gait recognition problem are binary silhouette images to eliminate the texture and color information from original images. In this section, we provide a domain-specific feature extraction procedure for the contour points of the human silhouette, and design four different feature spaces for similarity measurement between images. Our classification framework can be applied to other problems by simply modifying the feature spaces.

2.1 Contour Point Detection

Our image features are defined on contour points, as the locations of these points encode a given silhouette image. We design eight filters to detect contour points in silhouette images, as shown in Fig. 2. These detectors are linear filters with various orientations. A foreground point is marked as a contour point only when the convolution between the silhouette image and any of the eight filters is greater than a certain threshold. Compared

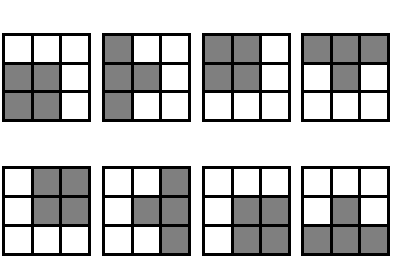

Fig. 2. Eight feature detectors for contour points

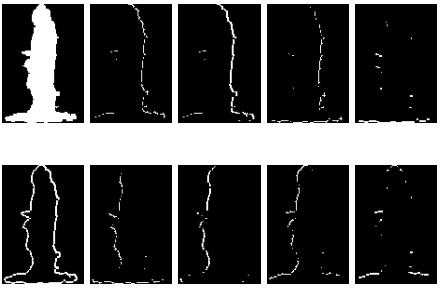

Fig. 3. Upper left panel: original silhouette image. Lower left: combined contour point detection results. Second to last columns: responses to the feature detectors in Fig. 2.

with simple morphological operations such as removing foreground pixels whose four-connected neighbors are foreground as well, these detectors can find the outer contour of the silhouette image and also the orientation of the corresponding boundary. Fig. 3 illustrates the results of this contour point detection process. For the silhouette image in the upper left panel, the responses to the eight filters are shown in the second to last columns. Combining these individual filter responses leads to the complete contour point plot shown in the lower left panel of Fig. 3.

2.2 Relative Shape Vector

To measure the relative location of a contour point within the silhouette, we introduce a new shape descriptor, called Relative Shape Vector (RSV). For any point p within a set of contour points C, all other points in the set $C - \{p\}$ should be located in two of the eight overlapping windows as shown in Fig 4(a). Counting the number of contour points located in each window results in an eight-dimensional vector $(c_1(p, C), \cdots, c_8(p, C))$. The RSV for the point p with respect to set C is a normalized version of this vector, defined as follows:

$$RSV(p, C) = (\alpha_1(p, C), \cdots, \alpha_8(p, C)), \tag{1}$$

where $\alpha_i(p, C) = c_i(p, C)/\sum_{k=1}^{8} c_k(p, C), i = 1, \ldots, 8$. The RSVs for contour points A, B, and C in Fig. 4(b) are given in Fig. 4(c). Note that this is similar to the shape context descriptor [21], except that we have a set of overlapping windows and the histogram

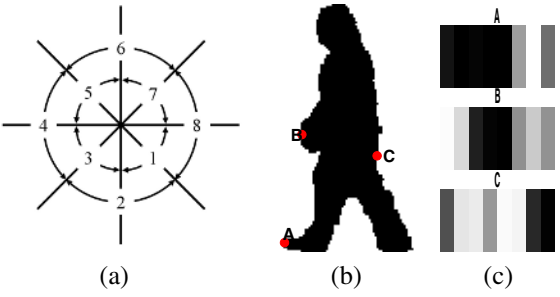

Fig. 4. (a) The eight windows used in Relative Shape Vector calculation. The number inside each section is the window index. (b) Three contour points on a silhouette image. (c) The corresponding eight-dimensional RSVs. Brighter shades means higher value.

is only calculated based on orientations, regardless of distances. Therefore, RSVs preserve redundant information with overlapping windows, and describe less details of the shape by ignoring relative distances, making it less sensitive to changes in the contours.

2.3 The Four Feature Spaces

As the RSV describes the relative location of a given contour point with respect to others, points on different images but at roughly the same relative locations will have similar RSV values. Therefore, grouping in the RSV space results in segmentation of contour points according to their relative locations on a silhouette. Fig. 5 shows the grouping results for two different images using the k-mean algorithm [22], where $k = 13$ is chosen empirically. We see that points at the same relative locations are labeled as the same group. These grouping results enable us to define four feature spaces for the silhouette images:

– Feature Space 1: Gravity Center Distance For each group of points in a given image, we obtain its gravity center. The first feature space contains all distances between different gravity centers, normalized with largest one. With 13 groups, the dimensionality of this feature space is 78.

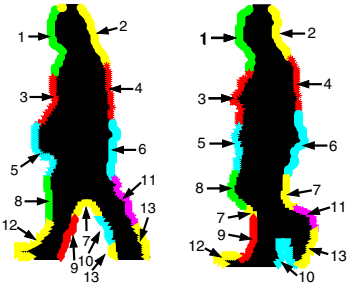

Fig. 5. Grouping results of contour points in two silhouette images. Group numbers are labeled alongside the contour points.

- Feature Space 2: Relative Center Distance
  For each gravity center, we rank its distances to other 12 centers, and obtain two relative distances as the smallest and the second largest distances with respect to the largest one among these 12 distances. This 26-dimensional feature measures relative distances within each image.
- Feature Space 3: Local Orientation Statistics
  The third feature space is designed to capture the boundary orientation information. The orientation of the silhouette image at each contour point is contained in the filter responses to the eight filters shown in Fig. 2. The average response from all points in each group can be interpreted as a histogram of silhouette orientation at the contour points. This 65-dimensional feature captures the local orientation statistics.
- Feature Space 4: Local Shape Statistics
  Another feature we want to capture is the local distribution of contour points within each group. These local shape statistics include the two eigenvalues of the covariance matrix of these points, the ratio of the eigenvalues, and the orientation of the principle components.

The above four feature spaces provide rich descriptions of the shape within a silhouette image and different opportunities for recognition. Similar silhouetted images are close to each other in the feature spaces. Therefore, we are able to cluster images according to their similarities within these four feature spaces.

3 Multipartite Graph Clustering and Filtering

In this section, we describe the clustering method for identifying uninformative frames in a given feature space. Given the representation of frames in a feature space, our goal is to divide the training data frames from multiple subjects into two classes: a set with subject-specific frames, and another with non-subject-specific ones (shared by many subjects). It is likely that filtering out frames that are similar (or uninformative) across subjects would increase the accuracy of classification. Finding such uninformative frames can be cast as a clustering problem on a graph representing similarity between frames from different subjects.

In a traditional clustering scenario, frames close to each other are grouped as a cluster. However, in our problem, similarity between frames from the same subject should not to be considered, as we are only interested in finding clusters containing similar frames from different subjects. Such a similarity structure is captured by a multipartite graph, where we denote a subject as a partite set and frames from a subject as vertices in the corresponding partite set. An illustration of this representation is in Fig. 6, where we show a complete three-partite graph. The multipartite graph representation is suitable because it only considers the similarity relationships between frames from different subjects. The presence of clusters in the multipartite graph reveals the similarity among frames from different objects and those containing very similar frames from a large number of subjects are considered as uninformative.

In graph theory, clusters are related to dense subgraphs called cliques [7]. Therefore, our problem of identifying similar frames across different subjects can be formulated

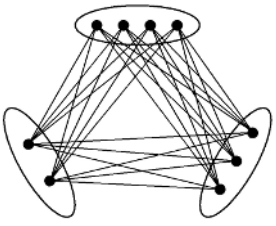

Fig. 6. A complete three-partite graph

as finding cliques in a multipartite graph [6]. However, the maximum clique finding problem, and its various weighted formulations for multipartite graphs, are proven to be NP-complete [8,6]. Due to the sensitivity of the filtering problem and the additional requirement for finding similar frames across different subjects only, existing clustering methods are inadequate. Heuristics are often used to find clusters in data [7] but these heuristic methods are not well suited to our approach since they cannot guarantee finding best multipartite clusters. Therefore, we need an optimization based procedure whose exact solution can be found. In this paper, we propose such a multipartite graph clustering method for finding clusters of uninformative frames. The objective function in our multipartite clustering method have favorable properties which guarantee that a globally optimal solution can be found with an efficient algorithm.

3.1 Multipartite Graph Clustering

For a problem with k subjects in the training data, let $V_i,\ i \in \{1, 2, \ldots, k\}$ denote a set of vertices which represent frames from subject i. We construct an undirected weighted multipartite graph $G = (V, E, W)$, where $V = \cup_{i=1}^{k} V_i$, and $E \subseteq \cup_{i \neq j} V_i \times V_j$ is the set of weighted, undirected edges connecting vertices from different partite sets. The weight $w_{ij} \in W$ on the edge $e_{ij} \in E$ represents the similarity between the vertices i and j.

Let H denote a subset of V such that H contains vertices from at least two partite. We define a score function $F(H)$ to measure the proximity among elements in H. Then, our multipartite quasi-clique, or a cluster, H^*, is defined as the subset with the largest score value, i.e,

$$H^* = \arg\max_{H \subseteq V} F(H). \tag{2}$$

The efficiency of extracting the optimal set H^* is closely related to the algebraic properties of $F(H)$. When the set function $F(H)$ is *quasi-concave*, i.e,

$$F(H_1 \cup H_2) \geq \min(F(H_1), F(H_2)) \ \ \forall H_1, H_2 \subseteq V \tag{3}$$

the optimal solution can be efficiently obtained, if $F(H)$ can be efficiently computed. Here, $F(H)$ is designed using a linkage function $\pi(i, H)$ that measures the degree of similarity of the frame $i \in H$ to other frames in H, leading to the following definition:

$$F(H) = \min_{i \in V} \pi(i, V), \forall i \in V \ \ \forall H \subseteq V \tag{4}$$

In other words, $F(H)$ is the $\pi(i,H)$ values of the least similar (outlier) frame in H. Then, according to (2), the subset, H^* contains frames such that similarity of the least similar frame in H is maximum. It can be proved that $F(H)$ as defined in (4) is quasi-concave if and only if the linkage function $\pi(i,H)$ is *monotonically increasing* [23], i.e,

$$\pi(i,H) \geq \pi(i,H_1) \;\; \forall i, \forall H_1, \forall H : i \in H_1 \subseteq H \subseteq V \tag{5}$$

The linkage function is constructed from the pair-wise similarity values between frames from different subjects. Besides being monotone increasing, the linkage function should be designed such that $\pi(i,H)$ and hence $F(H)$ can be efficiently computed; this would guarantee a polynomial time procedure to find H^*. At the same time, the linkage function should capture an appropriate notion of similarity so that the optimal solution H^* captures the multipartite quasi-clique nature of relationships between frames contained in H^*. Although a large family of linkage functions satisfying the above requirements exists, we use a simple and intuitive linkage function to demonstrate our approach. If m_{ij} $(\geq o)$ is the similarity value between frame i from subject $s(i)$ and frame j from subject $s(j)$ $(s(j) \neq s(i))$, then the linkage function is defined as

$$\pi(i,H) = \sum_{\substack{\ell=1 \\ \ell \neq s(i)}}^{k} \sum_{\substack{j \in H \\ s(j) \neq s(i)}} m_{ij} \tag{6}$$

In other words, the linkage function $\pi(i,H)$ aggregates the similarity between the frame i from subject $s(i)$ and all other frames in H that do not belong to subject $s(i)$.

Table 1. Pseudocode to extract H^*

Step 0: Set $t := 1$; $H_1 := V$; $H^* := V$;
Step 1: Find $M_t := \{i : \pi(i,H_t) = \min_{j \in H_t} \pi(j,H_t)\}$;
Step 2: if $((H_t \setminus M_t = \emptyset) \vee (\pi(i,H_t) = 0 \;\; \forall i \in H_t))$ STOP.
else $H_{t+1} := H_t \setminus M_t$; $t := t+1$;
if $(F(H_t) > F(H^*))$ $H^* = H_t$;
go to Step 1.

The algorithm to find the optimal solution H^* is described in Table 1 [23]. This iterative algorithm begins with the calculation of $F(V)$ and the set M_1 containing the subset of frames that satisfy $F(V) = \pi(i,V)$, i.e, $M_1 = \{i \in V : \pi(i,V) = F(V)\}$. The frames in M_1 are removed from V to obtain $H_2 = V \setminus M_1$. At iteration t, it considers the set H_{t-1} as the input, calculates $F(H_{t-1})$, identifies the subset M_t such that $F(H_{t-1}) = \pi(i_t, H_{t-1}), \forall i_t \in M_t$, and removes this subset from H_{t-1} to produce $H_t = H_{t-1} \setminus M_t$. The algorithm terminates at iteration T when $H_T = \emptyset$ or $\pi(i,H_T) = 0 \; \forall i \in H_T$. It outputs H^* as the subset, H_j with the smallest j such that $F(H_j) \geq F(H_l) \; \forall l \in \{1,2,\ldots,T\}$.

The above formulation gives us one multipartite cluster, however, many such clusters would be present in the multipartite graph. Assuming that these clusters are unrelated to each other, we can use a simple heuristic to find cluster H^*, remove it from

the set V, and find another cluster in the remaining set $V \setminus H^*$. This can be applied iteratively to find all the multipartite clusters.

3.2 Determining Uninformative Clusters

There are two criteria for determining which clusters should be labeled as uninformative. First of all, frames within such clusters should be highly similar to each other. Secondly, these clusters should contain frames from most of the subjects.

From the above procedure, similarity among frames within a cluster is given by the score value $F(H^*)$ of cluster H^*. Since this score value is the aggregated similarity between a frame and all other frames in the cluster, it would, in general, be larger for larger clusters. Hence we use the average similarity value of a cluster defined as $F(H^*)/|H^*|$, instead of $F(H^*)$ for estimating homogeneity within a cluster. Clusters with large average similarity values, but containing frames from just a few subjects are actually informative for classification as they characterize the subjects therein. Hence, only those clusters containing frames where most subjects have an average similarity value above a certain threshold are labeled as uninformative.

4 Details of Classification

Clusters obtained in the previous section provide a hierarchical structure for the entire data, based on which we build our classifiers as shown in Fig. 1. The first level is to determine whether a probe frame is informative, whereas identification within each informative cluster is specified in the second level. In addition, for a group of frames from an unknown subject, we aggregate the identification results from all informative ones to reach a final identification.

4.1 Hierarchical Classifiers

There are a number of standard methods available for binary or multi-class classification. Here we use Bayesian logistic regression(BLR) [24] as the first level of our hierarchical classifier structure and the nearest neighbor as the second level classifier.

First Level - Bayesian Logistic Regression. To determine the cluster for a probe frame, we train a binary classifier for each cluster obtained from Section 3. The Bayesian logistic regression (BLR) [24] is applied to achieve state of the art effectiveness while avoiding over-fitting.

Given a training set containing n samples, $s_1, \ldots, s_n$, with corresponding labels $y_1, \ldots, y_n$, the BLR aims at finding the best parameter value which maximizes the conditional probability model of the form

$$p(y = +1|\beta, s_i) = \psi(\beta^T s_i), \text{ where } \psi(r) = \frac{exp(r)}{1 + exp(r)}$$

is a logistic regression model, and β is a parameter to be determined through the learning process. We choose Laplace distribution as a prior distribution for each component of β. This sparse favoring prior guarantees most components of β have a 0 or near zero value. Fast and accurate prediction could be achieved through the method provided in [24].

Second Level - Nearest Neighbor. If a probe frame is labeled as an uninformative cluster (the cluster shadowed in Fig. 1), this frame does not contain characteristic information, and will be screened out from classification. The second level of classifier is used only for the frames classified as informative.

Within each informative cluster, there are multiple frames from several subjects. We choose the nearest neighbor method due to its simplicity and low bias.

4.2 Classification Rule

To classify an unknown subject $X = \{(x_{11}, x_{12}, x_{13}, x_{14}), \ldots, (x_{n1}, x_{n2}, x_{n3}, x_{n4})\}$ with n frames where $x_{k.} = (x_{k1}, x_{k2}, x_{k3}, x_{k4})$ is the representation for the frame k in the four different feature spaces, we apply four different classifiers C_1, C_2, C_3 and C_4 in corresponding feature spaces. Let m_{ij} denote the number of informative frames assigned by the classifier C_j to subject i in the training data-set. We integrate the classification results from the individual classifiers for estimating the membership r_i of the unknown subject X to the subject i. We used the ordered weighting assignment (OWA) [25] where weights for each classifier are ordered according to the confidence levels of the classifier. Precisely, for estimating the membership of X to i, the m_{ij}'s, $j = 1, \ldots, 4$ are sorted as $m_{ij_1} \geq \ldots \geq m_{ij_4}$ and ordered weighted assignment, r_i^{OWA} is calculated as $r_i^{OWA} = \sum_{l=1}^{4} w_l m_{ij_l}$. It must be noted that m_{ij_l} is different from m_{ij}. The procedure to find the weights w_l will be described shortly. The final classification r^* for X is obtained as the subject which shows highest level of assignment [26] , i.e., $r^* = \arg \max_i r_i^{OWA}$, where i is a subject from the training data.

The weight vector $\mathbf{w} = (w_1, w_2, w_3, w_4)$ used in the above classification rule must satisfy $\sum_{l=1}^{4} w_l = 1$ and $w_l \geq 0 \ \forall l$, where w_l is quantized into ten discrete values, i.e., $w_l \in \{0.1, 0.2, \ldots, 0.9, 1.0\}$. The optimal weights under this restriction are obtained, by exhaustively searching all possible combinations, as the one which maximizes the accuracy of the above classifier under 5-fold cross-validation on the training data.

5 Experimental Results and Analysis

We apply our method to the silhouette images from Human ID Gait Challenge Database collected at the University of South Florida (USF) in May and November 2001. There are a total of 122 subjects in the data, walking on different surfaces (G/C), with different shoes (A/B), with/without carrying a briefcase (BF/NB), and captured with two cameras (R/L) at different times (t_1/t_2). There are 12 carefully designed experiments to test whether identification can be made under different walking conditions, as summarized in Table 2 [27].

The image frames from different subjects are clustered using the multipartite graph clustering procedure described in section 3. On average about 87% of the training data are classified into clusters. Most of the clusters contain frames from a few subjects, therefore, they are considered as informative from classification perspective. Across different feature spaces, we are able to label (on average) about 14% of all frames as uninformative. Figure 7 shows some frames detected as uninformative. As we can see,

Table 2. 12 probe sets with the common gallery of (G, A, R, t_1) containing 122 subjects

Experiment	Probe	Difference
A	G, A, L	View
B	G, B, R	Shoe
C	G, B, L	Shoe View
D	C, A, R	Surface
E	C, B, R	Surface Shoe
F	C, A, L	Surface View
G	C, B, L	Surface Shoe View
H	G, A, R, BF	Briefcase
I	G, B, R, BF	Shoe, Briefcase
J	G, A, L, BF	View, Briefcase
K	G, A/B, R, NB, BF, t_2	Time
L	C, A/B, R, NB, BF, t_2	Surface Time

Fig. 7. A representative instance of uninformative frames from different subjects, in the USF data, within an (uninformative) multipartite cluster

most frames are captured during the mid-swing/mid-stance phase of a gait cycle. This is consistent with our intuition because at this stage of a gait cycle, most body parts overlap with each other and could not show details in the silhouette images.

Further, classifiers are constructed based on the clustering results in the four different feature spaces. The corresponding recognition rates, together with the aggregation results of different feature spaces, are given in Table 3. In addition, Fig. 8 illustrates the aggregation recognition results by the cumulative match score (CMS).

We find that the recognition rates in experiments A and C are consistently better in space 1 (gravity center distance, see section 2.3); however space space 3 (local orientation statistics) provides better recognition in experiments D through L, while space 4 (local shape statistics) seems to capture essential features in experiment B. This suggests that the four feature spaces capture considerably different aspects of the silhouette images, although in most cases, the local orientation statistics is the most relevant feature for recognition. Therefore aggregating results from all four features spaces leads to improved recognition rates, as shown in Table 3 by utilizing all available information.

In Table 4, our results are compared with others reported previously. Our approach considerably outperforms the baseline algorithm [27] in 10 out of the 12 cases, while matching in another one. Our method achieve a much higher recognition rate for experiments K and L, (18% and 15% respectively, comparing to 3% by baseline algorithm), which are the most difficult experiments in the challenge set. The only favorable case

Table 3. Recognition rates for different experiments in each of the four feature spaces as well as the aggregation results

Exp	Space 1	Space 2	Space 3	Space 4	Aggregate
A	70	61	66	52	82
B	81	78	78	83	89
C	69	56	54	63	76
D	19	21	31	20	36
E	36	31	40	34	53
F	12	12	16	13	23
G	22	16	22	16	34
H	32	25	42	25	44
I	40	29	50	40	57
J	25	19	28	20	37
K	12	12	12	12	18
L	6	3	12	9	15

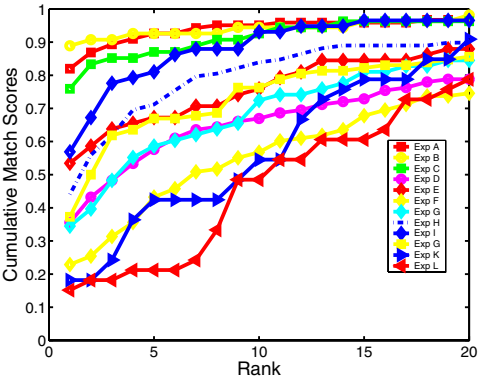

Fig. 8. CMS curves for USF data

for the baseline method is when the subject carries a briefcase in the probe sequence; we attribute this to insufficient information to distinguish subjects when the briefcase contour dominates the subject contour. In comparison to the results from HMM approach [3], our method performs better or equivalently in 5 of the 7 experimental conditions.

We also compare our method with k-nearest neighbor without screening uninformative examples. We choose $k = 5$ in our comparison, and the results show that all of the 12 cases, uninformative frames adversely affect the recognition rate. Further, we employ principle component analysis (PCA), a typical feature selection method, to reduce the feature space before the k-nearest neighbor classification. Though we maintain

Table 4. Comparison of recognition rates of our method with other methods

Exp Id	HMM (UMD) (71 subj.)	Baseline (122 subj.)	kNN (122 subj.)	PCA (122 subj.)	Our work (122 subj.)
A	99	73	75	75	82
B	89	78	78	78	89
C	78	48	56	56	76
D	36	32	34	34	36
E	29	22	40	40	53
F	24	17	16	16	23
G	18	17	22	22	34
H	N/A	61	43	43	44
I	N/A	57	48	48	57
J	N/A	36	32	30	37
K	N/A	3	12	12	18
L	N/A	3	6	6	15

95% total variance of the original data, as we can see, there is no gain in applying PCA, and the recognition rate even decreases in Exp J.

Our method significantly outperforms all other previously reported methods in cases when the subject walks on a different surface or at different time comparing to the training data. This shows our method has more prediction power for unknown walking conditions. Eliminating the common frames across subjects enables us to avoid training on "noisy" (i.e. uninformative from the classification perspective) frames, thereby resulting in a higher recognition rate. In contrast, approaches that attempt to capture strict dynamics of a walking style, such as HMM based methods, cannot isolate the uninformative frames, and hence may perform poorly.

6 Conclusions and Future Work

A new combinatorial framework for supervised classification problem is proposed in this paper. In this framework, the similar samples from a large number of subjects are taken to be uninformative. These uninformative samples are obtained as clusters on a multipartite graph using a novel combinatorial clustering approach. To classify a test subject (probe sequence), its observations (image frames) are first classified as either uninformative or informative; so only the informative frames from the test subject are used for classification.

This framework is applied to the gait recognition problem, and the results demonstrate the efficacy of our method. Currently, we treat sequences of images as a group of frames, without considering the dynamics. However, our framework can be extended to encode the dynamics by introducing another feature space. Furthermore, for color/texture images, different feature spaces should also be developed for other

applications. It would be interesting to see the performance of HMMs when they are trained on more informative data for classification.

Acknowledgment

This work is supported by the National Science Foundation under contract numbers 0200983, 0205671, 0313184, and CCR 0325398. In addition, RZ and AV would like to thank the DIMACS Graduate Student Awards for partial support of this work.

References

1. Cheson, B.D.: Chronic Lymphoid Leukemias. Marcel Dekker (2001)
2. Phillips, P., Sarkar, S., Robledo, I., Grother, P., Bowyer, K.: The gait identification challenge problem: Data sets and baseline algorithm. In: CVPR. (2002)
3. Sundaresan, A., RoyChowdhury, A., Chellappa, R.: A hidden markov model based framework for recognition of humans from gait sequences. In: ICIP. (2003)
4. Webb, A.: Statistical Pattern Recognition. Wiley (2002)
5. Comon, P.: Independent component analysis, a new concept? Signal Processing **36** (1994) 287–314
6. Dawande, M., Keskinocak, P., Swaminathan, J.M., Tayur, S.: On bipartite and multipartite clique problems. J of Algorithms **41** (2001) 388–403
7. Everitt, B.S., Landau, S., Leese, M.: Cluster Analysis, 4th edition. Oxford University Press Inc. (2001)
8. Hochbaum, D.S.: Approximating clique and biclique problems. J of Algorithms **29** (1997) 174–200
9. Cunado, D., Nixon, M., Carter, J.: Using gait as a biometric, via phase-weighted magnitude spectra. In: 1st Int. Conf. audio and video based biometric person authentification. (1997)
10. Yam, C.Y., Nixon, M.S., Carter, J.N.: On the relationship of human walking and running: Automatic person identification by gait. In: ICPR. (2002)
11. Tanawongsuwan, R., Bobick, A.: Gait recognition from time-normalized joint-angle trajectories in the walking plane. In: CVPR. Volume II. (2001) 726–731
12. Niyogi, S., Adelson, E.: Analyzing and recognizing walking figures in xyt. In: CVPR. (1994)
13. Zhang, R., Vogler, C., Metaxas, D.: Human gait recognition. In: IEEE Workshop on Articulated and Nonrigid Motion, in conjunction with CVPR'04. (July 2004)
14. Collins, R., Gross, R., Shi, J.: Silhouette-based human identification from body shape and gait. In: International Conference on Automatic Face and Gesture Recognition. (2002)
15. Murase, H., Sakai, R.: Moving object recognition in eigenspace representation: gait analysis and lip reading. Pattern Recognition Letters **17** (1996) 155–162
16. Huang, P., Harris, C., Nixon, M.: Human gait recognition in canonical space using temporal templates. IEE Proceedings - Vision, Image and Signal Processing **146** (1999) 93–100
17. Kale, A., Cuntoor, N., Yegnanarayana, B., Rajagopalan, A., Chellappa, R.: Gait analysis for human identification. In: Proceedings of the 3rd International conference on Audio and Video Based Person Authentication. (2003)
18. Little, L., Boyd, J.: Recognizing people by their gait: the shape of motion. Videre **1** (1996) 1–32
19. Wang, L., Tan, T., Ning, H., Hu, W.: Silhouette analysis-based gait recognition for human identification. IEEE PAMI **25** (2003) 1505–1518

20. Sunderesan, A., Chowdhury, A.K.R., Chellappa, R.: A hidden markov model based framework for recognition of humans from gait sequences. In: IEEE ICIP. (2003)
21. Belongie, S., Malik, J., Puzicha, J.: Shape matching and object recognition using shape contexts. IEEE Transactions on Pattern Analysis and Machine Intelligence **24** (2002) 509–522
22. Tou, J.T., Gonzalezn, R.C., eds.: Pattern Recognition Principles. Addison-Wesley, Norwell, MA (1974)
23. Vashist, A., Kulikowski, C., Muchnik, I.: Ortholog clustering on a multipartite graph. In: Workshop on Algorithms in Bioinformatics (WABI). (2005)
24. Genkin, A., Lewis, D.D., Madigan, D.: Large-scale bayesian logistic regression for text categorization. J of Machine Learning submitted. (2004)
25. Yager, R., Kacprzyk, J., eds.: The Ordered Weighted Averaging Operators: Theory and Applications. Kluwer Academic Publishers, Reading, MA (1997)
26. Kittler, J., Hatef, M., Duin, R.P., Matas, J.: On combining classifiers. IEEE PAMI. **20** (1998) 226–239
27. Sarkar, S., Phillips, P., Liu, Z., Vega, I., Grother, P., Bowyer, K.: The HumanID gait challenge problem: Data sets, performance, and analysis. IEEE PAMI. **27** (2005) 162–177