2024 33rd IEEE International Conference on
Robot and Human Interactive Communication (RO-MAN)
August 26-30, 2024. Pasadena, California, USA.

2024 33rd IEEE International Conference on Robot and Human Interactive Communication (ROMAN) | 979-8-3503-7502-2/24/$31.00 ©2024 IEEE | DOI: 10.1109/RO-MAN60168.2024.10731422

Feature Selection for Hand Gesture Recognition in Human-Robot Interaction

Matthew McCarver[1], Jing Qin[2] and Biyun Xie[3]

***Abstract*— Hand gesture recognition has been playing an important role in robotic applications, which allows robots to communicate with humans in an effective way. However, it typically desires to process high-dimensional data, such as images or sensor measurements. To address the computational challenges due to the data growth, it is desirable to select most relevant features during recognition by reducing the redundancy of the data. In this paper, we propose a novel feature selection approach based on the separable nonnegative matrix factorization (NMF) framework for hand gesture recognition. In particular, we adopt a nonconvex regularization term, i.e., the ratio of matrix nuclear norm and Frobenius norm. The proposed method reduces the data dimension by utilizing the data low-rankness in an adaptive way. To address the nonconvexity of the proposed model, we reformulate it by introducing an auxiliary variable and then apply the alternating direction method of multipliers (ADMM). Furthermore, a variety of numerical experiments on binary and grayscale hand gesture images demonstrate the efficiency of the proposed feature selection approach in improving the quality of factorization and its potential impact on robotic applications.**

***Index Terms*— feature, hand gesture recognition, low-rank, nonnegative matrix factorization, robot**

I. INTRODUCTION

Nonverbal communication through gestures, motions, facial expressions, etc., plays an important role in human-human interaction. It is also true for human-robot interaction, particularly in noisy environments where verbal communication is not allowable. In this work, we focus on human hand gesture recognition, which will enable robots to understand human intentions accurately through their hand gestures and take the right actions by following the human's instructions.

There are two major types of hand gesture recognition systems based on glove sensors or images/videos. The glove-based recognition system involve wearing a glove or using some form of apparatus that a hand interacts with in order to capture various data, such as motion and pressure plates [1], [2]. The other is visual-based system, either through images or videos are captured via a camera and then interpreted. Although the glove based or tactile hand gesture recognition has a good performance, the equipment can be expensive and is not as convenient or handy as a camera. In this paper, we focus on vision-based recognition systems.

The traditional vision-based gesture recognition system usually directly process images or image-based features learned by some feature descriptors such as Histogram of Oriented Gradients (HOG), Local Binary Patterns (LBP) and Scale-Invariant Feature Transform (SIFT). However, although a dataset with high dimensional features has numerous attributes, it is possible that only a few features are directly relevant to the classification and recognition tasks. Moreover, the growth in data dimensionality poses challenges to classification performance due to the presence of redundant or irrelevant features. Lastly, memory usage and training time are adversely affected when dealing with high-dimensional datasets. Therefore, it is crucial to reduce the feature dimensionality while maintaining the desired recognition accuracy. Recently, many dimensionality reduction methods have been developed to preprocess the input data, resulting in faster processing times without compromising recognition accuracy. Refer to [3], [4] for literature survey of dimension reduction techniques, [5] for a comprehensive review of dimensionality reduction techniques for feature selection and extraction, and [6] for a more recent review on hand gesture recognition in robotic applications.

Nonnegative Matrix Factorization (NMF) has been serving as one of the most popular and efficient dimensionality reduction methods because it reduces the dimensionality of the original data by factorizing the data matrix into two lower-dimensional matrices. These matrices typically have fewer columns than the original data matrix, effectively reducing the dimensionality of the data. NMF is particularly useful for feature extraction and has been widely used in applications such as image and text processing, where the data of interest can be represented as a matrix. In particular, when the factor matrices are separable, NMF becomes a special case known as separable NMF. In separable NMF, the columns of the factor matrices are both nonnegative and sparse, which means they consist mostly of zeros with a few non-zero elements. Recently, many fast algorithms have been developed for solving the separable NMF problem, including Successive Projection Algorithm (SPA) [7], FGNSR [8], and Frank-Wolfe based algorithm [9].

In this paper, we propose a novel feature selection approach based on the separable NMF framework for selecting the most relevant features from a collection of hand gesture images. In particular, we consider a low-rankness promoted regularization technique, i.e., the ratio of matrix nuclear norm and Frobenius norm. By applying the difference of convex function algorithm (DCA), we derive an efficient algorithm with convergence guarantee. As long as the learned features

[1] Matthew McCarver is with Department of Mathematics, University of Kentucky, Lexington, KY 40506, USA mhmc239@uky.edu

[2] Jing Qin is with Faculty of Mathematics, University of Kentucky, Lexington, KY 40506, USA jing.qin@uky.edu (corresponding author)

[3] Biyun Xie is with Faculty of Department of Electrical and Computer Engineering, University of Kentucky, Lexington, KY 40506, USA biyun.xie@uky.edu

Authorized licensed use limited to: UNIVERSITY OF KENTUCKY. Downloaded on August 13,2025 at 21:47:46 UTC from IEEE Xplore. Restrictions apply.

with reduced size are available, a data classifier, such as K-nearest neighbors (KNN) and support vector machines (SVM), can be further applied to classify the images on the reduced feature space. We have conducted a variety of numerical experiments on the binary and gray-scale hand gesture images, which have shown the potential of the proposed approach in realistic robotic applications.

The rest of the paper is organized as follows. Section II describes the proposed gesture recognition method in detail with complexity analysis. Various numerical experiments on binary and grayscale hand gesture images are shown in Section III. Finally, the conclusion and some future works are summarized in Section IV.

II. PROPOSED METHOD

In this section, we describe our proposed gesture recognition approach in details, which consists of two stages. At the first stage, we propose a novel algorithm for selecting the relevant features based on the ratio regularization of the matrix nuclear norm and Frobenius norm. Then at the second stage, classification and recognition tasks are performed on the reduced feature space.

Throughout the paper, we use lowercase letters to denote scalars, e.g., $a \in \mathbb{R}$, boldfaced letters to represent vectors $\mathbf{x} \in \mathbb{R}^n$, capital letters to represent matrices $X \in \mathbb{R}^{m\times n}$, and $\Delta = \{\mathbf{x} \in \mathbb{R}^n \,|\, 0 \leq x_i \leq 1, \sum_{i=1}^n x_i = 1\}$ to represent the unit simplex. We say a matrix $X \in \Delta$ if $\mathbf{x}_i \in \Delta$ where $\mathbf{x}_i$ is the i-th column of the matrix X. We adopt the Matlab notation for matrix indexing. For example, $X(i,:)$ is the i-th row of the matrix X and $X(:,j)$ is the j-th column of the matrix X. The nuclear norm of a matrix X, denoted by $\|X\|_*$, is defined as the sum of all singular values of X. The Frobenius norm of a matrix X, denoted by $\|X\|_F$, is defined as the square root of all the squares of the entries in X. The projection of a matrix X onto the set Λ is denoted as $\Pi_\Lambda(X)$. Let $\mathbb{R}_+^{m\times n}$ be the set of all nonnegative m-by-n matrices.

A. Feature Reduction

Separable NMF aims to decompose a feature matrix into a product of two nonnegative matrix factors while enforcing the separability constraint on the factors. In particular, given $M \in \mathbb{R}_+^{m\times n}$ where each row represents a data point, we assume that there exists a subset $\mathcal{K} \subset \{1, 2, \ldots, n\}$ of r column indices and $H \in \mathbb{R}_+^{r\times n}$ such that

$$M \approx M(:,\mathcal{K})H.$$

This is called the *separability assumption* [10], which has been widely adopted in many separable NMF algorithms [8], [9]. Under the separability assumption, computing the nonnegative matrix factors is reduced to finding the column index subset $\mathcal{K}$. Once $\mathcal{K}$ is identified, H can be computed via alternating least squares optimization [11]. In addition, the obtained index set $\mathcal{K}$ indicates how to select features for further processing.

To find $\mathcal{K}$, we can first rewrite $M(:,\mathcal{K})H$ as MX where $X \in \mathbb{R}_+^{n\times n}$ is obtained by padding H with zero rows. Then the problem of finding $\mathcal{K}$ turns into finding X and then selecting rows of X based on the row-sparsity. For example, a constrained row-sparsity minimization problem has been proposed for seeking X [12], [13]

$$\min_{X\in\mathbb{R}_+^{n\times n}} \|X\|_{\text{row},0} \quad \text{s.t.} \quad \|M - MX\|_F \leq \varepsilon,\, \mathbf{1}^T X = \mathbf{1}^T, \tag{1}$$

where $\|X\|_{\text{row},0}$ counts the number of nonzero rows of X and ε controls the error between M and its approximation MX. The constraint that the columns of X sum to one is imposed to guarantee the solution uniqueness. Due to the NP-hardness, convex relaxation techniques are usually applied to (1) which leads to the separable NMF algorithms based on the convex relaxed models [8] [9]. Motivated by the recent success of a ratio norm, i.e., ℓ_1/ℓ_2 in sparse signal recovery [14] and its low-rank extension [15], we adopt a non-convex low-rankness-promoting regularizer. In particular, we consider the following model to estimate X:

$$\min_{X\in\mathbb{R}_+^{n\times n}} \lambda\frac{\|X\|_*}{\|X\|_F} + \frac{1}{2}\|M - MX\|_F^2 \quad \text{s.t.} \quad X \in \Delta, \tag{2}$$

where Δ is the unit simplex, i.e., $X \in \Delta$ if and only if $\mathbf{1}^T X = \mathbf{1}^T$ and each entry of X is nonnegative. By following the procedure in [14], we split the ratio as the difference of two convex functions and then consider the difference

$$\min_{X\in\mathbb{R}_+^{n\times n}} g(X) - \langle X, \nabla_X h(X^{(k)}\rangle$$

where $g(X) = \lambda\|X\|_* + \frac{1}{2}\|MX - M\|_F^2$ and $h(X) = \alpha\|X\|_F$. By applying the difference of convex function algorithm (DCA) and the fact that $\Delta \subseteq \mathbb{R}_+^{n\times n}$, we obtain the updating scheme at the k-th iteration

$$\begin{aligned} X_{k+1} = \operatorname*{argmin}_{X\in\Delta} \lambda\|X\|_* + \frac{1}{2}\|MX - M\|_F^2 \\ + \frac{\beta}{2}\left\|X - X_k - \frac{\alpha_k}{\beta}\frac{X_k}{\|X_k\|_F}\right\|_F^2. \end{aligned}$$

Here the coefficient α_{k+1} is updated via

$$\alpha_{k+1} = \frac{\|X_{k+1}\|_*}{\|X_{k+1}\|_F}. \tag{3}$$

To solve the X-subproblem, we define the augmented Lagrangian function

$$\begin{aligned} &L_\rho(X,V;Z) = \lambda\|V\|_* + \frac{1}{2}\|MX - M\|_F^2 \\ &+\frac{\beta}{2}\left\|X - X_k - \frac{\alpha_k}{\beta}\frac{X_k}{\|X_k\|_F}\right\|_F^2 + \frac{\rho}{2}\|X - V + Z\|_F^2 + \mathbb{I}_\Lambda(V). \end{aligned} \tag{4}$$

By applying the alternating direction method of multipliers

Authorized licensed use limited to: UNIVERSITY OF KENTUCKY. Downloaded on August 13,2025 at 21:47:46 UTC from IEEE Xplore. Restrictions apply.

(ADMM), we have the following updates:

$$
\begin{cases}
X^{j+1} = \underset{X\in\Delta}{\operatorname{argmin}} \frac{1}{2}\|MX - M\|_F^2 + \frac{\rho}{2}\|X - V^j + Z^j\|_F^2 \\
\qquad + \frac{\beta}{2}\|X - X_k - \frac{\alpha_k}{\beta}\frac{X_k}{\|X_k\|_F}\|_F^2 \\
V^{j+1} = \underset{V\in\Delta}{\operatorname{argmin}} \lambda\|V\|_* + \frac{\rho}{2}\|X^{j+1} - V + Z^j\|_F^2 \\
Z^{j+1} = Z^j + X^{j+1} - V^{j+1}
\end{cases}
\tag{5}
$$

Here the X-update is obtained by solving the critical equation, i.e., setting the gradient with respect to X as zero and then solving for X. We have

$$
\begin{aligned}
X^{j+1} = \left(M^TM + (\beta+\rho)I\right)^{-1}\Big(M^TM + \beta X_k \\
+\alpha_k \frac{X_k}{\|X_k\|_F} + \rho(V^j - Z^j)\Big) := Q^{-1}W.
\end{aligned}
\tag{6}
$$

Since $Q = M^TM + (\beta+\rho)I$ is symmetric and positive definite, we can apply the Cholesky factorization to the matrix Q and get the upper triangular matrix factor R such that $Q = R'R$. Then we can use forward and backward substitutions to solve the linear system $R'RX = W$ for X, which can reduce the computational cost from $\mathcal{O}(n^3)$ to $\mathcal{O}(n^2)$. For the V-update, we can use the singular value thresholding (SVT) operator [16] and get

$$
V^{j+1} = \Pi_\Delta\left(\mathcal{D}_{\lambda/\rho}(X^{j+1} + Z^j)\right), \tag{7}
$$

Here the SVT operator with the parameter λ/ρ is given by $\mathcal{D}_{\lambda/\rho} = U\hat{\Sigma}V^T$ where $\hat{\Sigma} = \operatorname{diag}(\{(\sigma_i - \frac{\lambda}{\rho})_+\})$ provided that σ_i's are the singular values of $X^{j+1} + Z^j$ and the thresholding operator is defined as $(t)_+ = \max\{0, t\}$. We summarize the entire algorithm in Algorithm 1, which includes the post-processing step for selecting the column indices. For the post processing, one can either select $\mathcal{K}$ to be the index set of maximum r diagonal entries of the resulting matrix X, or the row index set with the maximum ℓ_2-norms of X. For the initialization of α_0 and X_0, there are different ways. For example, one can get an initial X_0 using a fast algorithm such as FGNSR [8]. Then once X_0 is obtained, we set $\alpha_0 = \|X_0\|_*/\|X\|_F$. Note that we terminate the inner loop if the relative error between two consecutive inner iterates is smaller than η, and terminate the outer loop if the relative error between the two consecutive outer iterates is smaller than ε.

B. Computational Complexity Analysis

We discuss the computational cost of Algorithm (1). For the X-update (6), we can pre-compute the matrix Q only once which has computational complexity $\mathcal{O}(n^3)$. Then the computational complexity of finding the Cholesky factorization of Q is $\mathcal{O}(n^3)$. Once the upper triangular matrix R is available, updating of X via (6) involves $\mathcal{O}(n^2)$ complexity at each inner iteration. Hence the entire X-update is of complexity $\mathcal{O}(n^3)$. The V-update involves the SVT operator. It has been shown that SVT of a $m \times n$ matrix has computational complexity $\mathcal{O}(mn^2 + n^3)$ [17], thus our $n \times n$ matrix will have complexity $\mathcal{O}(2n^3)$. Next the projection of an n-dimensional vector onto the unit simplex [18] can be computed in $\mathcal{O}(n \log n)$. Since we are projecting an $n \times n$ matrix onto the unit simplex it will be computed with complexity $\mathcal{O}(n^2 \log n)$. Thus the total computational complexity of our algorithm is $\mathcal{O}(3n^3 + n^2 \log n)$.

Algorithm 1 Ratio Norm Minimization for Separable NMF (RN-SNMF)

```
Input: $M \in \mathbb{R}_+^{m\times n}$, number $r$ of columns to extract, parameters $\rho, \beta, \lambda$, maximum number of iterations $K/J$ for DCA/ADMM, and error tolerance $\varepsilon/\eta$ for DCA/ADMM.
Output: $X \in \mathbb{R}_+^{n\times n}$, column indices $\mathcal{K}$.
$X \leftarrow X_0, \alpha \leftarrow \alpha_0$                                  ▷ Initialize
for $k = 1 : K$ do
    $X_k \leftarrow X_{k-1}$                   ▷ Keep previous iterates in memory
    $\alpha_k \leftarrow \alpha_{k-1}$
    for $j = 1 : J$ do                                     ▷ ADMM Updates
        Update $X^{j+1}$ via (6)
        Update $V^{j+1}$ via (7)
        Update $Z^{j+1} = Z^j + X^{j+1} - V^{j+1}$
        if $\|X^j - X^{j+1}\|_F/\|X^j\|_F < \eta$ then
            break
        end if
    end for
    $X_{k+1} = X^j$                                        ▷ Update $X_{k+1}$
    $\alpha_{k+1} = \frac{\|X_{k+1}\|_*}{\|X_{k+1}\|_F}$                               ▷ Update $\alpha$
    $e = \frac{\|X_k - X_{k+1}\|_F}{\|X_k\|_F}$                                ▷ Relative Error
    if $e \le \varepsilon$ then                           ▷ Stopping criterion
        break
    end if
end for
$X = X_k$
$\mathcal{K} = \text{post-processing}(X, r)$   ▷ Post-processing to get $\mathcal{K}$
```

III. NUMERICAL EXPERIMENTS

In this section, we will test the proposed Algorithm (1) on binary and grayscale data sets of hand gesture images. To quantify the performance, we adopt *classification accuracy*, which is defined as $1 - loss$ where loss is the true misclassification cost function used for our classifiers. There are two types of features that we use: (1) histograms of oriented gradients (HOG) feature vector where we get the HOG feature vector for each image and each row of our data matrix is the HOG feature vector for each image, and (2) local binary pattern (LBP) feature vectors for each image so that each row of our data matrix corresponds to the LBP feature vector of that image. The HOG and LBP extractions are implemented in Matlab as well as all the classifiers. The cell size for both HOG and LBP are set as 8. The parameters for Algorithm (1) are $\lambda = 10, \beta = 100, \rho = 100$ in the binary test and optimal λ varies among $\{1, 10^{-1}, \ldots, 10^{-4}\}$ in the grayscale test. The tolerance for the stopping criteria are selected as $\eta = \varepsilon = 10^{-4}$, the maximum number of outer iterations $K = 500$ and the maximum number of inner iterations $J = 5$. We test our algorithm against other related

Authorized licensed use limited to: UNIVERSITY OF KENTUCKY. Downloaded on August 13,2025 at 21:47:46 UTC from IEEE Xplore. Restrictions apply.

separable NMF algorithms, including successive projection algorithm (SPA) [11], FGNSR [8] and Frank-Wolfe (FW) [9]. All the numerical experiments were run in Matlab R2023b on a desktop computer with Intel 12th Gen Intel(R) Core(TM) i5-12500T CPU RAM 16GB with Windows 11.

A. Binary Image Test

For the binary image data downloaded from [19], there are three hand gestures in the database: fist, open-hand, and two finger which have 2003, 2010, and 2005 images, respectively. Each binary image is of size 150 × 150; see Fig. 1 for sampled images. We compute the HOG on the entire data set of 6000 images, and we select a subset of the binary images from each gesture (200 images each), and convert them into a matrix to obtain a matrix of size 600×10404. For each image, we extract the HOG features using Matlab command `extractHOGFeatures` and then obtain matrices of size 6000×10404 and 600×10404. We then remove zero columns from the obtained data matrices and then normalize the columns such that the sum of each column is one, which ensures that the columns of the input matrix lie in the unit simplex Δ. Thus, we end up with matrices of size 6000×10404 for the binary HOG data, and 600×10404, for the binary subset HOG data.

Next we test all the comparing methods with varying factorization ranks, i.e., r being 5%, 10%, 20%, 30%, 40%, and 50% of the columns of a data matrix. For the post-processing step, we select the r largest diagonal entries of X. After the column index set $\mathcal{K}$ is obtained, we perform classification on the reduced feature matrix with a specified factorization rank. Given the $m \times r$ reduced image-feature matrix, we split the data, holding 80% of the images as training data, and the other 20% of the images as testing data using Matlab command `cvpartition`. We then train a SVM model and a KNN model for classification of the binary images on the training reduced image-feature matrix in Matlab using `fitcecoc` and `fitcknn`, respectively. We run 50 trials for each r value, and average the accuracies given from each trial. The results for accuracy for the whole data set is given in Fig. 2 with the runtimes of the algorithms given in Table I. For the subset data, the HOG accuracies are given in Fig. 3 and their runtimes are given in Table II. One can see that the runtime of SPA depends on the matrix size while the proposed algorithm is scalable and quite stable for various matrix sizes. We see that the accuracies the same as SPA, and outperform FGNSR and FW for larger r values. However, the runtime for RN-SNMF is more stable than SPA, as one can see in Table I and II the times for SPA increase dramatically as the matrix size increases, which indicates that RN-SNMF is more scalable than SPA.

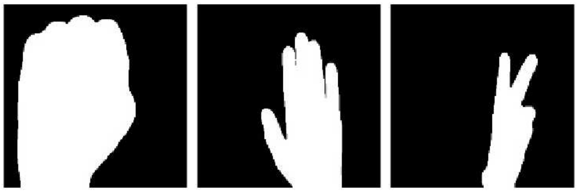

(a) Fist (b) Open-hand (c) Two fingers

Fig. 1: Sample images in the binary hand gesture data.

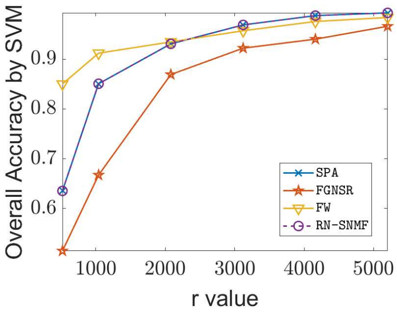


Fig. 2: Classification Accuracies for Binary Data Using HOG Features

r	SPA	FGNSR	FW	RN-SNMF
521	39.07	616.92	1.07E+04	410.91
1041	129.98	897.24	1.05E+04	510.42
2081	397.07	2.30E+03	1.11E+04	472.88
3122	800.69	3.92E+03	1.10E+04	481.41
4162	1.33E+03	6.71E+03	1.12E+04	510.60
5202	1.85E+03	9.64E+03	1.15E+04	396.87

TABLE I: Runtime (in seconds) for Binary HOG Full Data

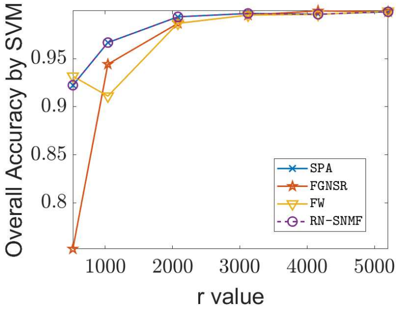


Fig. 3: Classification Accuracies for Binary Subset Data Using HOG Features

r	SPA	FGNSR	FW	RN-SNMF
520	1.50	407	3.53E+03	438.24
1040	1.88	405.75	3.56E+03	336.05
2081	4.00	405.69	3.54E+03	369.86
3121	6.35	392.99	3.52E+03	377.73
4162	8.09	385.72	3.57E+03	369.74
5202	9.41	393.31	3.51E+03	491.12

TABLE II: Runtime (in seconds) for Binary HOG Subset Data

In addition to HOG features, we conduct numerical experiments using the LBP features that are extracted from images. For each image, we obtain the LBP feature matrices of size 6000×19116 and 600×19116. Similar to the HOG feature case, we remove zero columns in the LBP feature matrix and then normalize each column so that they sum to one, ensuring that all the columns of the normalized feature matrix lie in the unit simplex Δ. Thus we end up with matrices of size 6000×18955, and 600×18088 for the respective full and subset LBP features corresponding to the binary data.

Next we test all the algorithms with varying factorization

Authorized licensed use limited to: UNIVERSITY OF KENTUCKY. Downloaded on August 13,2025 at 21:47:46 UTC from IEEE Xplore. Restrictions apply.

ranks, i.e., r being 5%, 10%, 20%, 30%, 40%, and 50% of the columns of a data matrix. For the post-processing step, we select the r largest diagonal entries of X. By the same procedure on the HOG features, We obtain classification accuracies based on the LBP features. The LBP accuracies are given in Fig. 4 and their runtimes are listed in Table III. For the subset data, LBP accuracies are given in Fig. 5 and their runtimes (in seconds) are given in Table IV. Fig. 5 shows that RN-SNMF preforms as well as the other compared algorithms in terms of SVM classification accuracy. When KNN is used as a classifier it outperforms FW and FGNSR for smaller r values. The runtime for RN-SNMF is comparable to FGNSR and much faster than FW, and is insensitive to the matrix size unlike SPA whose runtime increases significantly as the matrix size increases.

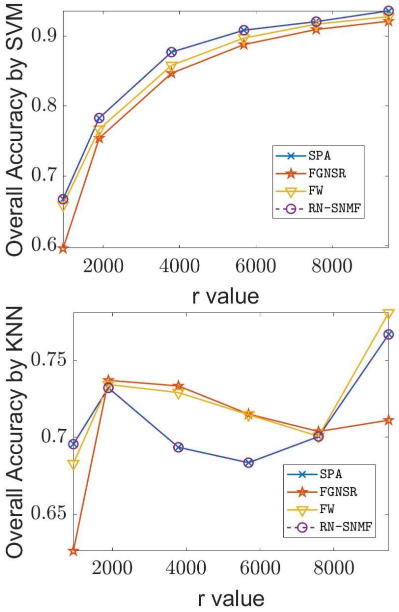


Fig. 4: Classification Accuracies for Binary Data Using LBP Features

r	SPA	FGNSR	FW	RN-SNMF
948	129.73	2.04E+03	3.71E+04	2.06E+03
1896	324.60	3.32E+03	3.61E+04	1.63E+03
3791	1.03E+03	1.00E+04	3.57E+04	1.91E+03
5687	2.02E+03	2.72E+04	3.58E+04	2.15E+03
7582	1.33E+03	2.95E+04	3.38E+04	1.75E+03
9478	3.23E+03	3.07E+04	3.48E+04	1.96E+03

TABLE III: Runtime (in seconds) for Binary LBP Data

r	SPA	FGNSR	FW	RN-SNMF
905	2.77	1.60E+03	1.13E+04	1.79E+03
1809	5.41	1.34E+03	1.15E+04	1.71E+03
3618	9.88	1.22E+03	1.14E+04	1.64E+03
5427	13.71	1.26E+03	1.17E+04	1.72E+03
7236	2.13E+01	1.48E+03	1.18E+04	1.47E+03
9044	2.70E+01	1.66E+03	1.19E+04	1.88E+03

TABLE IV: Runtime (in seconds) for Binary LBP Subset Data

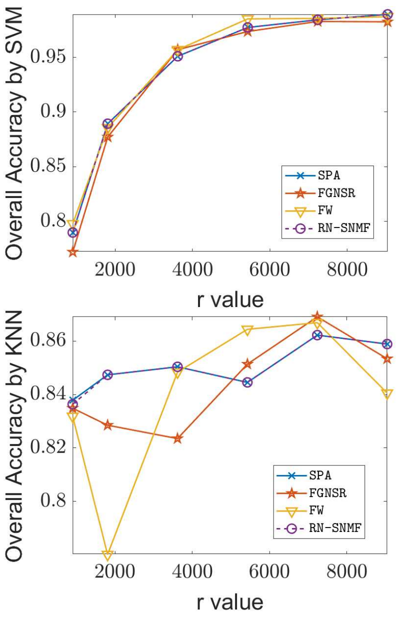


Fig. 5: Classification Accuracies for Binary Subset Data Using LBP Features

B. *Grayscale Image Test*

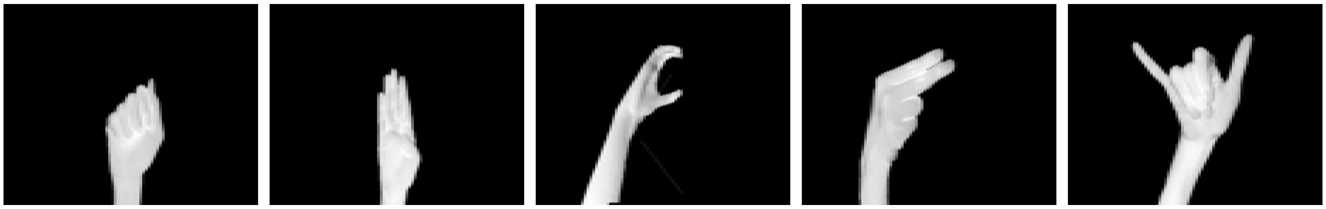
For this experiment, we use the HGM-4 multi-cameras dataset [20][1]. In particular, we choose five classes of images representing hand gestures for the letters A, B, C, H, and Y, as these gestures are most distinguishable for fair comparisons; see one example image for each gesture in Fig. 6. Each class of gestures have 40 images, each with size 160 × 90. In addition, we expand the data set by image rotation. Specifically, for each image, we rotate the images by a degree from $\{-2^\circ, -1^\circ, 1^\circ, 2^\circ\}$ and then extract the HOG features from all the images including rotated ones. Thus we can generate a feature matrix of size 1000×6840. We remove zero columns and normalize each column so they lie in Δ as done in the previous section, which leads to a feature matrix of size 1000×6840. For the post-processing step that selects the index set $\mathcal{K}$, we select the row indices corresponding to the r largest ℓ_2 norms of the rows of X. We compute the classification accuracy for each algorithm with 50 trials. The comparison of average accuracy by using the SVM classifier for each method is illustrated in Fig. 7 and the runtime for each algorithm is given in Table V, which shows our method can achieve an optimal balance between classification accuracy and runtime.

r	SPA	FGNSR	FW	RN-SNMF
342	0.29	137.386	831.01	409.89
684	0.50	144.90	816.90	495.50
1368	0.55	157.29	865.85	491.78
2052	1.74	127.96	840.62	405.86
2736	1.69	135.19	915.68	413.31
3420	1.81	128.87	865.63	488.89

TABLE V: Runtime (in seconds) for Gray-Scale HOG

[1] The data set is available at `https://data.mendeley.com/datasets/jzy8zngkbg/4`

Authorized licensed use limited to: UNIVERSITY OF KENTUCKY. Downloaded on August 13,2025 at 21:47:46 UTC from IEEE Xplore. Restrictions apply.



(a) A (b) B (c) C (d) H (e) Y

Fig. 6: Gray-Scale hand gestures

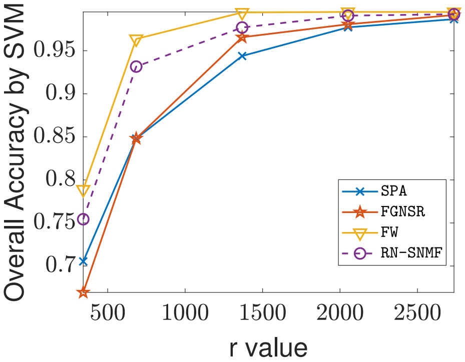


Fig. 7: Classification Accuracies for Gray Scale Data Using HOG Features

IV. Conclusions and Future Work

In this paper, we investigate feature dimension reduction for hand gesture recognition. We propose a novel feature selection method within the separable NMF framework, utilizing a low-rankness promoted regularizer in the form of a ratio of matrix nuclear norm and Frobenius norm. To solve the resulting minimization problem, we employ DCA and ADMM, which allow for closed-form solutions to each subproblem. Our numerical experiments on binary and grayscale gesture images, using various feature types, demonstrate the effectiveness of our proposed method in accurately recognizing hand gestures.

In the future, we plan to extend this framework to dynamic gesture recognition and apply it to complex RGB or RGBD data sets. In addition, to address the computational expense of finding the SVD of large matrices, we could use randomized SVD to speed up feature selection in real-time recognition for human-robot interaction.

Acknowledgments

The research of McCarver and Qin is supported by the NSF grant DMS-1941197, and and the research of Xie is supported by the NSF under Grant #2205292 as well as NASA and the NASA Kentucky EPSCoR Program under NASA award number 80NSSC22M0034.

References

[1] S. Jiang, P. Kang, X. Song, B. Lo, and P. Shull, "Emerging wearable interfaces and algorithms for hand gesture recognition: A survey," *IEEE Reviews in Biomedical Engineering*, vol. PP, pp. 1–1, 05 2021.

[2] L. Guo, Z. Lu, and L. Yao, "Human-machine interaction sensing technology based on hand gesture recognition: A review," *IEEE Transactions on Human-Machine Systems*, vol. 51, no. 4, pp. 300–309, 2021.

[3] I. K. Fodor, "A survey of dimension reduction techniques," Lawrence Livermore National Lab.(LLNL), Livermore, CA (United States), Tech. Rep., 2002.

[4] G. T. Reddy, M. P. K. Reddy, K. Lakshmanna, R. Kaluri, D. S. Rajput, G. Srivastava, and T. Baker, "Analysis of dimensionality reduction techniques on big data," *Ieee Access*, vol. 8, pp. 54 776–54 788, 2020.

[5] R. Zebari, A. Abdulazeez, D. Zeebaree, D. Zebari, and J. Saeed, "A comprehensive review of dimensionality reduction techniques for feature selection and feature extraction," *Journal of Applied Science and Technology Trends*, vol. 1, no. 1, pp. 56–70, 2020.

[6] L. Guo, Z. Lu, and L. Yao, "Human-machine interaction sensing technology based on hand gesture recognition: A review," *IEEE Transactions on Human-Machine Systems*, vol. 51, no. 4, pp. 300–309, 2021.

[7] N. Gillis and S. A. Vavasis, "Fast and robust recursive algorithmsfor separable nonnegative matrix factorization," *IEEE transactions on pattern analysis and machine intelligence*, vol. 36, no. 4, pp. 698–714, 2013.

[8] N. Gillis and R. Luce, "A fast gradient method for nonnegative sparse regression with self-dictionary," *IEEE Transactions on Image Processing*, vol. 27, no. 1, pp. 24–37, 2018.

[9] T. Nguyen, X. Fu, and R. Wu, "Memory-efficient convex optimization for self-dictionary separable nonnegative matrix factorization: A frank–wolfe approach," *IEEE Transactions on Signal Processing*, vol. 70, pp. 3221–3236, 2022.

[10] S. Arora, R. Ge, R. Kannan, and A. Moitra, "Computing a nonnegative matrix factorization – provably," in *Proceedings of the Forty-Fourth Annual ACM Symposium on Theory of Computing*, ser. STOC '12. New York, NY, USA: Association for Computing Machinery, 2012, p. 145–162.

[11] N. Gillis, "The why and how of nonnegative matrix factorization," *Regularization, Optimization, Kernels, and Support Vector Machines*, vol. 12, 01 2014.

[12] E. Esser, M. Moller, S. Osher, G. Sapiro, and J. Xin, "A convex model for nonnegative matrix factorization and dimensionality reduction on physical space," *IEEE Transactions on Image Processing*, vol. 21, no. 7, pp. 3239–3252, 2012.

[13] E. Elhamifar, G. Sapiro, and R. Vidal, "See all by looking at a few: Sparse modeling for finding representative objects," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, 2012, pp. 1600–1607.

[14] C. Wang, M. Yan, Y. Rahimi, and Y. Lou, "Accelerated schemes for the l_1/l_2 minimization," *IEEE Transactions on Signal Processing*, vol. 68, pp. 2660–2669, 2020.

[15] K. Gao, Z.-H. Huang, and L. Guo, "Low-rank matrix recovery problem minimizing a new ratio of two norms approximating the rank function then using an admm-type solver with applications," *Journal of Computational and Applied Mathematics*, vol. 438, p. 115564, 2024.

[16] J.-F. Cai, E. J. Candès, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM Journal on Optimization*, vol. 20, no. 4, pp. 1956–1982, 2010.

[17] E. J. Candès, C. A. Sing-Long, and J. D. Trzasko, "Unbiased risk estimates for singular value thresholding and spectral estimators," *IEEE Transactions on Signal Processing*, vol. 61, no. 19, pp. 4643–4657, 2013.

[18] Y. Chen and X. Ye, "Projection onto a simplex," *arXiv:1101.6081*, 2011.

[19] S. Goyal, "Hand gesture recognition database," 2020. [Online]. Available: https://www.kaggle.com/datasets/shivamgoyal1899

[20] V. T. Hoang, "Hgm-4: A new multi-cameras dataset for hand gesture recognition," *Data in Brief*, vol. 30, p. 105676, 2020.

Authorized licensed use limited to: UNIVERSITY OF KENTUCKY. Downloaded on August 13,2025 at 21:47:46 UTC from IEEE Xplore. Restrictions apply.