HANKEL SUBSPACE METHOD FOR EFFICIENT GESTURE REPRESENTATION

Bernardo B. Gatto[1], *Eulanda M. dos Santos*[2], *Anna Bogdanova*[3]

{*bernardo*[1], *emsantos*[2]}@*icomp.ufam.edu.br*, *bogdanova@heritage.tsukuba.ac.jp*[3]
Federal University of Amazonas, Brazil[1,2]
Tsukuba University, Japan[3]

ABSTRACT

Gesture recognition technology provides multiple opportunities for direct human-computer interaction, without the use of additional external devices. As such, it had been an appealing research area in the field of computer vision. Many of its challenges are related to the joint complexity of human gestures, which produce inconsistent distributions under different viewpoints. In this paper, we introduce a novel framework for gesture recognition, which achieves high discrimination of spatial and temporal information while significantly decreasing the computational cost. The proposed method consists of four stages. First, we generate an ordered subset of images from a gesture video, filtering out those that do not contribute to the recognition task. Second, we express spatial and temporal gesture information in a compact trajectory matrix. Then, we represent the obtained matrix as a subspace, achieving discriminative information, as the trajectory matrices derived from different gestures generate dissimilar clusters in a low dimension space. Finally, we apply soft weights to find the optimal dimension of each gesture subspace. We demonstrate practical and theoretical gains of our compact representation through experimental evaluation using two publicity available gesture datasets.

***Index Terms*—** Hankel matrices, gesture recognition

1. INTRODUCTION

Gesture recognition is an especially compelling research area in the computer vision field because it enables users to communicate with machines without additional hardware allowing more accessible and effective human-computer interaction. To develop such systems in which preassigned set of human joint motions can be used to transmit meaningful information or to navigate virtual environments, a highly efficient gesture recognition framework is needed. Moreover, to ensure accuracy across the variety of domains, gesture recognition applications often require state-of-the-art machine learning techniques.

The task to operate machines through gesture recognition involves many challenges. For example, same gesture identities produce inconsistent distributions from different viewpoints due to their complex joint structures. The recognition and estimation of a gesture are further complicated by masked or corrupted regions in the imagery, illumination conditions, camera angle and pose. In order to solve these problems, several methods have been introduced for gesture recognition. Among them, discriminative canonical correlation analysis (DCC) [1, 2], hidden Markov models (HMM) [3], orientation histograms [4], color based-models [5], dynamic time Warping (DTW) [6], and silhouette geometry based models [7].

Majority of methods in the literature approach gesture recognition with initial extracting of relevant information for classification, thus employing a variety of preprocessing techniques. For instance, HOG [8], SIFT [9], SURF [10] and LBP [11] can perform efficient feature extraction. However, preprocessing may increase the framework complexity, restricting its applications in the conditions where the hardware is limited.

Furthermore, some applications use external hardware to improve the recognition performance. In [12], Kinect^{TM} sensor significantly improves the overall application performance by making use of the depth data to decrease the complexity of segmenting the gesture joints. Alternatively, there are constantly expanding options to utilize more sophisticated devices, such as gloves with accelerometers [13] or Leap Motion Controller^{TM} [14]. Although these methods demonstrate high performance, we understand that such devices may increase the cost of the system both computationally and economically. Therefore, in our proposed method, we prefer to narrow our instruments to only machine learning techniques on raw images, without making use of external hardware or preprocessing techniques.

Among the solutions that do not require preprocessing techniques, subspace-based methods had recently come to the forefront in the research literature. Such methods have been employed in several computer vision applications, including face recognition [15, 16], object recognition [17, 18] and hand shape recognition [19, 20]. The advantage of the subspace-based methods lies in the utility of working directly on the raw images, as they do not require feature extraction to classify image set distributions. Instead, subspace-based methods work directly on the pixel level of the images, creating a very light, hence efficient, representation. For instance, in [21, 22], it has been argued that this approach is very robust since the subspace of an image-set generates distinctive clusters in a low dimensional space. Based on this observation, our assumption is that we can also represent gesture images as subspaces by employing Hankel matrix achieving more straightforward covariance representation.

Although subspace-based methods demonstrate high accuracy and low computational cost when applied to image set recognition, in general, they are incapable of handling the temporal information, as required for an efficient gesture representation. To address this issue, we suggest a new method based on clustering and sample selection, which preserves the temporal information while reducing the computational complexity. This new presentation makes use of the Hankel matrix formulation, where the image patterns can be stored in such a manner that the ordering of the images is preserved. In our method, we choose a representative sample from each image gesture set and then construct its corresponding Hankel matrix. The resulting covariance matrix is much smaller compared to traditional methods, and its basis vectors can be more easily extracted.

Another weakness of subspace-based methods is that the different class subspaces, usually, are handled equally regardless of their intrinsic dimensions. More precisely, subspace-based methods assume that all of the classes have the same dimensions, which leads to several problems, such as the loss of discriminative and representative features. For instance, we can infer that different distributions have different accumulated energy in each eigenvector. Some classes may have a high compactness ratio in only the first 4 eigenvectors, achieving a very efficient representation. However, some classes may have a high spread ratio energy over its eigenvalues, where only 4 eigenvectors are not sufficient to represent such classes. Therefore, we also propose an automatic method to weight the basis vectors of each image class, to better preserve its intrinsic dimension.

The contributions of this study to the literature are: (1) A novel framework for gesture recognition, with no preprocessing techniques, requiring low computational resources. (2) A new representation for gesture recognition, where the samples are dynamically selected, creating a very compact representation. In addition, by employing the Hankel matrix, this new representation is able to preserve temporal information. (3) An automatic approach for basis vector weighting based on accumulated energy strategy. In this solution, we employ all the basis vectors available for classification, without parameter tuning.

This paper is organized as follows. Section 2 introduces the compact Hankel matrix representation for gesture recognition and the construction of the automatic soft weights for subspaces. Section 3 shows the experimental results. Finally, Section 4 presents the conclusions.

2. PROPOSED METHOD

We begin this section by addressing the problem of gesture recognition from the sets of images. Next, we describe the application of the Hankel matrix to represent the ordered image set and introduce the procedure of creating Hankel subspaces. After that, we show the method of extracting the basis vectors of a given Henkel matrix, while selecting the samples for improving the processing time. This will be followed by the introduction of the dynamic soft weights and the discussion of their advantages over the conventional method. Finally, we explain the procedure of matching two Hankel subspaces to compute their similarity.

2.1. Problem Formulation

Given a set of gesture images, which can be described as $A = \{A_i\}_{i=1}^{M}$, where A_i is an image, we define that A is ordered, so $A_1 \preceq A_2 \preceq A_3 \preceq \ldots \preceq A_M$. Then, we assume that there is a linear mapping that represents A set in terms of its variance, preserving its spatial and temporal information. This linear transformation is performed in such way that the M gesture images are converted into $k-$dimensional orthonormal vectors ordered by their accumulated energy. This new representation, $\mathit{\Phi}_A = \{\phi_i\}_{i=1}^{k}$, provides a more compact manner to represent set A and its computational classification cost is therefore, greatly reduced. The $\mathit{\Phi}_A$ set spans a reference subspace P_A. In literature, $k \ll M$, where discriminative information may be lost. In our proposed method, $k = M$, as all the obtained basis vectors will be employed to create a subspace and a weight will be assigned to each basis vectors ϕ_i regarding its variance. Finally, for a given gesture image set $Y = \{Y_i\}_{i=1}^{N}$, where $Y_1 \preceq Y_2 \preceq Y_3 \preceq \ldots \preceq Y_N$, the task is to compute a subspace Q_Y that represents Y in terms of its variance, preserving its spatial and temporal information and calculate the similarity between Q_Y and P_A.

2.2. Hankel Matrix-based Gesture Representation

Let $A = \{A_i\}_{i=1}^{M}$ representing a gesture to be handled as a time series of vectors, where $A_1 \preceq A_2 \preceq A_3 \preceq \ldots \preceq A_M$. This temporal series can be regarded as the output of a Linear Time Invariant (LTI) system of unknown parameters [23].

It is well known [24] that, given a sequence of output measurements $A = \{A_i\}_{i=1}^{M}$, its associated truncated block-Hankel matrix is:

$$\widetilde{H}_A = \begin{bmatrix} A_1, & A_2, & A_3, & \ldots, & A_{m+1} \\ A_2, & A_3, & A_4, & \ldots, & A_{m+2} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ A_{n-1}, & A_n, & A_{n+1}, & \ldots, & A_M \end{bmatrix}, \quad (1)$$

where n is the maximal order of the system, M is the temporal length of the sequence, and it holds that $M = n + m - 1$.

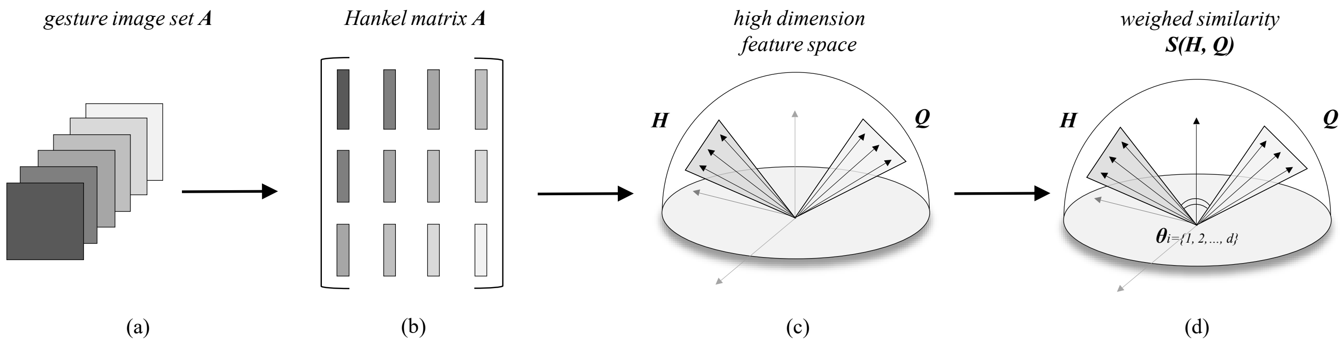


Fig. 1. Conceptual figure of our method. (a) An ordered subset of images representing a gesture A is handled, where a selection criterion is employed to reduce the number of images. (b) Then, the Hankel matrix A is created from the set of the selected images. (c) After that, we extract the basis vectors from the Hankel matrix A to produce the subspace H and its soft weights. (d) Finally, the soft weights are employed to achieve the structural similarity between H and a reference subspace **Q**.

Finally, the Hankel matrix can be normalized as follows:

$$H_A = \frac{\widetilde{H}_A}{\sqrt{||\widetilde{H}_A \cdot \widetilde{H}_A^T||_F}}. \tag{2}$$

2.3. Creating Hankel Subspaces

In order to represent an ordered image set $A = \{A_i\}_{i=1}^M$ in terms of a subspace and to preserve its spatial and temporal information, we introduce the concept of Hankel subspace for gesture recognition.

Subspace-based methods exploit the fact that a set of images lies in a cluster, which can be efficiently represented by a set of orthonormal basis vectors [21]. Our assumption is that the same formulation can be regarded for Hankel subspaces and, accordingly, we can achieve a novel representation for gesture based image recognition.

Therefore, given a normalized Hankel matrix H_A from the ordered image set $A = \{A_i\}_{i=1}^M$, we can compute an autocorrelation Hankel matrix as:

$$C_A = H_A H_A^T \tag{3}$$

When $C_A \in R^{k \times k}$, its eigendecomposition generates a set of eigenvectors $\Phi_A = \{\phi_i\}_{i=1}^k$ that spans a subspace P_A.

2.4. Selecting Samples

When creating a Hankel matrix, the number of images contained in a set and its dimension are crucial factors in terms of computational resources. In order to alleviate this issue, we introduce two approaches based on sample selection.

Random sample selection: In this approach, we randomly select images from the set, preserving its original order. We adopt this temporal sampling scheme in the image sequence since nearby images hardly change their appearance, containing a lot of redundant information to identify the gesture that is being performed. This strategy also allows us to perform the sample reduction with a straightforward implementation.

Clustering selection: The second approach employs a clustering strategy, where the centroids obtained by $k-$means clustering are employed to represent the set, decreasing its number of images. The use of $k-$means clustering was previously adopted for kernel dimensionality reduction in [25]. The advantage of using clustering is that the k centroids of the clusters will represent most of the relevant gesture information for discrimination, eliminating redundant images, achieving a good accuracy with low computational cost.

2.5. Computing the Soft Weights

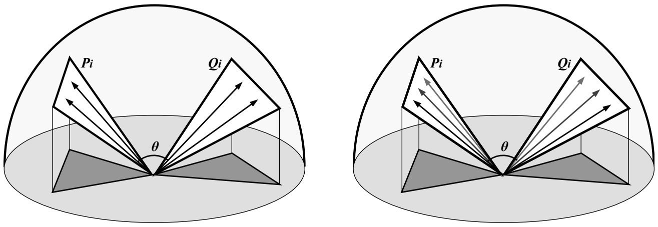
In gMSM, all the eigenvectors are employed to represent a subspace. However, each eigenvector has its own weight, which is computed as follows; let $\Lambda_m = diag(\lambda)$ be the eigenvalues of C_A in descending order, the design of the soft weights is performed according to these eigenvalues. Let $\Omega_A = diag(w)$ be a diagonal matrix of soft weights:

$$\omega = w_m(\lambda) = \min \left[\frac{\lambda}{\lambda_m}, 1 \right], \tag{4}$$

where w_m is the $m-$th eigenvalue in λ. This soft weighting evaluates the importance of each eigenvector as a basis in the subspace by the variance relative to λ_m. The m first values of the diagonal matrix Ω_A will be unity and the remainder will be proportionally decreasing with the $m-$th eigenvalue.

In gMSM, each class subspace P_i uses the same parameter m. In general, this value is set from 1 to 4 in order to evaluate the importance of the eigenvectors in each subspace.

In contrast to gMSM, in HMS we employ an automatic approach to set the value of m. We adopt a heuristic based



(a) (b)

Fig. 2. Difference between MSM and gMSM. (a) The concept of MSM, where the subspace dimensions of P_i and Q_i are empirically obtained. (b) The concept of gMSM, where soft weighting evaluates the importance of each eigenvector. gMSM employs all the basis vectors produced by PCA.

on the interpretation that eigenvectors corresponding to the eigenvalues larger than the average eigenvalues have high representative information. Let us denote λ_i as the $i-$th eigenvalue corresponding to the $i-$th eigenvector. The average eigenvalue μ_A is:

$$\mu_A = \frac{1}{k}\sum_{i=1}^{k}\lambda_i. \tag{5}$$

Next, let us consider that λ_j is the smallest eigenvalue corresponding to the $j-$th eigenvector that satisfies $\lambda_j \leq \mu_A$. Then, we set $m = j$. As in gMSM, these weights are unitary and the remainder eigenvectors will be proportionally decreasing. This approach has several advantages. First, the computational cost required to set the m parameter is largely reduced, as we do not have to set m by parameter tuning. Second, each class subspace P_i will achieve a different set of weights Ω_i, regarding the spread of energy over the eigenvectors. Conceptual illustration of the differences between MSM and gMSM is shown in Figure 2

2.6. Hankel Subspaces Matching

After obtaining the Hankel subspaces and its weights, we can compute the similarity between the subspaces. This procedure is achieved by applying canonical angles or principal angles [26]. In subspace-based methods, we consider that if the distance between two subspaces is small enough, then these subspaces similar to each other.

In practical terms, let $\Phi_A = \{\phi_i\}_{i=1}^{k}$ and $\Psi_Y = \{\psi_i\}_{i=1}^{k}$ span two $k-$dimensional subspaces P_A and Q_Y. Then, let $S(P_A, Q_Y) = \{0 \leq \theta_1 \leq \theta_2 \leq \ldots \leq \theta_n \leq \pi/2\}$ represents the set of angles between P_A and Q_Y.

A practical approach to determine $S(P_A, Q_Y)$ is by computing the $\Lambda = \{\lambda_1, \lambda_2, \ldots, \lambda_k\}$ eigenvalues of:

$$R = \Omega_A \Phi_A^T \Psi_B \Omega_B. \tag{6}$$

Then, the canonical angles:

Table 1. Evaluated methods and their average accuracy.

Methods	Cambridge [27]	HCI [11]
MSM [21]	$61.5\% \pm 0.6$	$56.7\% \pm 0.8$
DCC [1]	$82.0\% \pm 0.3$	$77.3\% \pm 0.4$
gMSM [17]	$75.5\% \pm 0.5$	$66.1\% \pm 0.7$
GDS [20]	$76.0\% \pm 0.3$	$71.4\% \pm 0.4$
HSM-I (random)	$77.6\% \pm 0.4$	$74.1\% \pm 0.6$
HSM-II ($k-$means)	$81.6\% \pm 0.3$	$76.4\% \pm 0.4$
gHSM	**84.9%** $\pm$ **0.3**	**79.1%** $\pm$ **0.4**

$$\theta_i = \{\cos^{-1}\lambda_1, \cos^{-1}\lambda_2, \ldots, \cos^{-1}\lambda_k\}, \tag{7}$$

are employed to compute the structural similarity between soft weighted Φ_A and Ψ_B Hankel subspaces as follows:

$$S(\Phi_A, \Psi_B)_M = \frac{1}{M}\sum_{i=1}^{M}\cos^2\theta_i, \tag{8}$$

the structural similarities between Hankel subspaces are more robust to noise, such as illumination variations and point-of-view in sets of gesture images.

3. EXPERIMENTAL EVALUATION

In this section we show the experimental results of our proposed method. We employed Cambridge gesture dataset [27] for general gestures classification and Human-Computer Interaction (HCI) dataset [11] containing computer interface gestures. In our experiments, we performed leave-one-out cross-validation. After reporting the results for HSM-I (random sample selection), HSM-II ($k-$means sample selection) and gHSM (generalized version of HSM-II), we compare HSM and variants with several state-of-the-art subspace-based methods: mutual subspace method (MSM) [21], discriminative canonical correlation analysis (DCC) [1], generalized mutual subspace method (gMSM) [17] and generalized difference subspace (GDS) [20]. As HSM-I depends on a random selection and HSM-II depends on the initial conditions of $k-$means clustering, for these methods, we performed each experiment 20 times. We report the average of these results.

The Cambridge gesture dataset: consists of 9 classes of gestures. In total, there are 900 video sequences which are partitioned into 5 different illumination subsets. We reduce the size of the video frame to 20×20 pixels and then converted the images to grayscale. Each class contains 100 image sequences with 5 different illuminations and 10 arbitrary motions performed by 2 subjects.

Human-Computer Interaction (HCI) dataset: consists of both static and dynamic hand gestures according to mouse functionalities: cursor, left click, right click, mouse activation, and mouse deactivation. The dataset is divided into 2 sets, the first one has no information regarding the temporal segmentation of the frames and the second is properly segmented. In our experiments, we employed the second image set, where region of interest and label information are available. This set contains 30 labeled video sequences, which are performed by 6 different individuals, each video sequence contains in average 75 images. We reduce the size of the video frame to 20×20 pixels and then converted the images to grayscale.

Table 1 shows the results of the different evaluated methods for gesture recognition. Among the methods that do not employ Hankel matrices, DCC and GDS exhibit high discriminative power comparing to MSM and gMSM. This is a result that both DCC and GDS employ discriminative spaces, where more informative features may be extracted. On the other hand, MSM and gMSM rely only on affine subspaces, where no discriminative scheme is adopted.

HSM-I achieved competitive accuracy, similar to DCC. This indicates that the temporal information extracted by the Hankel representation is very powerful, even when random samples are selected, the main concern here is that the selected samples should preserve its temporal order. HSM-II achieved a higher accuracy than HSM-I, demonstrating that $k-$means clustering is more efficient than random sample selection. This is an expected result, as selecting the centroids obtained by $k-$means is more likely to preserve the structural information of the gesture manifold than random selection. gHSM demonstrated the higher accuracy among the evaluated methods, indicating that the weighted structural similarity between subspaces extracted from Hankel matrices is very efficient for gesture recognition from image sets.

From the Table 1 we observe that all the methods presented a sharp drop in accuracy when comparing the results from the Cambridge dataset and HCI dataset. This is a consequence of the different background from each dataset. In Cambridge dataset, the gesture images where collected in a controlled background, different from the HCI dataset, where the images were recorded in an unconstrained background.

As final remark, we would like to emphasize that HSM and variants do not employ any learning scheme, different from DCC and GDS, where a discriminant space is employed in order enhance the discriminability among the gesture classes. This demonstrates the effectiveness of employing the Hankel subspace for gesture representation.

4. CONCLUSIONS

We presented a novel framework for gesture recognition from sets of images. In our work, we introduce a compact representation based on sample selection, Hankel matrix, and automatic soft weighting. We have evaluated two strategies for reducing the number of samples per image set class. While in the first one, we select the sample images randomly, in the second approach, we pick the most representative samples by using k-means clustering. The selected images are then adopted to construct a Hankel matrix preserving the information about the temporal relation between them. We proceed by extracting the eigenvectors from the Hankel matrix, achieving a highly compact representation. Further, we proposed a modified version of the soft weighs procedure aimed at optimizing the use of the basis vectors from the Hankel matrix. We were able to experimentally show that instead of manually selecting the number of eigenvalues, where their weights are set to 1, using the average of the eigenvalues as a soft threshold we can accomplish high recognition rates, without searching for all of the basis vector combinations. This novel approach shows higher recognition rate compared to the conventional method. Finally, comparing the sample selection strategies, the random selection demonstrated high efficient time and competitive recognition rate. By selecting the samples using $k-$means clustering, we achieved a superior recognition rate compared to all of the investigated methods.

One possible way to improve the performance of the proposed framework is to apply a discriminative selection criterion to achieve a more optimal assortment of samples from each class of the gesture images. In our method, we select the samples by their representative importance in relation to the class distribution, without taking into consideration the discriminative power in relation to the other classes. Another avenue for future research would be to extend our work to deal with nonlinear patterns since the current project is mainly based on linear transformations. This objective can be realized by adopting a nonlinear variant of PCA, such as kernel PCA.

5. REFERENCES

[1] Tae-Kyun Kim, Josef Kittler, and Roberto Cipolla, "Discriminative learning and recognition of image set classes using canonical correlations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 6, 2007.

[2] Tae-Kyun Kim, Shu-Fai Wong, and Roberto Cipolla, "Tensor canonical correlation analysis for action classification," in *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*. IEEE, 2007, pp. 1–8.

[3] Zhong Yang, Yi Li, Weidong Chen, and Yang Zheng, "Dynamic hand gesture recognition using hidden markov models," in *Computer Science & Education (ICCSE), 2012 7th International Conference on*. IEEE, 2012, pp. 360–365.

[4] Ying Yin and Randall Davis, "Real-time continuous gesture recognition for natural human-computer interaction," in *Visual Languages and Human-Centric Computing (VL/HCC), 2014 IEEE Symposium on*. IEEE, 2014, pp. 113–120.

[5] Hui-Shyong Yeo, Byung-Gook Lee, and Hyotaek Lim, "Hand tracking and gesture recognition system for human-computer interaction using low-cost hardware," *Multimedia Tools and Applications*, vol. 74, no. 8, pp. 2687–2715, 2015.

[6] Antonio Hernández-Vela, Miguel Ángel Bautista, Xavier Perez-Sala, Víctor Ponce-López, Sergio Escalera, Xavier Baró, Oriol Pujol, and Cecilio Angulo, "Probability-based dynamic time warping and bag-of-visual-and-depth-words for human gesture recognition in rgb-d," *Pattern Recognition Letters*, vol. 50, pp. 112–121, 2014.

[7] Ahmet Birdal and Reza Hassanpour, "Region based hand gesture recognition," 2008.

[8] Jakub Konecnỳ and Michal Hagara, "One-shot-learning gesture recognition using hog-hof," *Journal of Machine Learning Research*, vol. 15, pp. 2513–2532, 2014.

[9] Peter Sykora, Patrik Kamencay, and Robert Hudec, "Comparison of sift and surf methods for use on hand gesture recognition based on depth map," *AASRI Procedia*, vol. 9, pp. 19–24, 2014.

[10] Jun Wan, Guodong Guo, and Stan Z Li, "Explore efficient local features from rgb-d data for one-shot learning gesture recognition," *IEEE transactions on pattern analysis and machine intelligence*, vol. 38, no. 8, pp. 1626–1639, 2016.

[11] Ana I Maqueda, Carlos R del Blanco, Fernando Jaureguizar, and Narciso García, "Human–computer interaction based on visual hand-gesture recognition using volumetric spatiograms of local binary patterns," *Computer Vision and Image Understanding*, vol. 141, pp. 126–137, 2015.

[12] Manavender R Malgireddy, Ifeoma Inwogu, and Venu Govindaraju, "A temporal bayesian model for classifying, detecting and localizing activities in video sequences," in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2012 IEEE Computer Society Conference on*. IEEE, 2012, pp. 43–48.

[13] Paweł Pławiak, Tomasz Sośnicki, Michał Niedźwiecki, Zbisław Tabor, and Krzysztof Rzecki, "Hand body language gesture recognition based on signals from specialized glove and machine learning algorithms," *IEEE Transactions on Industrial Informatics*, vol. 12, no. 3, pp. 1104–1113, 2016.

[14] Frank Weichert, Daniel Bachmann, Bartholomäus Rudak, and Denis Fisseler, "Analysis of the accuracy and robustness of the leap motion controller," *Sensors*, vol. 13, no. 5, pp. 6380–6393, 2013.

[15] Gregory Shakhnarovich and Baback Moghaddam, "Face recognition in subspaces," in *Handbook of Face Recognition*, pp. 19–49. Springer, 2011.

[16] Jiwen Lu, Yap-Peng Tan, and Gang Wang, "Discriminative multimanifold analysis for face recognition from a single training sample per person," *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 1, pp. 39–51, 2013.

[17] Takumi Kobayashi, "Generalized mutual subspace based methods for image set classification," in *Asian Conference on Computer Vision*. Springer, 2012, pp. 578–592.

[18] Bernardo B Gatto, SS Waldir, M Eulanda, and Dos Santos, "Kernel two dimensional subspace for image set classification," in *Tools with Artificial Intelligence (ICTAI), 2016 IEEE 28th International Conference on*. IEEE, 2016, pp. 1004–1011.

[19] Bernardo B Gatto and Eulanda M dos Santos, "Image-set matching by two dimensional generalized mutual subspace method," in *Intelligent Systems (BRACIS), 2016 5th Brazilian Conference on*. IEEE, 2016, pp. 133–138.

[20] Kazuhiro Fukui and Atsuto Maki, "Difference subspace and its generalization for subspace-based methods," *IEEE transactions on pattern analysis and machine intelligence*, vol. 37, no. 11, pp. 2164–2177, 2015.

[21] Ken-ichi Maeda, "From the subspace methods to the mutual subspace method," in *Computer Vision*, pp. 135–156. Springer, 2010.

[22] Bernardo Bentes Gatto, Lincon Sales de Souza, and Eulanda M dos Santos, "A deep network model based on subspaces: A novel approach for image classification," in *Machine Vision Applications (MVA), 2017 Fifteenth IAPR International Conference on*. IEEE, 2017, pp. 436–439.

[23] Eduardo Sontag, "Nonlinear regulation: The piecewise linear approach," *IEEE Transactions on automatic control*, vol. 26, no. 2, pp. 346–358, 1981.

[24] Mats Viberg, "Subspace-based methods for the identification of linear time-invariant systems," *Automatica*, vol. 31, no. 12, pp. 1835–1851, 1995.

[25] Kazuhiro Hotta, "Local co-occurrence features in subspace obtained by kpca of local blob visual words for scene classification," *Pattern Recognition*, vol. 45, no. 10, pp. 3687–3694, 2012.

[26] Françoise Chatelin, *Eigenvalues of Matrices: Revised Edition*, vol. 71, SIAM, 2012.

[27] Tae-Kyun Kim and Roberto Cipolla, "Canonical correlation analysis of video volume tensors for action categorization and detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 8, pp. 1415–1428, 2009.