2024 IEEE International Conference on Systems, Man, and Cybernetics (SMC)
October 6-10, 2024. Sarawak, Malaysia

Modeling Human Strategy for Flattening Wrinkled Cloth Using Neural Networks

Nilay Kant[1], Ashrut Aryal[1], Rajiv Ranganathan[2], Ranjan Mukherjee[1], and Charles Owen[3]

***Abstract*— This paper explores a novel approach to model strategies for flattening wrinkled cloth learning from humans. A human participant study was conducted where the participants were presented with various wrinkle types and tasked with flattening the cloth using the fewest actions possible. A camera and Aruco marker were used to capture images of the cloth and finger movements, respectively. The human strategies for flattening the cloth were modeled using a supervised regression neural network, where the cloth images served as input and the human actions as output. Before training the neural network, a series of image processing techniques were applied, followed by Principal Component Analysis (PCA) to extract relevant features from each image and reduce the input dimensionality. This reduction decreased the model's complexity and computational cost. The actions predicted by the neural network closely matched the actual human actions on an independent data set, demonstrating the effectiveness of neural networks in modeling human actions for flattening wrinkled cloth.**

***Index Terms*— Aruco marker, flattening cloth, human-machine cooperation, human strategy, learning, neural network, wrinkle**

I. Introduction

Flattening deformed or wrinkled cloth is critical in several industries, such as textiles [9], [18], garments [11], [12], [23], and surgical robots [19]. Smooth, wrinkle-free fabric is essential for procedures like cutting, sewing, and packaging. Automating this task with robots presents significant challenges due to the complex and dynamic nature of fabric manipulation [5]. The inherent flexibility of fabrics and the wide variety of wrinkle types require high levels of precision and adaptability. Humans excel at this task, and this research aims to model human strategies for flattening wrinkled cloth to enhance automation capabilities in industries where fabric handling is essential.

Researchers have employed various approaches to address the challenge of cloth flattening using robots. For example, a heuristic-based approach in [20] identified clusters of wrinkles with k-means filtering on the range map, targeting the largest wrinkle and applying an appropriate force to eliminate it without creating new ones. Seita et al. [19] utilized deep imitation learning with a fabric simulator and an algorithmic supervisor, which provided paired observations and actions based on complete fabric state information. However, this approach often failed on already smooth fabrics, causing unnecessary pulls that introduced wrinkles. A dynamic method involving high-velocity actions namely, 'pick', 'stretch', and 'fling' using a dual-arm robot was proposed in [6], where a self-supervised learning framework that learns from visual observations was employed. A two-phase algorithm for automating the unfolding and flattening of laundry using interactive perception was presented in [22]. This paper presents an approach to flattening wrinkled cloth by learning from human behavior, a method that contrasts with existing techniques.

Understanding human behavior is crucial for developing effective human-robot interactions and enhancing the safety and efficiency of such systems [17], [24]. The need for robots to be highly aware of their surroundings and human counterparts increases with more frequent interactions [8]. For instance, robots can be trained with statistical models of human behavior to make decisions aligned with individual human styles [14]. Learning from human behavior has proven beneficial in various fields. In autonomous driving, Olier et al. [15] developed a method that enables vehicles to replicate human driving behaviors through deep learning and Bayesian filtering, enhancing autonomous monitoring. Kuperwajs et al. [13] used deep neural networks to improve cognitive models of human planning by analyzing gameplay patterns, refining decision-making models. Wang et al. [21] introduced an RGB-based architecture combining CNNs and LSTM units with a temporal-wise attention model to efficiently recognize human actions.

Motivated by the advantages of machines replicating human behaviors, we examine how humans flatten wrinkled cloth by collecting data on finger placement, pull direction, and pull length from human participants, relative to the cloth's wrinkled state. This data is used to train a regression neural network, which effectively models these human actions and can be applied to robotic automation for flattening wrinkled cloth, in sewing industry, for example.

II. Methods

A. Participants

Ten college-aged individuals in engineering, with no known motor impairments volunteered for the experiment. Participants provided written informed consent, and the experimental procedures were approved by the Institutional Review Board at Michigan State University.

[1]First, second and fourth authors are with the Department of Mechanical Engineering, Michigan State University, East Lansing, Michigan, USA.

[2]Third author is with the Department of Kinesiology, Michigan State University, East Lansing, Michigan, USA.

[3]Fifth author is with the Department of Computer Science and Engineering, Michigan State University, East Lansing, Michigan, USA. The corresponding author is Ranjan Mukherjee: mukherji@egr.msu.edu

This work was supported by the Michigan State University Strategic Partnership Grant and by the National Science Foundation, Grant No. CMMI-2326227.

2024 IEEE International Conference on Systems, Man, and Cybernetics (SMC) | 978-1-6654-1020-5/24/$31.00 ©2024 IEEE | DOI: 10.1109/SMC54092.2024.10832048

B. Task Description

The objective was to flatten a rectangular, uniformly-colored, wrinkled cloth placed on a flat table. To ensure that the cloth did not translate freely over the table, the midpoint of the top edge of the cloth was pinned to the table. Four distinct wrinkle patterns were employed in the study: vertical, horizontal, inclined, and mixed - see Fig.1. The point where the cloth was pinned to the table is indicated by a black dot. Each wrinkle type was presented five times to each participant, resulting in twenty trials per participant. Wrinkle patterns were generated before each trial by laying the cloth on a rod placed on the table and then removing the rod; this allowed us to generate the different types of wrinkles consistently across different trials and participants.

The objective of the study was to examine the strategy used by humans for flattening wrinkled cloth. Participants were encouraged to accomplish flattening of the cloth with the minimum number of iterations. When a wrinkle pattern was presented, participants were directed to use their index finger to pull the cloth from any point on an edge of the cloth along a straight line. Prior to each pull, participants were instructed to position their index fingers at a designated home position on the table - see Fig.2. The sequences of actions comprised of lifting the finger from the home position, translating it to the desired position on the edge of the cloth, executing the pull, and returning the finger to the home position, constituted one iteration. Participants were permitted to proceed at their own comfortable pace and to pause in the home position to strategize their next iteration. The iterative process was continued till the cloth was in a state of near flatness (as determined visually by the experimenter). It should be noted that each participant was given a practice trial in the beginning to get familiar with the process. Additionally, the order in which the wrinkle patterns were presented to the participants was randomized.

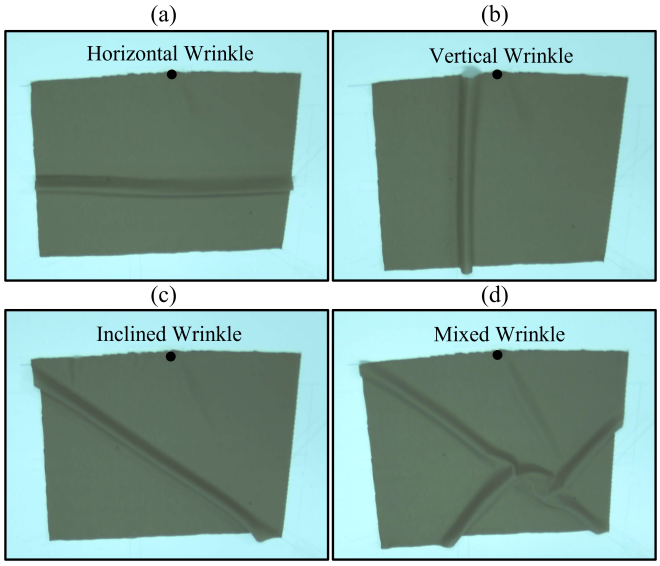

Fig. 1. Four wrinkle types used in experiments: (a) horizontal, (b) vertical, (c) inclined and (d) mixed. Black dot shows location where the cloth is pinned to the table.

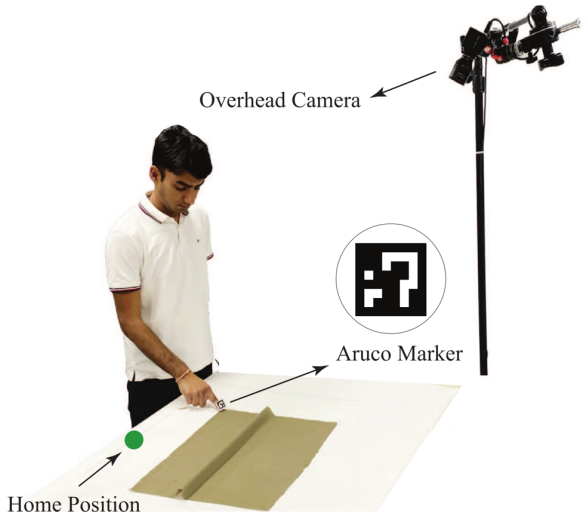

Fig. 2. Experimental setup for data collection

C. Data Acquisition

The experimental setup for the study is shown in Fig.2, and is comprised of an overhead camera and an Aruco marker on the participant's finger. We employed a 12 MP Arducam IMX477 camera system and an Arducam 2.8-12mm Varifocal C-Mount lens at 147cm distance from the center of the cloth. The camera was configured to capture images at a resolution of 3840×2160 pixels, resulting in a 8MP output. The lens was set to a $12mm$ focal length for a 37.7 degree horizontal field of view. Image acquisition was conducted using a local PC equipped with the OpenCV Python library. To track the motion of the index finger, Aruco markers [4] were used. Real-time motion tracking was achieved using the OpenCV Aruco library. The camera system was used to simultaneously sense both the state of the cloth and finger movements. The collected data for each trial included image frames along with the corresponding 3D-coordinates of the Aruco marker. This data was utilized to extract the image of the wrinkled cloth prior to each iteration and the attributes of the iteration, which include the Cartesian coordinates of the location where the finger is placed for the pull operation, and the direction and length of the pull performed by the participant.

III. Image Processing for Wrinkle Extraction

Our aim was to establish a connection between the state of the wrinkled cloth and the corresponding action (attributes of iteration) for its removal by a human participant. It was postulated that variables such as orientation, location, shape, and height of the wrinkles would impact the decision-making process involved in pulling the cloth. Utilizing an overhead camera setup, images of the state of the cloth were captured prior to each iteration. These images, however, contained superfluous information such as the region outside

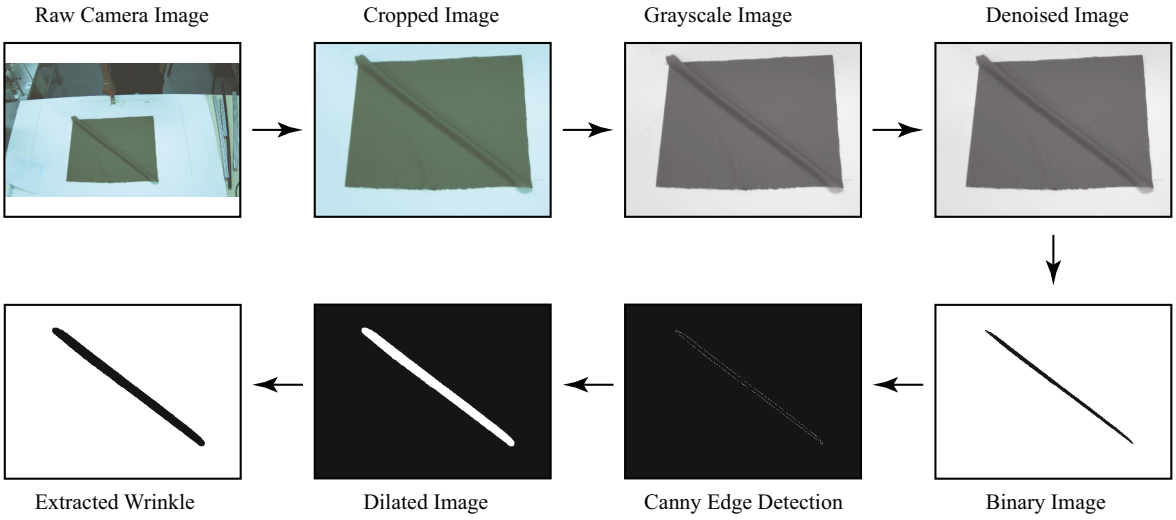

Fig. 3. Image processing steps utilized for extracting wrinkle features.

the perimeter of the cloth, which does not contain relevant information for decision-making. Additionally, we reasoned that while wrinkles play a significant role in decision making, finer details such as thread patterns are less influential. Therefore, our methodology focused on extracting solely the wrinkle features from the camera image frame captured prior to each iteration.

To extract wrinkles from a raw camera image, a series of image processing techniques were employed. The raw image captured by the camera, comprising of RGB channels, was first cropped to isolate the cloth portion while retaining background elements on the table and scaled to a size of 100 x 100 pixels. The image was then converted to monochrome. To reduce image noise, a non-local means-based filtering technique [2] was applied. Wrinkled cloth exhibits distinct edges when viewed from an overhead camera. To separate these edge-like regions from the cloth background, the image was binarized using Otsu's adaptive thresholding method [16]. Notably, the obtained threshold value was scaled down (60 %) for consistent background removal across all trials. Following the separation of wrinkle features from the background, Canny's edge detection algorithm [3] was utilized to highlight the wrinkle boundaries. Next, the morphological operation of dilation [7] with a disk structuring element was used to refine and emphasize the wrinkle feature. Finally, the dilated image was inverted to achieve a clean extraction of wrinkles against a white background. A schematic of the image processing[1] steps for an inclined wrinkle is presented in Fig. 3.

IV. Learning Human Strategy

A. Consistency in Human Strategy

The human participant data for the four wrinkle types is shown in Fig.4. The red dots denote the position of finger placement prior to a pull operation and the black arrows[2] depict the pull length and direction. It can be seen in Fig.5 that the human strategy is consistent in terms of position, direction, and length of pull for three of the four wrinkle types, namely, horizontal, vertical and inclined. The human strategy for the mixed wrinkle type - see Fig.4(d), was inconsistent. The initial pull locations were clustered in three

[1]All image processing operations were performed using the computer vision toolbox in Matlab.

[2]All the arrows were concentrated in a small region and therefore a few of them are shown to avoid over-crowding

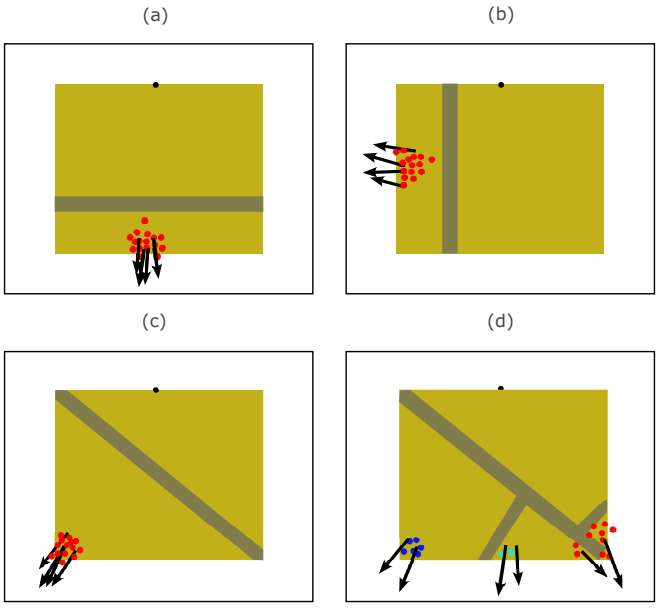

Fig. 4. Schematic illustrating human strategy for wrinkle removal from (a) horizontal, (b) vertical, (c) inclined, and (d) mixed wrinkle types. Red dots denote the location of finger placement in the first three sub-plots where there is only one dominant region of finger placement. In the fourth sub-plot, different colors represent each of the three major regions of finger placement. Black arrows depict the length and direction of pull operation.

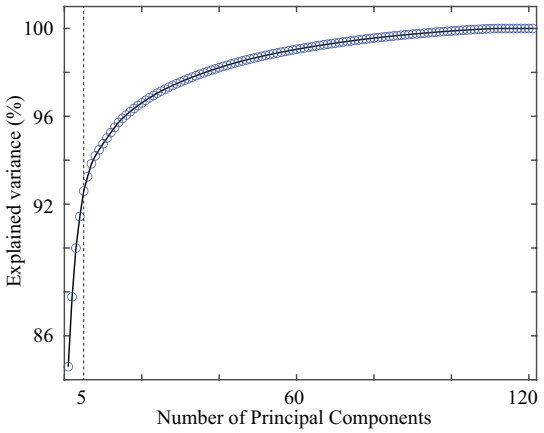

Fig. 5. Plot of explained variance of the input image data.

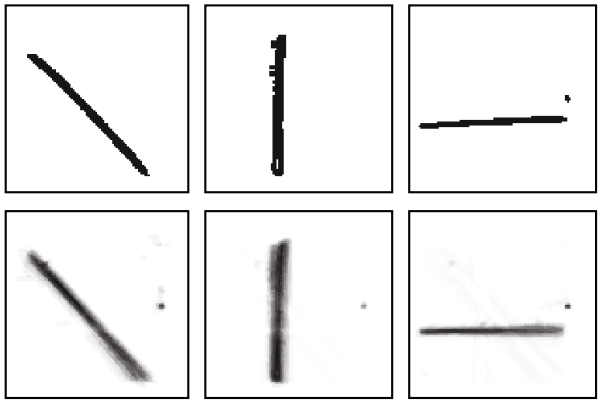

Fig. 6. Comparison of original and reconstructed images. Top row shows inclined, vertical, and horizontal wrinkle. Bottom row shows the corresponding images reconstructed with only five principal components.

different regions at the bottom edge of the cloth - the two corners and the midpoint. These strategies were also equally effective in flattening the cloth, as participants, on average, flattened the cloth in two iterations. This observation suggests the existence of multiple optimal solutions for flattening a mixed wrinkle. Due to the absence of a consistent human strategy, we excluded the mixed wrinkle category from our current analysis.

B. Image Dimensionality Reduction

Our objective is to learn the human strategy for manipulating wrinkled cloth based on its visual representation. To this end, we utilize a regression neural network framework for supervised learning. It is important to note that we assume human actions depend solely on the current cloth state and are not influenced by past cloth states or actions. The images processed in Section III are of dimensions of 100×100 pixels. Our dataset contains 112 processed images containing three types of wrinkle patterns: vertical, horizontal, and inclined. We also included 10 images of completely flat cloth, which would require no pulling action. This resulted in an input dataset of dimension 122×10000, where each row of the array represents an image. The corresponding human action to each of these 122 trials is characterized by four parameters: the coordinates (x and y) of finger placement for cloth manipulation, the length of the pull (d), and the direction of the pull (θ). The human actions linked with flat cloth was set as a vector with four zero entries. This resulted in an output dataset of dimension 122×4.

In theory, a regression neural network could be directly trained on this data. However, our training dataset is constrained due to the limited number of participants. Given the high dimensionality of the input data, relying solely on the neural network to extract relevant features would significantly increase the complexity of the network, thereby expanding the number of parameters to be learned. This may not be feasible given the limitations imposed by the size of the dataset. To determine the number of relevant features in our input dataset, we employ principal component analysis (PCA) [1]. The cumulative explained variance plot, shown in Fig.5, indicates that five principal components contribute to over 92% variation within the images. Therefore, we opted for 5 principal components and projected our original dataset, which had a dimension of 10000 (pixels), to a significantly reduced dimension of 5. To visualize the effectiveness of this dimensionality reduction, plots of three sample wrinkle types are presented in the first row of Fig.6; the second row displays the same images reconstructed using only the five principal components. It can be observed that the five principal components distinctly retain the wrinkle attributes.

The input dataset was reduced from 122×10000 to 122×5 using PCA prior to training. To address variations in scale and units among output variables, z-score standardization was applied. Before training, a randomization procedure was implemented on both input and output data to shuffle samples, thereby mitigating potential biases inherent in the dataset. Subsequently, the dataset was partitioned into training and validation sets at a ratio of 75% to 25%.

The neural network architecture, shown in Fig.7 comprised of four fully connected hidden layers. The initial layer featured 10 neurons, followed by a layer with 15, 24, and 10 neurons, respectively. The sigmoid activation function was incorporated after each fully connected layer to capture nonlinear relationships between input and output data. The output layer, consisting of 4 neurons, was also fully connected, followed by a regression layer to facilitate continuous valued predictions, aligning with the output dimensionality.

C. Neural Network Description and Training

Three optimization algorithms namely, stochastic gradient descent, RMSprop, and Adam were evaluated in training the neural network. Among these, the Adam optimizer [10] exhibited superior training performance. This can be attributed to the observation that the gradient exhibited fluctuations during the training process. Consequently, the Adam optimizer, known for its robustness to noisy gradients, was selected for learning the neural network weights. The training data was shuffled every epoch. The gradient decay

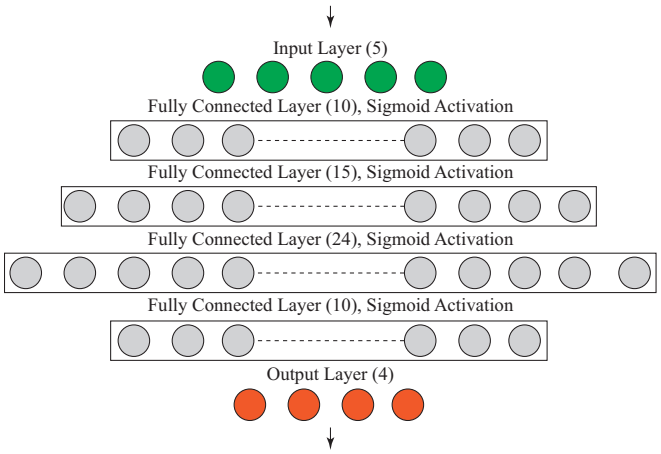

Fig. 7. Neural network architecture.

factor was set to 0.9, and the squared gradient decay factor to 0.999, ensuring smooth convergence and stability during training. The initial learning rate was set to 0.01, with no predefined learning rate schedule, but a drop factor of 0.1 is applied every 10 epochs to adaptively adjust the learning rate. L_2 regularization with a weighting of 1.0×10^{-4} was employed to reduce over fitting. Training was carried out for a maximum of 100 epochs, utilizing mini-batches of size 12. Initially, the root mean squared error was approximately 2.8. Around the 90th epoch, the average root mean squared error stabilized within a close range of 0.8.

V. Results

After training the neural network, its performance was assessed using the validation dataset. The actual and predicted plots of the four human actions are displayed in Fig.8. Figs.8 (a) and (b) show the actual and predicted x and y coordinates of finger placement before each pull, while Figs.8 (c) and (d) show the length of pull and direction of pull for both actual data and those predicted by the neural network. The root mean square (RMS) errors of the neural network prediction for x and y coordinates, pull length (d), and pull angle (θ) are provided in Table I.

For better visualization, a schematic representing the human actions is presented in Fig. 9. The yellow box represent the cloth when it is perfectly flat. The green arrow lines depict the actual human actions, while blue arrow lines signify actions predicted by the neural network. The starting points of the arrow lines indicate the initial finger placement before pulling while its length denote the pull length. Three clusters of arrow lines are observable: the upper left cluster correspond to vertical wrinkles, the lower corner cluster correspond to inclined wrinkles, and the bottom-right cluster

TABLE I
RMS Prediction Error of Human Actions

x (m)	y (m)	d (m)	θ (rad)
0.013	0.017	0.008	0.3

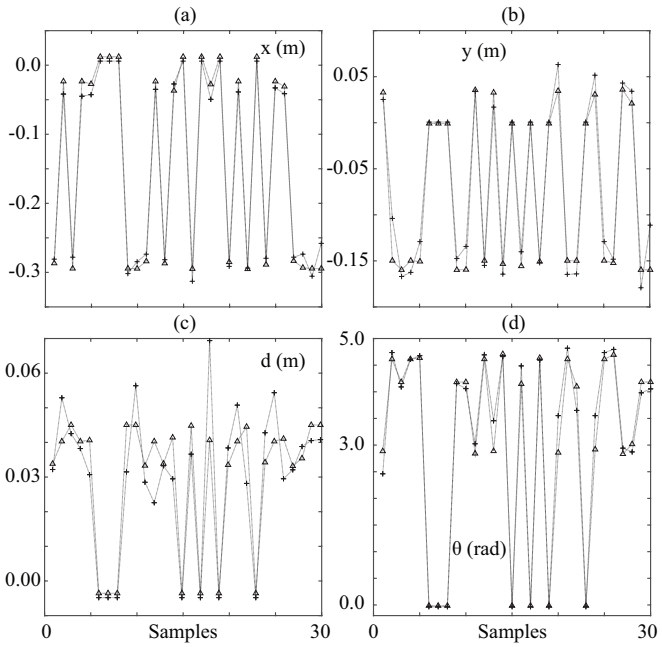

Fig. 8. Plot showing neural network performance in learning human strategy. The symbols + and △ denote the actual and predicted human actions, respectively.

correspond to horizontal wrinkles. It can be observed that the pull direction is approximately perpendicular to the orientation of the wrinkles. Also, for cloth in a flat state, the human actions are clustered near the origin with the length of the arrows almost equal to zero. This is shown by a blue dot.

The actual and predicted human actions are closely clustered together for each of the three wrinkle types. Note that some arrows in Fig.9 may appear either within or outside the cloth even though participants pulled the cloth from its edges; however this discrepancy in the visualization is due to the fact that the edges of the cloth shift when wrinkles are formed and this results in a visible cloth area that is smaller than depicted. Additionally, since the cloth was fixed at one point, it occasionally underwent slight rotation

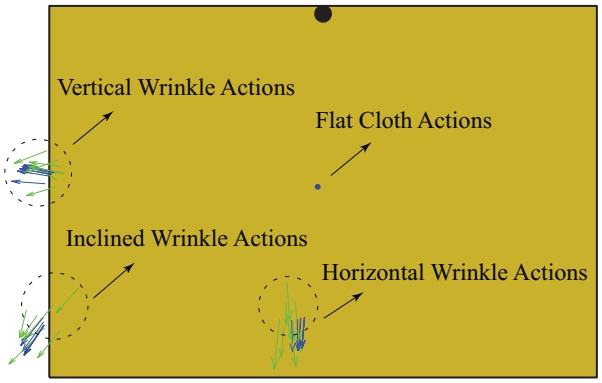

Fig. 9. Real and predicted human actions in relation to flattening wrinkled cloth. The green and blue arrows represent the actual and predicted human actions, respectively.

when wrinkles were formed prior to data collection, rather than remaining perfectly horizontal. These variations in cloth state, along with corresponding variations in human action, are inherently included in the neural network modeling. As evident from Figs. 8 and 9, the neural network closely models and predicts human strategies for flattening wrinkled cloth.

VI. Discussion

Flattening wrinkled cloth is commonly done by humans, yet the decision-making process underlying this task remains unclear. To address this, we conducted a human participant study where participants were tasked with flattening preformed wrinkled cloth, considering four distinct wrinkle patterns: horizontal, vertical, inclined, and mixed. Participants were instructed to pull the cloth from its edges along a straight line, while their finger motion trajectories and cloth images were recorded using a camera. Four variables, namely, Cartesian coordinates of finger placement before pulling, pull length, and pull direction, constituted human action variables. Our objective was to map input wrinkled cloth images to these action variables using a supervised regression neural network model. Image processing and PCA were utilized for feature extraction due to the limited number of human participants. The network was then trained and validated using human participant data, demonstrating promising results in predicting human actions.

A limitation of this study is that it modeled human actions for only a single wrinkle on cloth due to the challenges and costs of acquiring data from human trials. To achieve good prediction performance, we kept the network complexity low, as increasing it would require more data. Additionally, for mixed wrinkle patterns, we observed three variations in human strategy, each removing wrinkles with the same number of iterations, leading us to exclude this type from our analysis. This suggests that completely learning the human approach to any wrinkle type may be nondeterministic. However, even simple wrinkle patterns provide valuable insights into human cloth-flattening approaches, which can inform the design of generalized, human-inspired algorithms. This is the focus of our current research. Future work will aim to use the network to assist a robotic arm in real-time cloth flattening during sewing.

References

[1] Hervé Abdi and Lynne J Williams. Principal component analysis. *Wiley interdisciplinary reviews: computational statistics*, 2(4):433–459, 2010.

[2] Antoni Buades, Bartomeu Coll, and J-M Morel. A non-local algorithm for image denoising. In *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, volume 2, pages 60–65. Ieee, 2005.

[3] John Canny. A computational approach to edge detection. *IEEE Transactions on pattern analysis and machine intelligence*, (6):679–698, 1986.

[4] Sergio Garrido-Jurado, Rafael Muñoz-Salinas, Francisco José Madrid-Cuevas, and Manuel Jesús Marín-Jiménez. Automatic generation and detection of highly reliable fiducial markers under occlusion. *Pattern Recognition*, 47(6):2280–2292, 2014.

[5] David Gershon. Strategies for robotic handling of flexible sheet material. *Mechatronics*, 3(5):611–623, 1993.

[6] Huy Ha and Shuran Song. Flingbot: The unreasonable effectiveness of dynamic manipulation for cloth unfolding. In *Conference on Robot Learning*, pages 24–33. PMLR, 2022.

[7] Robert M Haralick and Linda G Shapiro. *Computer and robot vision*, volume 1. Addison-wesley Reading, MA, 1992.

[8] Roohollah Jahanmahin, Sara Masoud, Jeremy Rickli, and Ana Djuric. Human-robot interactions in manufacturing: A survey of human behavior modeling. *Robotics and Computer-Integrated Manufacturing*, 78:102404, 2022.

[9] Petros I Kaltsas, Panagiotis N Koustoumpardis, and Pantelis G Nikolakopoulos. A review of sensors used on fabric-handling robots. *Machines*, 10(2):101, 2022.

[10] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[11] P Koustoumpardis and N Aspragathos. A review of gripping devices for fabric handling. *hand*, 19(July 2004):20, 2004.

[12] Panagiotis Koustoumpardis, Nikos Aspragathos, and Paraskevi Zacharia. *Intelligent robotic handling of fabrics towards sewing*. Citeseer, 2006.

[13] Ionatan Kuperwajs, Heiko H Schütt, and Wei Ji Ma. Using deep neural networks as a guide for modeling human planning. *Scientific reports*, 13(1):20269, 2023.

[14] Stefanos Nikolaidis, Ramya Ramakrishnan, Keren Gu, and Julie Shah. Efficient model learning from joint-action demonstrations for human-robot collaborative tasks. In *Proceedings of the tenth annual ACM/IEEE international conference on human-robot interaction*, pages 189–196, 2015.

[15] Juan Sebastian Olier, Pablo Marín-Plaza, David Martín, Lucio Marcenaro, Emilia Barakova, Matthias Rauterberg, and Carlo Regazzoni. Dynamic representations for autonomous driving. In *2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pages 1–6. IEEE, 2017.

[16] Nobuyuki Otsu et al. A threshold selection method from gray-level histograms. *Automatica*, 11(285-296):23–27, 1975.

[17] Loizos Psarakis, Dimitris Nathanael, and Nicolas Marmaras. Fostering short-term human anticipatory behavior in human-robot collaboration. *International Journal of Industrial Ergonomics*, 87:103241, 2022.

[18] Johannes Schrimpf and Lars Erik Wetterwald. Experiments towards automated sewing with a multi-robot system. In *2012 IEEE International Conference on Robotics and Automation*, pages 5258–5263. IEEE, 2012.

[19] Daniel Seita, Aditya Ganapathi, Ryan Hoque, Minho Hwang, Edward Cen, Ajay Kumar Tanwani, Ashwin Balakrishna, Brijen Thananjeyan, Jeffrey Ichnowski, Nawid Jamali, et al. Deep imitation learning of sequential fabric smoothing from an algorithmic supervisor. In *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 9651–9658. IEEE, 2020.

[20] Li Sun, Gerarado Aragon-Camarasa, Paul Cockshott, Simon Rogers, and J Paul Siebert. A heuristic-based approach for flattening wrinkled clothes. In *Towards Autonomous Robotic Systems: 14th Annual Conference, TAROS 2013, Oxford, UK, August 28–30, 2013, Revised Selected Papers 14*, pages 148–160. Springer, 2014.

[21] Lei Wang, Yangyang Xu, Jun Cheng, Haiying Xia, Jianqin Yin, and Jiaji Wu. Human action recognition by learning spatio-temporal features with deep neural networks. *IEEE access*, 6:17913–17922, 2018.

[22] Bryan Willimon, Stan Birchfield, and Ian Walker. Model for unfolding laundry using interactive perception. In *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 4871–4876. IEEE, 2011.

[23] Ryder C Winck, Steve Dickerson, Wayne J Book, and James D Huggins. A novel approach to fabric control for automated sewing. In *2009 IEEE/ASME International Conference on Advanced Intelligent Mechatronics*, pages 53–58. IEEE, 2009.

[24] Jie Yang, Yangsheng Xu, and Chiou S Chen. Human action learning via hidden markov model. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 27(1):34–44, 1997.