ONLINE WHOLE-WORD AND STROKE-BASED MODELING FOR HAND-WRITTEN LETTER RECOGNITION IN IN-CAR ENVIRONMENTS

You-Chi Cheng[⋆] *Kehuang Li*[⋆] *Zhe Feng*[†] *Fulinag Weng*[†] *Chin-Hui Lee*[⋆]

[⋆] Georgia Institute of Technology, Atlanta, GA 30332-0250, USA
[†] Robert BOSCH Research and Technology Center, Palo Alto, CA 94304, USA

ABSTRACT

A finger-written, camera-based, hand gesture recognition framework for English letters in an in-vehicle environment based on Hidden Markov models is proposed. Due to the nature of the constrained hand-movement situations on the steering column, we are confronted with at least two challenging research issues, namely varying illumination conditions and noisy hand gestures. The first difficulty is alleviated by utilizing the contrast for background-foreground separation and skin model adaptation. We also adopt sub-letter stroke modeling to reduce the noisy frames of the beginning and ending parts of the letter gestures followed by the trajectory re-normalization . Moreover, the geometric relationship between letter pairs is also utilized to distinguish highly confusable letters. Finally, score fusion between whole-letter and sub-stroke models can be used to further improve the performance. When compared with the baseline system with simple features, our experimental results show that an overall relative error reduction of 66.03% can be achieved by integrating the above four new pieces of information.

Index Terms— Hand gesture, letter recognition, skin color detection, stroke modeling, Hidden Markov models

1. INTRODUCTION

User interfaces for in-vehicle instrument control require not only good user experience but also minimal distraction due to safety reasons [1]. While most systems were done by control knobs or touch screens with proper displaying systems [2, 3] for user interactions, multiple modalities, such as speech and gesture recognition, also draws recent attentions [4, 5]. In such hands-busy and eyes-busy input conditions, gesture input plays a key role to speech input due to their complimentary nature. In this work, hand gestures of finger-written English letters for in-car environments are studied.

The first challenge for in-car hand gesture recognition is due to varying illumination conditions. Most generic gesture recognition systems are built for in-door environments [6, 7, 8, 9, 10]. Different color space representations [7], skin regions versus non-skin regions classification rules [8], automatic white balance calibration [6], or depth information [9] are often used to deal with the cluttered environmental conditions. Due to strong diversities for in-car environments, even with these solutions, the detection can still be quite challenging. Therefore, other hardware solutions like active lamp [11] is often used.

The second challenge for in-car hand gesture recognition is how to model the hand gesture. Most generic gesture recognition systems use sequential modeling techniques such as Hidden Markov Model (HMM)[8, 12] or Finite State Machine (FSM)[13]. The main challenge for in-car environment is its gesturing diversities. Since a user may put more attention on driving instead of looking at his or her own gesturing hands. These gestures are expected to be more noisy and may have meaningless gesture components such as making a long starting stroke before actually beginning to perform the real gesture. Directly normalizing the whole gesture like [14] may cause the meaningless strokes to be misunderstood as part of a gesture.

In this work, we developed solutions for both challenges. For the first challenge, in order to deal with the diversities encountered under in-car environment, we use our previously proposed adaptive skin detectors [15]. The main idea for this detector is to use both the corresponding background and foreground information along with Maximum a Posteriori (MAP) [16] adapted skin models to enhance the detector performance and robustness.

For the second challenge on modeling, besides the whole letter modeling, we utilize the stroke modeling. The idea of the stroke modeling is similar to the stroke modeling for the hand writing Chinese character recognition [17, 18, 14] or the phoneme modeling for speech recognition [19]. The stroke modeling may be able to take care of some gestures with different stroke orders. Furthermore, the stroke modeling enables the possibility to increase the discriminative power by explicitly modeling the intra-stroke correlation with geometric properties as a post processing stage, and it also has the potential to remove noisy strokes. Experimental results showed these advantages can be verified and these two practical challenges are effectively addressed.

2. RELATION TO PRIOR WORK

Most of the previous gesture recognition systems addressed issues of the object detection and the gesture recognition itself as two separate problems. That is, the input gesture frames must go through some feature extraction modules to produce some feature vectors for gesture models. The importance of the first challenge mentioned in Section 1 is that a bad object detector can often result in a noisy sets of feature vectors.

To design a robust hand or body part detector for human gesture recognition, there are two aspects of concerns. The first is mainly on the signal acquisition side. During the collection of the data set [6], automatic white balance calibration is first used to adjust the color distribution, but a postprocessing stage is still required. In an in-car gesture recognition designed in [11], LED light is used to minimize the effect of different illumination conditions. Infrared camera can also be used to obtain the depth information [9] so that the foreground object can be reliably extracted.

Thanks to Robert BOSCH RTC for funding.

Besides hardware solutions for cluttered environments, other issues are related to the skin color detection. For example, in [11, 8], more specific thresholding rules are applied. Other side information, such as motion and edges, is also proved to be useful for better detection results [8]. The effect of different color space representation and skin color models are discussed in [20] for hand detection, and it turns out that different color spaces are recommended in different application domains and models. Color tracking using an Expectation-Maximization (EM) type algorithm is shown to be useful [21] while adaptive techniques based on refining the model with false detection samples are also proposed in [22] to enhance the skin detection performance.

However, none of these methods directly use the fact that the detection target is from a sequence of frames. In our previous work [15], we explicitly included the background and foreground correlation into our skin detection modeling. Besides that, we used the Maximum a Posteriori (MAP) adaptation [16] to adjust the model conditions so that the mismatch between the training recording environment and that of the testing data can be reduced. With this algorithm, no hardware calibration or a carefully designed thresholding rule is required.

In addition to the skin detection, modeling of the extracted features is another important issue. The most commonly used tools are based on sequential modeling, such as HMMs [8, 12] or FSMs [13], which is more flexible than HMM. In these previous studies, a whole gesture is modeled by a single HMM or FSM, and the model that gives the highest score will be chosen as the recognized gesture. Typical input features for these application may involve position, velocity, and orientation [10]. However, for the recognition inside a car, a gesture can be noisy because the user didn't actually look at the gesture during writing. Furthermore, for gestures, such as English letters, one gesture may have several different ways of writing. Thus it needs more effort to collect the data for several different stroke orders for each letter. This motivates our study of stroke modeling.

Strokes are often used in the community of Chinese or other Asian written character recognition [17, 18, 14] in several different ways. This is due to its infeasibility to individually collect and model thousands of commonly used characters. Therefore, a lexicon is defined to regulate the way to compose a character with strokes similar to [23]. The same stroke from different characters can be shared to train a better stroke model. For the in-car gesture recognition, however, the main difference is that the extraneous movements between meaningful strokes will also be recorded and they may not be easily distinguished.

To deal with the difference between in-car stroke modeling and written text stroke modeling, we treated the intermediate strokes as meaningful strokes while building the lexicon. With this framework, the meaningless beginning and ending strokes of the gesture can be handled and re-normalization of meaningful strokes can be done accordingly. In addition, we may also incorporate the stroke relationship modeling by some geometric properties. Finally, a score fusion with the original whole-letter model is also applicable.

In summary, advantages of the proposed framework over prior arts are as follows:

1. Use the background and foreground information and adaptation to produce good hand detection;
2. The strokes can be trained with more effective training data;
3. It can potentially increase the discriminative power:
   (a) By removing meaningless strokes;
   (b) By explicit model the relationship among stroke;
4. Give flexibility to future data with different stroke order.

In the following we describe our hand gesture recognition system for the English letter recognition in detail, as shown in Fig. 1. After gesture frames are acquired by cameras, they are passed to the hand detection module, and its resulting region information is then fed to the trajectory understanding module to perform the actual gesture recognition. Our proposed techniques to address the two main difficulties for the in-car gesture recognition will also be presented.

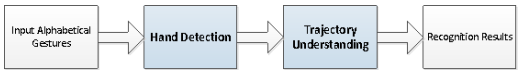


Fig. 1. Conceptual block diagram of the hand gesture recognition

3. HAND DETECTION ALGORITHM

3.1. In-Car Environmental Issues

As mentioned in Section 2, we applied the color information as a main clue for the hand region recognition. Unlike previous studies, we mainly focused on the in-car gesture recognition, concerning the various illumination conditions. Due to different illumination, reflection, and saturation, it is clear only using the color information may not be able to easily distinguish hands from background regions.

3.2. Proposed Solution for In-Car Hand Detection

We use our previously proposed algorithm [15] to deal with the issues encountered under in-car environments. As mentioned in Section 2, the goal is to integrate the background-foreground information to assist the MAP adapted skin color models. The algorithm is composed of the following steps:

1. Assume a purely background frame is always given, for the whole sequence, check the total variance σ by averaging the squared difference between every pixel of each frame and the background frame, if σ exceed some threshold, the frame is claimed to be a candidate frame with hand object.
2. For each local patch with selected block size (4×4, for example), compute magnitude of the block correlation coefficients $|\rho|$ with the corresponding background block.
3. Doing the MAP adaptation on the original color skin model.
4. Compute the MAP adapted log likelihood score *LL* of the current block's average RGB value. And scale the likelihood score *LL* to $NLL = 1/\{1+exp[-0.5\times(LL)+0]\}$.
5. Combine $|\rho|$ and *NLL* into a two dimensional vector, use a fusion classifier trained by these feature on the training data to make the final hand detection.

The goal of the third step is to reduce the model mismatch between the training and testing data. And the fifth step is used to combine both the background-foreground relationship with the color information, so that robustness across different illumination conditions can be achieved.

4. GESTURE MODELING

4.1. In-Car Gesture Modeling Issues

Although most traditional gesture recognition systems use HMMs as their primary modeling tool [12, 8], there are still some issues not

considered, just as mentioned in Section 2. Beside detection issues, users performing in-car gesture will not intend to perform gestures in a fixed region. That is, given a surface for doing gesture, like area near steering wheel, users may perform the gesture at any part of that surface, while meaningless gesture components in the beginning or ending of that gesture will also be captured.

In addition, previous techniques using HMM type tools should follow the Markov assumption, that is, only the relation between two immediate adjacent states or strokes will be explicitly modeled. Other stroke relationship, though not directly related to the stroke transition, may still have some discriminative power for the gesture recognition system. In this work, we will also address this issue.

In the following sub-sections, we will discuss several algorithms for in-car gesture modeling.

4.2. In-Car Gesture Modeling Algorithms

4.2.1. Whole Letter HMM

The most intuitive baseline system used the normalized position, velocity, and acceleration information as input vector sets o for HMMs with continuous state observation probability densities. The log likelihood score for the $i-th$ English letter can be computed as:

$$log[P(o|\Lambda_i^{whole\ word})]. \tag{1}$$

4.2.2. Stroke HMM

A lexicon is first defined for each English letter gesture. For example, letter P is defined as $\{START, \downarrow, \uparrow, \supset, END\}$. Multiple stroke sequences are also possible, but we only used one stroke sequence per letter in this study. With each stroke modeled by an HMM with the same set of features but with a smaller number of states, the log likelihood score for the ith letter can be computed as:

$$log[P(o|\Lambda_i^{strokes})] = \sum_{k \in D_i} log[P(o|\Lambda_k^{stroke} \cap \Lambda_i)], \tag{2}$$

where D_i is the stroke set allowed for the $i-th$ letter defined in the lexicon. An example segmentation result is shown in Fig. 2.

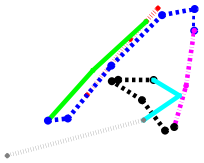
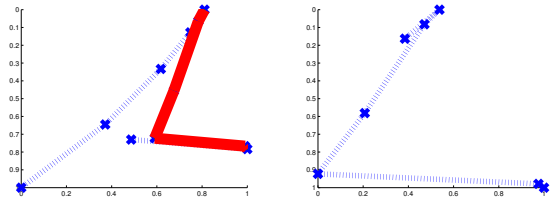


Fig. 2. An example of segmenting a letter with stroke HMM, lines with different colors are strokes decoded.

4.2.3. Stroke HMM with Two-Pass Decoding for Re-Normalization

As mentioned before, in-car gestures may include irrelevant starting and ending strokes, as shown in the left sub figure of Fig. 3. These strokes need to be removed and the meaningful gesture parts need to be re-scaled afterward because these varying length strokes can bias the gesture model. In order to cope with this issue, we use HMMs in a two-pass manner. We first use the HMMs with *START* and *END* models for decoding, and then remove these strokes and re-normalize the remaining position vectors, and pass the re-normalized vectors to a new set of HMMs without *START* and *END*. As shown in the right part of Fig. 3, this strategy works well.



(a) Original Trajectory (b) Re-Normalized Trajectory

Fig. 3. An example of applying two-pass decoding on the trajectory. The starting and ending part of the trajectory are effectively removed. Note in (a), the red solid lines are the meaningful strokes

4.2.4. Two-Pass Stroke HMM with Geometric Information

All previous configurations did not consider the explicit relation among strokes. By considering relationship among any two strokes, with the assumption of uniform stroke distribution, and considering the fact of probability scale differences, the modified score function for $i-th$ letter can be written as:

$$g(o|\Lambda_i) = \sum_{k \in D_i} log[P(o|\Lambda_k^{stroke} \cap \Lambda_i)] + \gamma \sum_{\substack{k,h \in D_i \\ k \neq h}} log[P(o|\Lambda_{k,h}^{geometric} \cap \Lambda_i)], \tag{3}$$

where γ is a positive constant with a typical value in the range of $[0,1]$. NOW we summarize the final two-pass algorithm as follows:

Algorithm 1: Two-pass letter recognition

input : sequence of position vectors o
letter models with starting/ending strokes Λ
letter models without starting/ending strokes Ξ
output: recognized English letter
for $\Xi_i \in \Xi$ **do**
  compute $log[P(o|\Xi_i)]$
end
$stroke\_sets \leftarrow \underset{\xi \in \Xi}{argmax}\{log[P(o|\xi)]\}$
// $\mathcal{N}$: scale normalization function
$o_- \leftarrow \mathcal{N}\{o - starting\ \&\ ending\ strokes\}$
for $\Lambda_i \in \Lambda$ **do**
  compute $log[P(o_-|\Lambda_i)]$
end
return $\underset{\lambda \in \Lambda}{argmax}\{log[P(o_-|\lambda)]\}$

In practical application, instead of picking up the maximum score over all candidates, we only choose two candidates with the first and second highest stroke scores. According to our observation, these candidates will give relatively big counts in the confusion matrix, so they are the confusion pairs to be considered. Furthermore, we also observed if geometric properties are not strong enough, fusing the whole letter score can be helpful, as will be discussed soon.

Table 1. Summary of Different Modeling Strategies

Configuration	Error Rate(%)
Whole-letter HMM (naïve skin model, baseline)	24.62
Whole letter HMM (MAP-adapted skin model)	12.57
Stroke HMM	12.73
Stroke HMM (+Re-normalization)	9.91
Stroke HMM (+Geometric)	9.17
Stroke HMM (+Whole letter fusion)	8.36
Stroke HMM (2-best)	5.54

5. EXPERIMENTAL RESULTS

5.1. Experimental Setup

We used a PointGray Dragonfly 2 camera with YUV 422 color and 1024×768 setup at 15 frames per second and 10 ms shutter speed. The acquired frames were first resized to a 256×192 to save storage spaces. We collected data from 17 different users and spread the sampling time over morning, noon, and evening sessions for all. A set of about 5,000 gesture sequences was recorded. To verify the proposed framework, the following experiments were conducted:

1. Test the performance of adaptive hand detector (Section 3.2) over naïve one.
2. Compare the stroke HMM (Section 4.2.2) and the whole-letter HMM (Section 4.2.1).
3. Check the impact of the stroke normalization (Section 4.2.3) and adding simple geometric information (Section 4.2.4).

Within this set 4,211 sequences, with the most commonly seen stroke order used for each of 26 letter gestures, were picked for evaluation. Half of the data for each letter and recording condition were used for training and testing. After trials and errors on the baseline system, 16-state single-mixture and 2-state 32-mixture HMMs were used for whole-letter and stroke modeling, respectively. To make a fair comparison among different configurations, we only evaluated the systems on a subsets of 1,877 gesture sequences with more than 16 frames of hand motion. Input to all models were sets of vectors composed of position with both axes normalized into the range of $[0,1]$ and corresponding velocity and acceleration.

5.2. Experimental Results

5.2.1. Whole-letter Modeling

With the recommended parameter setting and the baseline system mentioned in Section 4.2.1, we characterized the skin and non-skin models by Gaussian mixture models (GMMs) [24] with mixture counts determined by the technique in [25]. Then the proposed adaptive algorithm based on our previous work can significantly reduce the recognition error rate by 48.94% relatively, as shown in the first two rows of Table 1.

5.2.2. Sub-letter Stroke Modeling

Again, the recommended parameter setting was used, and testing was done on the same subset of data used in testing whole-letter models. We can see that the error rate for stroke modeling was 12.73%, which was comparable with 12.57%, produced by the whole-letter modeling, as shown in the 2^{nd} and 3^{rd} rows of Table 1.

One reason for this slight degradation in performance could come from the fact that the current stroke modeling was based on the same feature set as those used in whole-letter modeling. For example, both arcs in letter B were used to model the arc stroke that appear in B, P, R without considering their relative positions to the whole letter, this would make the resulting model not as stable. To enhance the stroke modeling, we adopted position re-normalization and added simple geometric information next.

5.2.3. Position Normalization and Letter Geometry

With gesture modeled by strokes defined in a lexicon, as shown in the 3^{rd} and 4^{th} rows of Table 1, applying the two-pass re-normalization scheme as proposed in Section 4.2.3 we reduced the error rate to 9.91%, a 22.15% relative error rate reduction over the original position-independent stroke models.

In addition, we applied the following geometric indices and used Equation 3 and Algorithm 1 for the letter recognition:

- Distances between endpoints from either stroke.
- Distances of stroke endpoints to the line connected by endpoints of the other stroke.
- Distance of the middle point of a stroke to the line connected by endpoints of the other stroke.
- Cosine value of two lines formed by endpoints of either strokes.

We only considered the letter candidates with the first and second highest stroke model scores. We can see that this setup with $\gamma = 0.08$ can further improve the performance by 7.47% and 27.97% relatively over the re-normalized and original stroke models, respectively, as shown in the 3^{rd}, 4^{th}, and 5^{th} rows of Table 1. Moreover, by adding corresponding scores from the original whole letter models with weight $(1, 0.08, 0.5)$, we can further reduce the error rate from stroke models with geometric properties to the best performance of a 8.36% error rate, which represents a 66.03% relative error reduction from the baseline whole-letter system.

Furthermore, if we considered the setup in Section 4.2.3 but allowed any hit from candidates with the 2 highest scores as a correct recognition, i.e., 2-best, we can have a much lower error rate of 5.54%. This makes application design easy when additional information such as a dictionary can be integrated into the figure-written letter recognition systems and also implies a potential improvement by using more discriminative geometric properties.

6. CONCLUSION AND FUTURE WORK

In this work, we deal with challenges encountered under in-car environments for the English letter hand gesture recognition. We alleviate the first challenge of varying lighting condition by integrating the background-foreground information and the adaptive skin color modeling. For the challenge of proper gesture modeling, we investigating the stroke modeling and designed a two-pass decoding algorithm and showed its effectiveness to remove the meaningless starting and ending strokes that constantly seen for in-car hand gestures. We then integrating the geometric properties to further reduce the error rate. Finally, by further fusing this model with original baseline score, a relative 66.03% error reduction compared to original HMM baseline is achieved. Moreover, the performance gap still exists when compared to the 2-best scenario. Therefore, geometric properties among strokes with better discriminative power and other stroke properties should be further studied, especially for easily confused letter pairs.

7. REFERENCES

[1] L. Garay-Vega, AK Pradhan, G. Weinberg, B. Schmidt-Nielsen, B. Harsham, Y. Shen, G. Divekar, M. Romoser, M. Knodler, and DL Fisher, "Evaluation of Different Speech and Touch Interfaces to In-Vehicle Music Retrieval Systems," *Accident Analysis & Prevention*, vol. 42, no. 3, pp. 913–920, 2010.

[2] M. Tonnis, V. Broy, and G. Klinker, "A Survey of Challenges Related to the Design of 3D User Interfaces for Car Drivers," in *in Proc. of IEEE 3D User Interfaces 2006*. IEEE, 2006, pp. 127–134.

[3] M.G. Jæger, M.B. Skov, N.G. Thomassen, et al., "You Can Touch, but You Can't Look: Interacting with In-Vehicle Systems," in *in Proc. of ACM* 26^{th} *Annual SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2008, pp. 1139–1148.

[4] S. Oviatt, P. Cohen, L. Wu, L. Duncan, B. Suhm, J. Bers, T. Holzman, T. Winograd, J. Landay, J. Larson, et al., "Designing the User Interface for Multimodal Speech and Pen-Based Gesture Applications: State-of-the-Art Systems and Future Research Directions," *Human-computer interaction*, vol. 15, no. 4, pp. 263–322, 2000.

[5] C. Muller and G. Weinberg, "Multimodal Input in the Car, Today and Tomorrow," *Multimedia, IEEE*, vol. 18, no. 1, pp. 98–103, 2011.

[6] T. Sim, S. Baker, and M. Bsat, "The CMU Pose, Illumination, and Expression (PIE) Database," in *in Proc. of IEEE Automatic Face and Gesture Recognition 2002*. IEEE, 2002, pp. 46–51.

[7] L. Bretzner, I. Laptev, and T. Lindeberg, "Hand Gesture Recognition Using Multi-Scale Colour Features, Hierarchical Models and Particle Filtering," in *in Proc. of IEEE Automatic Face and Gesture Recognition 2002*. IEEE, 2002, pp. 423–428.

[8] F.S. Chen, C.M. Fu, and C.L. Huang, "Hand Gesture Recognition Using a Real-Time Tracking Method and Hidden Markov Models," *Image and Vision Computing*, vol. 21, no. 8, pp. 745–758, 2003.

[9] X. Liu and K. Fujimura, "Hand Gesture Recognition Using Depth Data," in *in Proc. of IEEE Automatic Face and Gesture Recognition 2004*. IEEE, 2004, pp. 529–534.

[10] H.S. Yoon, J. Soh, Y.J. Bae, and H. Seung Yang, "Hand Gesture Recognition Using Combined Features of Location, Angle and Velocity," *Pattern Recognition*, vol. 34, no. 7, pp. 1491–1501, 2001.

[11] M. Zobl, M. Geiger, B. Schuller, M. Lang, and G. Rigoll, "A Real-Time System for Hand Gesture Controlled Operation of In-Car Devices," in *in Proc. of IEEE Multimedia and Expo 2003 (ICME 2003)*. IEEE, 2003, vol. 3, pp. III–541.

[12] M.Y. Chen, A. Kundu, and J. Zhou, "Off-Line Handwritten Word Recognition Using a Hidden Markov Model Type Stochastic Network," *Pattern Analysis and Machine Intelligence, IEEE Trans. on*, vol. 16, no. 5, pp. 481–496, 1994.

[13] P. Hong, M. Turk, and T.S. Huang, "Gesture Modeling and Recognition Using Finite State Machines," in *in Proc. of IEEE Automatic Face and Gesture Recognition 2000*. IEEE, 2000, pp. 410–415.

[14] C.L. Liu, S. Jaeger, and M. Nakagawa, "Online Recognition of Chinese Characters: the State-of-the-Art," *Pattern Analysis and Machine Intelligence, IEEE Trans. on*, vol. 26, no. 2, pp. 198–213, 2004.

[15] Y.C. Cheng, Z. Feng, F. Weng, and C.H. Lee, "Enhancing Model-Based Skin Color Detection: From Low-Level RGB Features to High-Level Discriminative Binary-Class Features," in *in Proc. of IEEE ICASSP 2012*. IEEE, 2012, pp. 1401–1404.

[16] J.L. Gauvain and C.H. Lee, "Maximum a Posteriori Estimation for Multivariate Gaussian Mixture Observations of Markov Chains," *Speech and Audio Processing, IEEE Trans. on*, vol. 2, no. 2, pp. 291–298, 1994.

[17] S.W. Lu, Y. Ren, and C.Y. Suen, "Hierarchical Attributed Graph Representation and Recognition of Handwritten Chinese Characters," *Pattern Recognition*, vol. 24, no. 7, pp. 617–632, 1991.

[18] H.Y. Kim and J.H. Kim, "Hierarchical Random Graph Representation of Handwritten Characters and its Application to Hangul Recognition," *Pattern Recognition*, vol. 34, no. 2, pp. 187–201, 2001.

[19] Q. Li, B.H. Juang, and C.H. Lee, "Automatic Verbal Information Verification for User Authentication," *Speech and Audio Processing, IEEE Trans. on*, vol. 8, no. 5, pp. 585–596, 2000.

[20] V. Vezhnevets, V. Sazonov, and A. Andreeva, "A Survey on Pixel-Based Skin Color Detection Techniques," in *Proc. Graphicon*. Moscow, Russia, 2003, vol. 3, pp. 85–92.

[21] Y. Wu and T.S. Huang, "Color Tracking by Transductive Learning," in *in Proc. of IEEE CVPR 2000*. IEEE, 2000, vol. 1, pp. 133–138.

[22] Q. Zhu, K.T. Cheng, C.T. Wu, and Y.L. Wu, "Adaptive Learning of an Accurate Skin-Color Model," in *in Proc. of IEEE Automatic Face and Gesture Recognition 2004*. IEEE, 2004, pp. 37–42.

[23] M. Nakai, N. Akira, H. Shimodaira, and S. Sagayama, "Substroke Approach to HMM-Based On-Line Kanji Handwriting Recognition," in *in Proc. of IEEE Document Analysis and Recognition 2001*. IEEE, 2001, pp. 491–495.

[24] M.J. Jones and J.M. Rehg, "Statistical Color Models with Application to Skin Detection," in *in Proc. of IEEE CVPR 1999*. IEEE, 1999, vol. 1.

[25] M.A.T. Figueiredo and A.K. Jain, "Unsupervised Learning of Finite Mixture Models," *Pattern Analysis and Machine Intelligence, IEEE Trans. on*, vol. 24, no. 3, pp. 381–396, 2002.