2011 Sixth IEEE International Symposium on Electronic Design, Test and Application

Adaptive Classifier For Robust Detection Of Signing Articulators Based On Skin Colour

Shujjat Khan[1], Donald Bailey[1], Gourab Sen Gupta[1], Serge Demidenko[2]
[1]School of Engineering and Advanced Technology (SEAT)
Massey University, New Zealand
[2]RMIT International University Vietnam, Ho Chi Minh City, Vietnam
Email: S.Khan@massey.ac.nz, d.g.bailey@massey.ac.nz, g.sengupta@massey.ac.nz, Serge.Demidenko@rmit.edu.vn

Abstract*—The proposed classifier is a novel skin detector that outperforms most of the existing approaches by dropping most of the non-skin pixels in its earlier stages of weak classifiers. Only the pixels with maximum skin likelihood are processed in later adaptive classifier. Parametric background modelling and validation based online training significantly improves the robustness of the whole classifier in any daily-life lighting conditions

Keywords- Skin detector, cascaded classifier, non-parametric model, Background model, adaptive histogram

I. INTRODUCTION

Real time human activity detection is the most significant task in vision-based applications such as facial expression, gesture recognition, surveillance and other Human-Computer Interaction (HCI) systems. Skin colour features are very basic and natural clue for detecting and analyzing any human action in a scene because of less computational complexity. Unlike other shape/texture features, skin colour classification is invariant to rotation and translation.

Recent work on gesture recognition highlights the significance of high precision and adaptive articulator detection whose accuracy severely influences the underlying recognition modules. For example, for better recognition of a Sign Language (SL), skin detector needs to segment the signing articulators (hands and face) out of the background with a larger success rate and less errors. During the literature review [1] of the problem, we come across few approaches with some absurd conditions whose violation hinders the system performance in a loose environment. Classifier training is another important concern for modern systems because most of the existing methods require a huge training set containing ground truth, which ultimately take sufficient time to train.

In this paper, we propose a novel cascaded skin classifier that outperforms most of the existing approaches by dropping as many non-skin pixels in earlier stages of classifier. This is an optimum solution for such application where a large area of input image is non-pixels. Only the pixels with maximum skin likelihood are processed in later classifier. Validation based online training also improves the robustness of whole classifier.

In section II, we discuss some existing skin colour detection methods and their implementation in gesture applications. Section III describes the proposed cascaded classifier followed by section IV and V containing result and conclusion respectively.

This research is sponsored by Massey University Research Fund (MURF 2010) and Bilateral Research Activities Program (BRAP 2008-2010), The Royal Society of New Zealand

II. ADAPTIVE SKIN DETECTOR

Human skin colour is formed by Red blood cells and Brownish Melanin hence their colour distribution forms a tight cluster in most of the colour spaces [2]. Simple form of skin classifiers consist of a defined boundary in a colour space representing skin-pixels while pixels lying out are classified as non-skin. Skin colour detection systems are severely affected by varying illumination, complex background, the signer's ethnicity (skin colour), and articulator occlusion. Some researchers use orthogonal colour spaces and illumination independent colour spaces for their experimentations [3]. In such chrominance based methods, some valuable skin colour will be lost whilst attempting to separate luminance and chrominance [2]. These methods may seem suitable for laboratory experimentation due to tightly controlled lighting but it will result in a large error rate if installed at public places like hospitals, courts and shops.

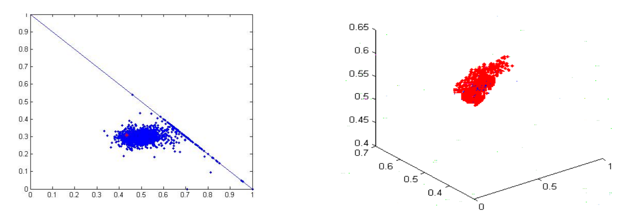

Figure 1. Skin colour distribution in Normalised-RG, and YCbCr Space.

For skin classification, a simple histogram based density estimator is a feasible solution provided a large amount of training data and sufficient system memory is available [4]. Farhad et al [5] utilises same skin modelling method to form an adaptive classifier by integrating temporal aspect of skinned organs (hands and face). The algorithm is based on the assumption that the skin regions are the only moving objects in the scene against a uniform static background. Hue-histogram is retrained by the in-motion pixels detected through frame differencing and optical flow methods. Another weighted histogram adaption methods based on ground truth data accumulated within a search window already overlaid during training session [6]. Qiang et al [7] developed an adaptive skin model which classifies the targeted skin pixels out of a larger

978-0-7695-4306-2/11 $26.00 © 2011 IEEE
DOI 10.1109/DELTA.2011.54

set of skin similar pixels. Skin similar pixels are those pixels which can belong to a wide range of skin colours. In an input image, true skin colour is parametrically detected by Gaussian modelling. Two separate Gaussians parametrically model both the classes, with the prominent Gaussian for skin pixels and the weaker one for false skin pixels in skin similar space. Main problem associated with this scheme is to model the background using a single Gaussian which is inappropriate for a cluttered background.

Figure 2. Two scenarios of skin colour and skin similar colour

Online training is another challenge faced by the vision researchers which requires number of representative samples and the strategy to retrain a system while it validates. In contrast to this, mostly skin classifiers may take huge amount of skin while shape based classifier may takes weeks to offline train. Generalization can also be associated with the outcome of training; i.e. how sensitive a system responds to an unseen sample. We will discuss our proposed skin detector in the context of these challenges.

III. CASCADED CLASSIFIER

Skin classifiers, if individually deployed, may cause more classification errors due to cluttered background, lack of pure ground truth training data, illumination and generalization for skin colour variation. The proposed classifier consists of a cascade of number of different filters and classifiers; background filtering, N-Rules RGB and combination of parametric and non-parametric classifiers followed by a contextual voting stage. Fig. 3 shows the overview of a cascaded skin classifier consists of 3 stages.

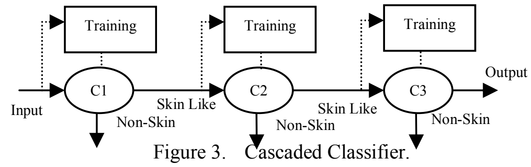

Figure 3. Cascaded Classifier.

Basic purpose of cascading number of weak classifiers is to achieve maximum detector's robustness and to reduce system computational complexity by dropping as many non-skin pixels in earlier stages. This scheme is quite useful especially when there is only a small proportion of skin area in the scene as compared to the background.

First stage of the cascade comprises of naïve N-Rules RGB classifier [2] which discards most of the non-skin pixels using following rules. In Fig 3, C1 is based upon Eq. 1.

$$C1:\begin{Bmatrix} 1 & R>G\ \ \&\ R>B \\ 0 & otherwise \end{Bmatrix} \tag{1}$$

Classifier C2 is an adaptive non-parametric Bayesian classifier which creates a 2-D Hue-Saturation probability histogram using Eq. 2 [1].

$$P(hs \mid skin) = \frac{s[hs]}{N} \tag{2}$$

Where $s[hs]$ represents skin pixel count in training images containing the ground truth and N is the total number of pixels. In verification stage, a Greyscale Skin Detection Mask (GSDM) is created by a process called "histogram back projection" in which each pixel is assigned a probability value representing its skin likelihood. Fig 4 shows a GSDM at the output of stage C2.

In next phase a contextual voting based classifier decides the pixel assignments to skin or non-skin class based on the voting of its neighbours. 4-neighbour classifier smoothes most of the undetected smaller skin patches inside larger skin region and filters out all isolated pixels as shown in Fig. 4.

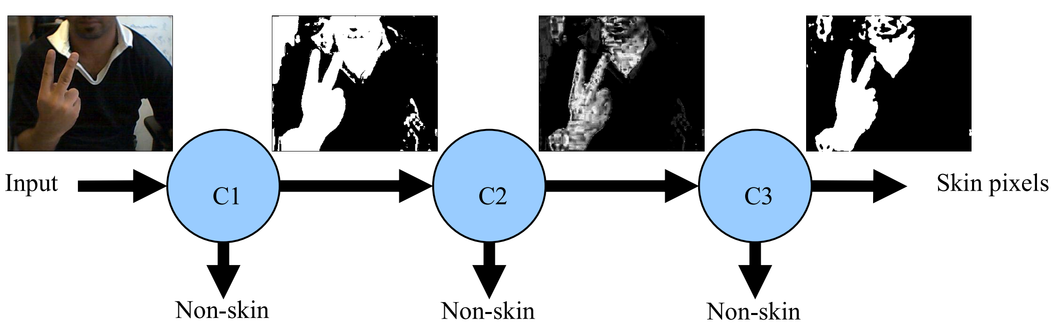

Figure 4. Adaptive classification.

Robustness of non-parametric skin classifiers heavily depends upon the nature and amount of training ground truth data. Classifier C2 is offline trained with approximately 5000 skin samples but for the operation lighting conditions, it is retrained by recurring skin samples over a defined training period.

Simple linear cascade of weak classifiers can increase the skin detection rate (true positive) on the expense of extra computations but overall system efficiency can be improved by carefully incorporating earlier classifiers of high rejection rate. For example, in most of the application, human articulators are the dominant moving objects so most of the non-skin components can be discarded by using the motion features.

A. *Movement detection*

Using Sum of Absolute Difference (SAD) method, we can detect the signer's movement and mask out all the stationary background pixels using a Change Detection Mask (CDM). SAD based frame differencing (Eq .1) is quite sensitive to very small inter-frame variations (lighting and small background activity) and causes high false positive in the CDM as shown in Fig. 4. Another approach called, Sum of Square Difference (SSD) not only detects prominent inter-frame variations but inherently suppresses minor false positives.

$$SAD = Abs(I_n - I_{n-1}) \tag{3}$$

$$SAD = (I_n - I_{n-1})^2 \tag{4}$$

B. *Adaptive classification*

Frame differencing methods are computationally inexpensive to spot the existence of target articulator (skin area) and these detected skin-like pixels are also used for histogram adaptation after passing through all the succeeding stages. This scheme also makes the whole classifier fully capable to adapt as per available lighting and shadows.

Skin detection system was tested with a range of signers with different skin colours and the results are shown in Fig. 5 Interesting observation of the detector is its learning capability, "how it learns the signer skin". Fig. 5 clearly shows the starting frames, where the system possesses not enough knowledge about the signer's skin but as the time passes, its detection rate increases significantly.

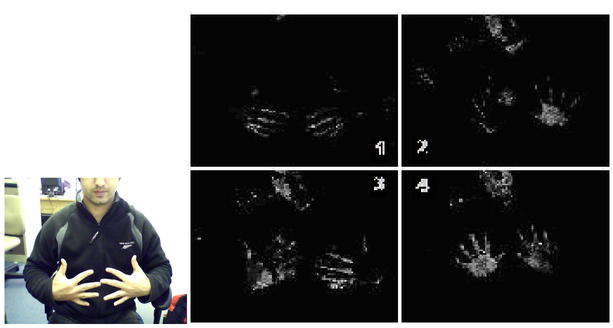

Figure 5. Test image and adaptive classification.

False positives in the classification are due to the skin-like colour objects (wooden desk and chair) present in the signing space. Frame differencing in early stage can not differentiate between the foreground and background objects so the skin-like background objects also become candidate skin area if they are present in signing space. In worst condition, a signer may continuously sign in front of a wooden wall or door (skin-like colour) and due to structure of our classifier, misclassified background samples will also contribute to new skin model. In this case, increasing the training period significantly deteriorates the classification performance. Fig. 6 demonstrates such situation where, signing in front of a skin-like background causes an overshoot in false positive.

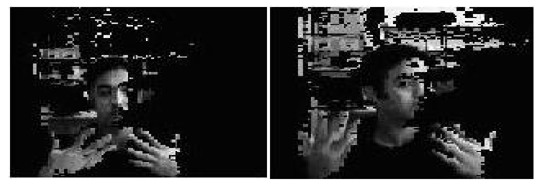

Figure 6. Model performance against a skin-similar background.

C. *Optimization*

In order to solve the skin-similar background problem, each pixel of the scene is parametrically modelled with a mean (μ) and standard deviation (σ) assuming the pixel distribution forms a natural distribution. Graphs in Fig. 7 show the Gaussian distributions of few random background pixels accumulated over a number of training images.

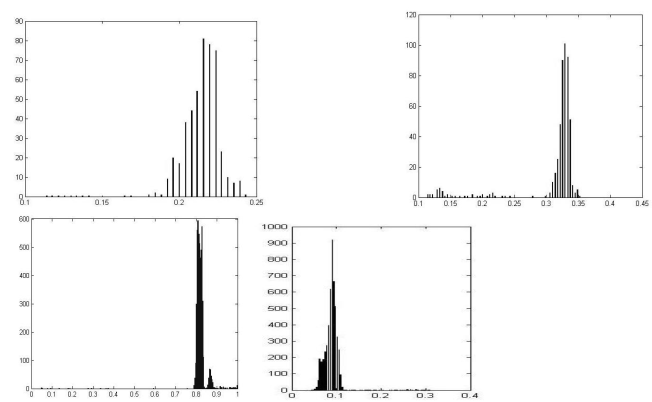

Figure 7. Distribution randomly selected Background pixels.

Background model is constructed in the training phase and each pixel in testing stage is compared with the model as per Eq. 5.

$$test = \begin{cases} foreground & u \pm test > T \\ background & otherwise \end{cases} \tag{5}$$

For this background model, Threshold T is selected as the number of standard deviation a test pixel is allowed to deviate from mean. We chose T= 2.5σ as a good compromise between true positives and false positives and the results are shown in Fig. 8.

Figure 8. Background classification ($T=2.5\sigma$)

A great advantage of optimizing the cascade with proper background modelling is its good classification for statics background and robustness against light variations (due to nature of illumination or door/window opening) and minor background activities (waving curtains). As obvious from the Fig .8, background classifier is an ideal choice which can significantly drop most of the background (non-skin) pixels in early stages.

IV. Conclusion

Pixel-based skin colour detection method classifies each pixel either as skin or non-skin independently while in region-based methods, neighbouring pixels also contribute deciding about an input pixel. In most of the skin classifiers, an input pixel is tested with pixel based classifier followed by region based approaches like morphological operations or connected component analysis. Our cascaded classifier approach drops most of the non-skin pixels by integrating parametric and non-parametric models in first few stages. With proper online and offline training, this combination of weak classifiers adaptively improves the classification performance in any daily-life lighting conditions.

References

[1] S. M. P. Kakumanu, N. Bourbakis, "A survey of skin-color modeling and detection methods", The journal of pattern recognition, 2007, pp. 106–1122.

[2] A. Cheddad, J. Condell, K. Curran, and P. M. Kevitt, "A skin tone detection algorithm for an adaptive approach to steganography", *Elsevier*, vol. 89, no. 12, 2009, pp. 2465-2478

[3] R. Akmeliawati, M. P. Leen Ooi, and Y. C. Kuang, "Real-Time Malaysian Sign Language Translation using Colour Segmentation and Neural Network ", *Instrumentation and Measurement Technology Conference (IMTC2007)*, Warsaw, Poland, pp. 1-6, 1-3 May, 2007

[4] M. J. Jones, and J. M. Rehg, "Statistical Color Models with Application to Skin Detection", *International Journal of Computer Vision (IJCV)*, vol. 46, no. 1, 2010, pp. 81-96.

[5] Farhad Dadgostar, and A. Sarrafzadeh, "An adaptive real-time skin detector based on Hue thresholding: A comparison on two tracking methods", Pattern Recognition Letters, Voll. 27, Issue 12, Spetember 2006, pp 1342-1352.

[6] A. K. B. See, and Y. K. Liaw, "Face detection and tracking utilizing enhanced CAMSHIFT model", *International journal of innovative computing, information and control*, vol. 10, 2006, pp. 1-12.

[7] Z. Qiang, C. Kwang-Ting, W. Ching-Tung, and W. Yi-Leh, "Adaptive learning of an accurate skin-color model", *Sixth IEEE International Conference on* Automatic *Face and Gesture Recognition*, Seoul, Korea,pp. 37-42, 17-19 May, 2004.