International Journal of Research in Advent Technology, Vol.6, No.8, August 2018
E-ISSN: 2321-9637
Available online at www.ijrat.org

A Comparative Study of Colour Spaces for Mouth Region Segmentation in Indian Context

Sandesh E PA[1] and Lajish V.L[2],Bibish Kumar K T[3], R.K.Sunil Kumar[4]
[1,2] *Department of Computer Science, University of Calicut - 673635*
sandeshepa@gmail.com, lajishvl@gmail.com
[3,4]*Department* of Physics, Govt. College, Madappally
bibishkrishna@gmail.com, seuron74@gmail.com

Abstract - Mouth region is the most visible and dynamic portion of human speech production system, which makes analysis and modelling of mouth region pixels crucial for developing effective man machine interfaces. The region mainly consists of lip, skin neighbourhood of lips, teeth, tongue and dark portions of mouth cavity. The colour variability of these mouth regions according to ethnicity is a well established fact. Hence mouth region segmentation algorithms designed based on colour information always try to include human images of different ethnicities. But so far the colour variability in Indian subcontinent and its effects on mouth region segmentation is not addressed adequately. In this study, statistical and probabilistic analysis of mouth region pixels is performed in different colour spaces on an own developed isolated phoneme audio visual speech corpora. The obtained analysis results are incorporated to a multiclass Bayesian classifier, which in turn is used to rank segmentation performance of colour components in each mouth region. The conclusions derived from this work can be used for developing effective visual speech synthesis and recognition systems applicable to human subjects belonging to Indian subcontinent.

Index Terms- Segmentation; Color models; Multi class mouth region segmentation; Bayesian classifiers

1. INTRODUCTION

Facial feature extraction and analysis is an active research area with potential applications in computer vision, visual speech recognition, data driven animation and in automatic sign language processing. Mouth motion accounts for the prominent non-rigid facial motion, especially during talking. Realistic and expressive mouth movements during talking are the most appealing aspect of animated characters and virtual agents. An analysis synthesis framework for facial motion animation, especially lip motion synthesis for speech animation is an emerging research topic [1-4]. Attempts are also made to track facial expression for imparting expressions to animation characters [5].

Automatic Speech Recognition has changed its mode of operation from audio-only speech recognition System to Audio-Visual Speech Recognition System (AV-ASR). The performance of Audio-only Speech Recognition System degrades drastically in the noisy environment [6-8]. In the light of this fact researchers in this field have incorporated the visual information with its audio counterpart which improve the robustness by providing complementary information [9,10]. Human's major visual speech information is provided by the lower part of the face, especially the mouth region. The reliability of a visual speech recognition system deeply depends on the accurate tracking of mouth.

Mouth area mainly consists of lip, skin neighborhood of lips, teeth, tongue and dark portions of mouth cavity. The range of human skin colour varies from white to black with discernable complexions including yellow, copper coloured and olive coloured [11]. The difference is primarily due to melanin content which varies greatly with ethnicities. The lip colour harmonizes with the background skin colour making lip segmentation using colour information quite difficult [12]. There are studies exploring the correlation of teeth colour with skin colour for various ethnicities [13]. So, while performing studies on colour based segmentation of skin or lip researchers ensure the presence of images representing different ethnicities [14-18]. But the ethnic group selected is mostly images of African, Mongolian and Anglo-Saxon ethnic origins. The wide ethnic and skin colour variability in Indian subcontinent is not addressed properly [19]. This work performs the statistical and probabilistic analysis of mouth region pixels in Indian context for different colour spaces. The results of analysis are incorporated to a multiclass Bayesian classifier for ranking the colour components according to its effectiveness in segmenting different mouth regions. The conclusions derived from this work can be used for developing effective visual speech synthesis and recognition systems which considers the colour ethnic peculiarities and orientations of Indian subcontinent. Even though lip segmentation and tracking has been studied deeply, only very few works has addressed the problem of segmenting different mouth regions including teeth and tongue. Tongue segmentation is addressed recently in some works especially for applications related to traditional Chinese medicine [20-25]. Lip segmentation attempts can be broadly classified into two classes, namely Model based approach and Colour based approach [26,27]. In model based approach mathematical models of lip contour are used as a set of model parameters for lip segmentation. Active shape and appearance model, snake model and deformable templates are widely used methods in this category [28-33]. In colour based approach, the colour triplet values of skin and lip pixels are used as the basic information for segmentation [34-38].

The paper is structured as follows. After a brief introduction in section 1, section 2 sketches the basics of colour based approaches to mouth area segmentation with a brief explanation of 5 colour spaces selected for the study. Section 3 describes the preparation of the audio visual speech corpora in the Malayalam language used in the work. The results of Statistical and probabilistic analysis of lip, skin, teeth, tongue and oral cavity pixels in different colour spaces are discussed in section 4. Section 5 presents mouth region segmentation in different colour spaces using naïve Bayesian classification. The paper is concluded in section 6 with future directions.

2. COLOUR SPACES USED FOR MOUTH REGION MODELLING

Lip, tongue, teeth and skin colour of human beings greatly varies with ethnicity. In Indian context the contrast between lip and skin colour is found to be low compared to other ethnicities. Statistical analysis of mouth region pixel colours in Indian context is an urgent requirement for developing native tools in the domain of visual speech recognition and synthesis. This paper is an attempt towards this direction. Colour can be represented in different colour spaces such as RGB, normalized RGB, HSI, CIE Lab and YCbCr. The discriminating power among mouth regions is different for different colour spaces. Selection of optimum colour space which performs segmentation process effectively is still being actively debated among researchers. The optimum colour space is found to be depending on the mouth region colour properties of the population or on the ethnicity. In this paper the statistical analysis of skin and lip pixels of peoples in Indian subcontinent is executed in 5 different colour spaces.

Eveno et al performed the most exhaustive survey for finding the appropriate colour space for lip – skin segmentation [39]. It compares 7 colour spaces and 12 additional special transforms and prepared a ranking of colour channels according to the suitability for segmentation. Axel et.al.suggests a method for lip segmentation based on rgb-colour histogram[40]. Threshold-based segmentation strategies give a simple and effective approach to implement lip segmentation. Gritzman, Ashley D. et al [41] propose a method called adaptive threshold optimisation (ATO) which selects the threshold by using feedback of shape information. Statistical-colour model approach [42, 43] is used to automatically find and track the face and other facial features in the image. Jian-qiang Du et.al uses the hue and intensity component based thresholding scheme for tongue segmentation [44]. The detailed review on skin, lip and other mouth area segmentation works identified 5 colour spaces which are most frequently used and reported to be most effective for the purpose. Section 2.1 gives a brief explanation of these 5 colour spaces, RGB, Normalized RGB, HSV, CIELa*b*, and YCbCr selected for this study.

2.1 *Colour spaces*

A colour space is a mathematical representation of a set of colours. Most colour spaces are derived from the RGB colour space. Different colour spaces are suitable for different image processing applications. RGB, Normalized RGB, HSV, CIE La*b*, and YCbCr are the five colour spaces selected for this study.

2.1.1. *RGB colour space*

RGB colour space is the simplest and most widely used method in computer graphics. Luminance of a given RGB pixel is a linear combination of the R, G and B values. High correlation between channels, significant perceptual non-uniformity and mixing of chrominance and luminance data make RGB a bad choice for colour analysis and colour based recognition algorithms.

2.1.2. *Normalized RGB colour space*

Normalised RGB is invariant to changes of surface orientation relative to the light source [45]. The normalised RGB values are represented as r, g, and b .These values are obtained by a normalisation procedure as shown bellow.

$$r = \frac{R}{R+G+B,}, g = \frac{G}{R+G+B}, b = \frac{B}{R+G+B} \qquad (1)$$

2.1.3. *HSV colour space*

Hue, Saturation and Value colour space is a non-linear transformation of RGB colour space into a cylindrical coordinate representation. Hue is the wavelength at which the enegy is maximum or the colour of the pixel is most prominent. Saturation is the slope of the bandwidth curve around the central maximum. colour wil be purer if saturation is high and vice-verse [46]. This class of colour model is the most intuitive or artistic way of describing a colour. This model is closest to the way humans percvieve colour. Hence it is the most widely used model in computer vision applications. The detailed algorithm for converting from RGB to HSV is given in the work of Jim et al [47].

2.1.4. *CIE L*a*b colour space*

In La*b*(CIE 1976 L*a*b*), the 'L' channel represents the human perception of luminosity and the 'a*' and 'b*' components provide colour information. The range of L channel varies between black and white and the channel 'a*' varies from red to green similarly channel 'b*' varies from blue to yellow. It encodes the perceptual difference in colours when viewed by humans [48].

2.1.5. *YCbCr colour space*

In YCbCr, the gray scale information is carried by the 'Y' component and colour information is provided by two colour channels, 'Cb' and 'Cr'. Cb is obtained as the difference between blue component and a reference value and Cr is obtained as the difference between red component and a reference value. The Y value is obtained as a weighted sum of R, G, and B triplet [48].

3. VISUAL SPEECH DATABASE CREATION AND PRE-PROCESSING

This work uses an own developed audio visual speech corpora of 20 trained female speakers. The human subjects are selected so as to include the entire colour variability of Indian subcontinent. The age of human subjects spans in the range of 20 to 45 years. The procurement is restricted to female speakers to avoid complexities introduced by facial hairs of male subjects. The data base consists of isolated phoneme utterances in Malayalam. Malayalam is a Dravidian language used by more than 38 million people and it is the mother tongue of Kerala, Mahe and lakshdweep . Each speaker utters the phonemes in Malayalam in a silence-utterance- silence manner. Malayalam has a rich phoneme set consisting of 51 phonemes. The video is captured using a SONY Camera HDR-CX405 handy cam having frame rate of 25fps with a resolution of 1280 x 720 in MP4 format and audio sampled at 44100Hz. The recording is made in an ordinary office room with normal lighting conditions.5 frames for each phoneme are manually selected for analysis from the recordings of each speaker with maximum visual phoneme presence. Section 3.1 discusses the pre-processing performed on the images to be used for statistical analysis.

3.1 *Pre-processing*

Removal of back ground pixels and labelling remaining pixels as either skin, lip, teeth, tongue or mouth cavity are the steps performed as part of pre-processing. The inner and outer lip contour is segmented by manually placing 36 landmark points on each image. The outer lip contour is represented by 20 points, while the inner lip contour is represented by 16 points. The lip region is segmented using these land mark points. Fig. 1 shows an image with 36 landmark points around inner and outer lip contours.

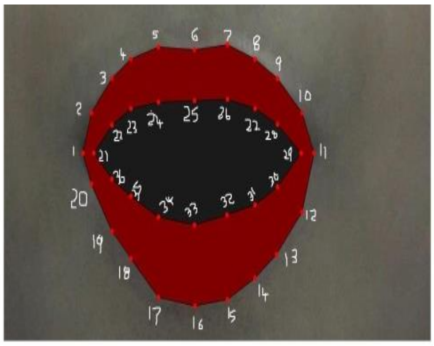

Fig. 1. 36 land mark points around inner and outer lip

The teeth and tongue regions inside the inner lip portion are segmented using semi automatic thresholding techniques. A mask is created for each image in the data base with separate labels for skin, lip, tongue, dark and teeth region (with corresponding labels 0,1,2,3and 4 respectively). Back ground pixels are eliminated by selecting a rectangular window based on the centroid of land mark points. The cropped images and the mask with labels are used for statistical analysis of different regions and will act as a ground truth values for testing purpose. Fig. 2 shows the output images and mask after preprocessing for two different images in the database. The second column displays the mask with separate labels for skin, lip, teeth, tongue and dark regions.

4. STATISTICAL ANALYSIS OF DIFFERENT MOUTH REGION PIXELS

This section consolidates the statistical analysis performed on the images for understanding the colour properties of different mouth regions. The results of analysis can be used for developing mouth region segmentation tools adapted to the colour peculiarities of Indian subcontinent. The colour information can also be used for developing native speech animation applications. The properties of 15 colour components (from five colour spaces) in five different mouth regions is performed on the images visual speech corpora explained in section 3.Mean, standard deviation ,class conditional probabilities and prior probabilities are estimated for pixels in each region.

Other than skin, lip, tongue and teeth a separate class is employed for representing dark pixels inside the mouth cavity. The statistics is obtained after the analysis of around one billion pixels of 20 speakers. Table 1 gives the mean and standard deviation of colour components belonging to 5 colour spaces in 5 different mouth regions.

RGB and YCbCr are defined in the 0- 255 range, while HSV and nRGB are defined in the range 0 - 1. In La*b* colour space L is defined over the range 0 - 100 and a* and b* is defined in the range 0 – 255. The dynamic range of mouth region pixels is below 50% of the possible range for all components in the RGB space. But the hue component of HSV space encompasses more than 75% of the possible range which is the maximum of all colour components. The class conditional probabilities and prior probabilities of different mouth region pixels are also estimated for 15 colour components. The estimated values are used for naive Bayesian classifier which will be discussed in the next section. The class conditional probability distribution of different regions for the Hue component in the HSV colour space and Cr component of YCbCr is shown in Fig.3 and Fig.4.

In the distribution of H component teeth pixels has a distinct domain represented by blue colour in the figure, while lip pixels has colour presence in both low and high values of the spectrum. There is considerable overlap in the teeth, tongue and mouth cavity pixels in this colour space. The distinctiveness of teeth pixels is repeated in Cr values also, but the dynamic range of Cr values is comparatively small. In order to rank the effectiveness of colour components for practical purposes such as segmentation more objective evaluations are needed. The next section describes the performance evaluation of different colour components obtained using an automatic classifier using naive Bayesian approach.

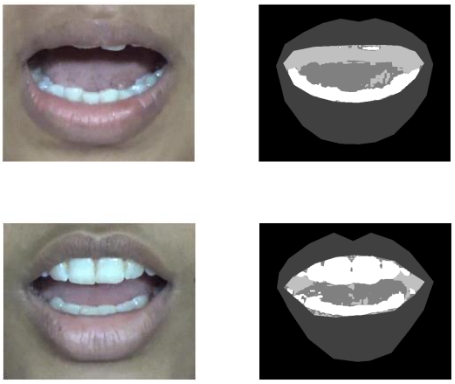

Fig. 2. Sample images and mask after pre-processing

5. SEGMENTATION BASED ON BAYSIAN CLASSIFIER

A 5 class Bayesian classifier is designed to test the distinguishing power of mouth region pixels in different colour spaces. Skin, lip, teeth, tongue, dark regions inside moth cavity is the 5 visually different areas in the mouth region. The class conditional probabilities and prior probabilities computed from the training database are used for designing a multiclass Bayesian classifier.

5.1 *Histogram model with native Bayesian classifier*

Bayesian classification is based on a training phase which calculates the class conditional probability or likely hood for each class and prior probability for each class. Bayesian classification has been employed for solving skin and lip segmentation problems [49-52].

The idea is to divide the colour space in to histogram bins, where each bin stores the count corresponding to the occurrence of that colour in the training database. The counts are converted in to class conditional probability measures. Here Bayesian classification is attempted to label each pixel either as skin, lip, teeth, tongue or dark region pixel. Let wi's be the set of classes (in this case 5 mouth regions) and x is the colour value of the current pixel Bayes' theorem helps us to calculate the posterior probability$P\left({}^{w_i}/_{x}\right)$, the probability of observing the ith mouth region given a colour value x

$$P\left({}^{w_i}/_{x}\right) = \frac{P\left({}^{x}/_{w_i}\right).P(w_i)}{P(x)} \qquad (2)$$

$P\left({}^{x}/_{w_i}\right)$ is the class conditional probability or likely hood,$P(w_i)$ is the prior probability for each class (both calculated from the training data) . The

Table 1. Mean and standard deviation of mouth regions in 15 colour components from 5 different colour spaces

Colour Component	Mean					Standard Deviation				
	Skin	Lip	Teeth	Tongue	Dark	skin	lip	Teeth	Tongue	Dark
	HSV									
H	0.116	0.682	0.564	0.770	0.865	0.370	0.353	0.045	0.079	0.095
S	0.131	0.134	0.137	0.125	0.238	0.024	0.041	0.043	0.031	0.088
V	0.521	0.513	0.737	0.522	0.289	0.054	0.088	0.136	0.106	0.110
La*b*										
L	53	49.48	71.01	50.52	25.86	5	7.68	12.74	10	11.63
a*	128.75	135.3	123.3	135.5	135.7	4.01	3.44	4.39	2.58	2.85
b*	134.29	127.8	120.6	121.6	125.5	3.88	2.47	2.94	2.32	2.47
RGB										
R	133	130.6	157.6	129.1	72.6	13.8	22.6	25.5	25.85	28.91
G	125.8	113.8	177.9	117.04	58.4	12.8	18.4	36.13	25.04	26.03
B	115.64	118.2	187.6	131.8	66.09	12.11	20.12	34.73	27.63	26.45
Normalised RGB										
R	0.3556	0.360	0.302	0.342	0.374	0.006	0.011	0.010	0.008	0.026
G	0.359	0.314	0.338	0.309	0.289	0.013	0.008	0.009	0.008	0.019
B	0.3088	0.325	0.359	0.348	0.336	0.010	0.008	0.009	0.008	0.019
YCbCr										
Y	108.93	102.5	148.5	105.0	54.59	10.77	16.80	28.32	21.86	22.96
Cb	122.45	127.3	135.1	132.4	129.1	3.01	1.98	2.81	2.11	1.55
Cr	131.85	134.9	118.4	132.0	133.5	2.26	3.19	5.05	2.17	2.67

probability of happening colour x, p(x) is the same for all classes hence it is neglected for computations for classification[53]. A pixel with colour value x is assigned a label (representing the class membership) calculated using equation (3)

$$label = Max_{i=0}^{4}\left(P\left({}^{x}/_{w_i}\right).P(w_i)\right) \quad (3)$$

In the testing phase a mask is created corresponding to each image, where each pixel is assigned a label based on its identified class. Fig. 5 shows the image and the mask obtained after Bayesian classification on an image which is not in the training set. The mask thus obtained for 15 colour components can be compared with the ground truth mask prepared as part of pre-processing. Region wise and over all classification accuracies are computed by comparing the mask obtained from Bayesian classification and ground truth mask.

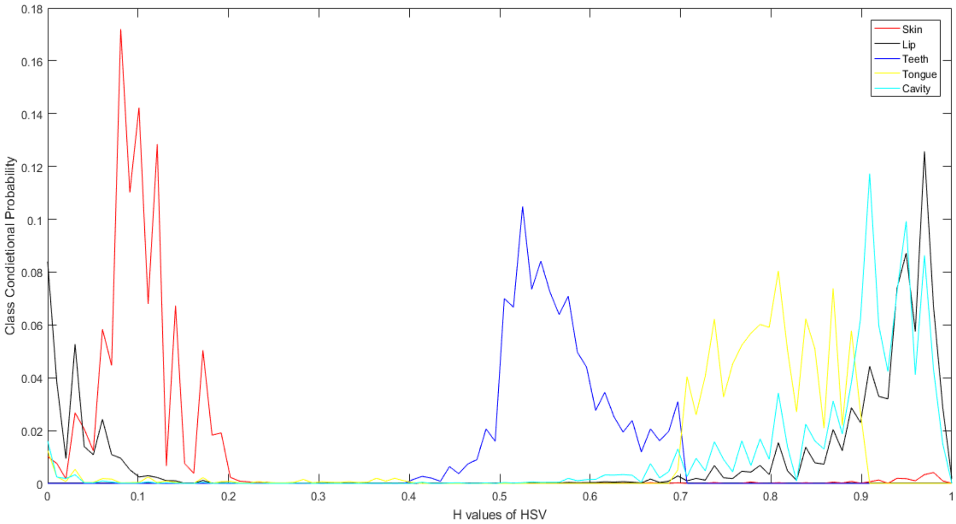

Fig. 3. Class conditional probability of H component of HSV

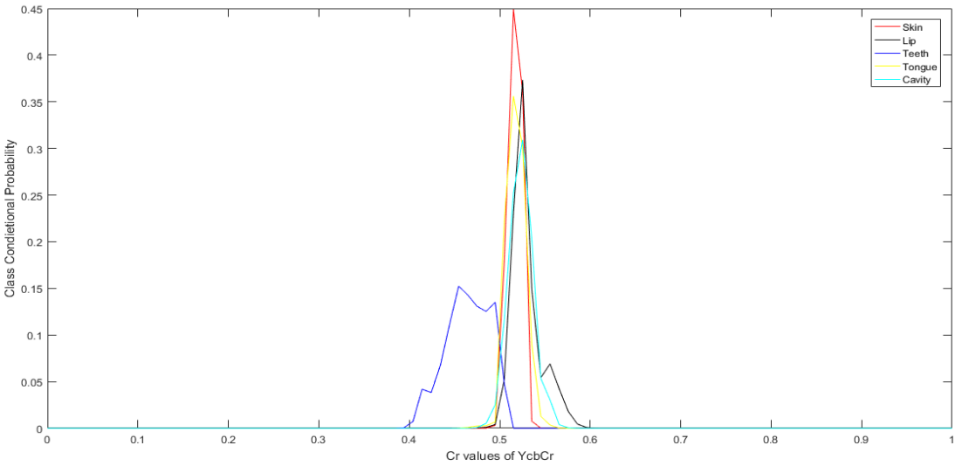

Fig. 4. Class conditional probability of Cr component of YCbCr (Normalised in the 0 – 1 range)

5.2 *Performance evaluation*

The selection of colour space is crucial for discriminating different mouth regions. The principle objective of this work is to find the optimal colour component for classifying mouth regions pixels of peoples of Indian subcontinent. The performance of the multiclass segmentation in different colour spaces is compared in this section. The segmentation mask obtained after Bayesian classification is used for comparison. The accuracy of classification is computed separately for each region by comparing with the ground truth mask. The 15 colour components are arranged according to their region wise performance in table 2.

The overall performance and tongue segmentation performance is best for hue component in the HSV colour space. Lab –a for skin, nRGB – G for lip, YCbCr – Cr for teeth and RGB- R for dark are other toppers in the list. The accuracy is found to be lowest for tongue segmentation. Fig. 6 is the graphical representation of overall segmentation accuracy of different colour components.

The percentage of overall accuracy ranges from 4.82(HSV - V) to 84.19(HSV-H). The skin accuracy is of the range 6.59(HSV - V) to 99.48(Lab-a), while the lip performance range is 2.02(HSV-S) to 92.73(nRGB-G). More than 99 percentages of teeth pixels are correctly identified in both YCbCr – Cr and HSV-H colour components, while teeth pixel identification is around zero for S and V components of HSV colour space. The range of accuracy for dark pixels varies from 0(Lab-a,b) to 96.8(RGB-R). The accuracy is lowest for tongue identification which is in the range 0(Lab-b) to 79.8(HSV-H).

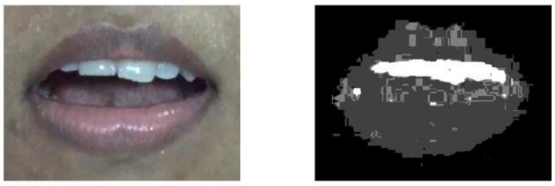

Fig. 5. Original image and mask created using Multiclass Beysian Classifier in b* Colour Component of La*b* space

6. CONCLUSION

This paper has carried out an in-depth analysis of mouth region colour pixel in five different colour spaces covering the entire complexion range of Indian population. The statistical analysis of different colour models (RGB, Normalised RGB, HSV, CIE LAB and YCbCr) shows the strength and weakness in modelling and segmenting different mouth regions. The effectiveness of 15 colour components in 5 colour spaces is evaluated against a multiclass beysian classifier and the colour components are ranked according to the performance. Separate ranking is employed for skin,lip,teeth,tongue and mouth cavity segmentation performance. Of these colour models Lab –a (for skin), nRGB – G(for lip), YCbCr – Cr (for teeth) HSV – H(for tongue)and RGB- R(for Mouth cavity) are found to be the best performing components. The overall classification accuracy is found to be best for H component of HSV colour space. Experimenting Combination of colour channels from different colour spaces in Indian context is the next logical step. The current practice in model based methods is employing gray scale values of RGB colour values as inputs. This ranking can be used for selecting more appropriate colour space for model based methods in mouth region segmentation and tracking applications.

REFERENCES

[1] Beaumesnil, Brice, Franck Luthon, and Marc Chaumont. "Liptracking and MPEG4 animation with feedback control." ICASSP'06: IEEE International Conference On Acoustics, Speech, and Signal Processing. Vol. 2. 2006.

[2] Beaumesnil, Brice, and Franck Luthon. "Real time tracking for 3D realistic lip animation." Pattern Recognition, 2006. ICPR 2006. 18th International Conference on. Vol. 1. IEEE, 2006.

[3] Weissenfeld, Axel, Kang Liu, and Jörn Ostermann. "Video-realistic image-based eye animation via statistically driven state machines." The Visual Computer 26.9 (2010): 1201-1216.

[4] Wu, Zhilin, Petar S. Aleksic, and Aggelos K. Katsaggelos. "Lip tracking for MPEG-4 facial animation." Multimodal Interfaces, 2002. Proceedings. Fourth IEEE International Conference on. IEEE, 2002.

[5] Seyedarabi, Hadi, Ali Aghagolzadeh, and Sohrab Khanmohammadi. "Facial expressions animation and lip tracking using facial characteristic points and deformable model." International Journal of Information Technology 1.4 (2004): 165-168.

[6] Benoit, Christian, et al. "Which components of the face do humans and machines best speechread?." Speechreading by humans and machines. Springer, Berlin, Heidelberg, 1996. 315-328.

[7] Burnham, Douglas, et al. Hearing eye II: the psychology of speechreading and auditory-visual speech. Routledge, 2013.

[8] Sumby, William H., and Irwin Pollack. "Visual contribution to speech intelligibility in noise." The journal of the acoustical society of america 26.2 (1954): 212-215.

[9] Sumby, William H., and Irwin Pollack. "Visual contribution to speech intelligibility in noise." The journal of the acoustical society of america 26.2 (1954): 212-215.

[10] Neti, Chalapathy, et al. "Large-vocabulary audio-visual speech recognition: A summary of the Johns Hopkins Summer 2000 Workshop." Multimedia Signal Processing, 2001 IEEE Fourth Workshop on. IEEE, 2001.Tongue Segmentation Using Active Contour Model

[11] Cotton, S. D., and Ela Claridge. "Do all human skin colours lie on a defined surface within LMS space?." SCHOOL OF COMPUTER SCIENCE RESEARCH REPORTS-UNIVERSITY OF BIRMINGHAM CSR (1996). Sadeghi, M., J. Kittler, and K. Messer. "Modelling and segmentation of lip area in face images." IEE Proceedings-Vision, Image and Signal Processing 149.3 (2002): 179-184.

[12] Haralur, Satheesh B., et al. "The tooth and skin colour interrelationship across the different ethnic groups." International journal of dentistry 2014 (2014).

[13] Hassanpour, Reza, Asadollah Shahbahrami, and Stephan Wong. "Adaptive Gaussian mixture model for skin colour segmentation." World Academy of Science, Engineering and Technology 41 (2008): 1-6.

[14] Oliveira, V. A., and A. Conci. "Skin Detection using HSV colour space." H. Pedrini, & J. Marques de Carvalho, Workshops of Sibgrapi. 2009.

[15] Kakumanu, Praveen, Sokratis Makrogiannis, and Nikolaos Bourbakis. "A survey of skin-colour modeling and detection methods." Pattern recognition 40.3 (2007): 1106-1122.

[16] Habili, Nariman, Cheng-Chew Lim, and Alireza Moini. "Automatic human skin segmentation based on colour information in the YCbCr colour space." Proceedings of the Information, Decision and Control Conference Final Program and Abstracts. 2002.

[17] Bhoyar, K. K., and O. G. Kakde. "Skin colour detection model using neural networks and its performance evaluation." Journal of computer science. 2010.

[18] Tabassum, Mirza Rehenuma, et al. "Comparative study of statistical skin detection algorithms for sub-continental human images." arXiv preprint arXiv:1008.4206 (2010)

[19] Zhai, Xueming, Hang-dong Lu, and Lizhong Zhang. "Application of image segmentation technique in tongue diagnosis." Information Technology and Applications, 2009. IFITA'09. International Forum on. Vol. 2. IEEE, 2009.

[20] Shi, Miao-Jing, et al. "Computerized tongue image segmentation via the double geo-vector flow." Chinese medicine 9.1 (2014): 7.

[21] Wu, Kebin, and David Zhang. "Robust tongue segmentation by fusing region-based and edge-based approaches." Expert Systems with Applications 42.21 (2015): 8027-8038.

[22] Cui, Zhenchao, et al. "Automated tongue segmentation based on 2D Gabor filters and fast marching." International Conference on Intelligent Science and Big Data Engineering. Springer, Berlin, Heidelberg, 2013.

[23] Ning, Jifeng, et al. "Automatic tongue image segmentation based on gradient vector flow and region merging." Neural Computing and Applications 21.8 (2012): 1819-1826.Mehrotra, Harshit, Gaurav Agrawal, and M. C. Srivastava. "Automatic lip contour tracking and visual character recognition for computerized lip reading." International Journal of Computer Science 4, no. 1 (2009): 62-71.

[24] Skodras, Evangelos, and Nikolaos Fakotakis. "An unconstrained method for lip detection in colour images." In Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on, pp. 1013-1016. IEEE, 2011.

[25] Luettin, Juergen, Neil A. Thacker, and Steve W. Beet. "Active shape models for visual speech feature extraction." In Speechreading by humans and machines, pp. 383-390. Springer Berlin Heidelberg, 1996.

[26] Sum, K. L., Wai H. Lau, S. H. Leung, Alan Wee-Chung Liew, and K. W. Tse. "A new optimization procedure for extracting the point-based lip contour using active shape model." In Acoustics, Speech, and Signal Processing, 2001. Proceedings. (ICASSP'01). 2001 IEEE International Conference on, vol. 3, pp. 1485-1488. IEEE, 2001.

[27] Nguyen, Quoc Dinh, and Maurice Milgram. "Semi adaptive appearance models for lip tracking." In Image Processing (ICIP), 2009 16th IEEE International Conference on, pp. 2437-2440. IEEE, 2009.

[28] Kass, Michael, Andrew Witkin, and Demetri Terzopoulos. "Snakes: Active contour models." International journal of computer vision 1, no. 4 (1988): 321-331.

[29] Jang, Kyung Shik. "Lip contour extraction based on active shape model and snakes." International Journal of Computer Science and Network Security 7, no. 10 (2007): 148-153.

[30] Liu, Xin, and Yiu-ming Cheung. "A robust lip tracking algorithm using localized colour active contours and deformable models." In Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on, pp. 1197-1200. IEEE, 2011.

[31] Kushwahaa, Riti, Neeta Naina, and Promila Jangraa. "Extraction of Lip Contour from Face." (2012).

[32] Liew, AW-C., Shu Hung Leung, and Wing Hong Lau. "Segmentation of colour lip images by spatial fuzzy clustering." IEEE transactions on Fuzzy Systems 11, no. 4 (2003): 542-549.

[33] Celenk, Mehmet. "A colour clustering technique for image segmentation." Computer Vision, Graphics, and image processing 52, no. 2 (1990): 145-170.

[34] Fu, King-Sun, and J. K. Mui. "A survey on image segmentation." Pattern recognition 13, no. 1 (1981): 3-16.

[35] Cheung, Yiu-ming, and Meng Li. "MAP-MRF based LIP segmentation without true segment number." In Image Processing (ICIP), 2011 18th IEEE International Conference on, pp. 769-772. IEEE, 2011.

[36] Gritzman, Ashley D., David M. Rubin, and Adam Pantanowitz. "Comparison of colour transforms used in lip segmentation algorithms." Signal, Image and Video Processing 9, no. 4 (2015): 947-957.

[37] Hassanat, Ahmad BA, and Sabah Jassim. "Colour-based lip localization method." In Proceedings of SPIE-The International Society for Optical Engineering. 2010.

[38] Eveno, Nicolas, Alice Caplier, and P-Y. Coulon. "New colour transformation for lips segmentation." In Multimedia Signal Processing, 2001 IEEE Fourth Workshop on, pp. 3-8. IEEE, 2001.

[39] Panning, Axel, Robert Niese, Ayoub Al-Hamadi, and Bernd Michaelis. "A new adaptive approach for histogram based mouth segmentation." Proceedings of the World Academy of Science, Engineering and Technology 56 (2009): 779-784.

[40] Gritzman, Ashley D., Vered Aharonson, David M. Rubin, and Adam Pantanowitz. "Automatic computation of histogram threshold for lip segmentation using feedback of shape information." Signal, Image and Video Processing10, no. 5 (2016): 869-876.

[41] Stiefelhagen, Rainer, Uwe Meier, and Jie Yang. "Real-time lip-tracking for lipreading." In Eurospeech. 1997.

[42] Stiefelhagen, Rainer, Jie Yang, and Alex Waibel. "A model-based gaze tracking system." International Journal on Artificial Intelligence Tools6, no. 02 (1997): 193-209.

[43] Du, Jian-qiang, et al. "A novel algorithm of color tongue image segmentation based on HSI." *BioMedical Engineering and Informatics, 2008. BMEI 2008. International Conference on*. Vol. 1. IEEE, 2008.

[44] Vezhnevets, Vladimir, Vassili Sazonov, and Alla Andreeva. "A survey on pixel-based skin colour detection techniques." In Proc. Graphicon, vol. 3, pp. 85-92. 2003.

[45] Bigun, J. "Vision with Direction: A Systematic Introduction to Image Processing and Vision." (2006).

[46] Schwiegerling, Jim. "Field guide to visual and ophthalmic optics." Spie, 2004.

[47] Ibraheem, Noor A., Mokhtar M. Hasan, Rafiqul Z. Khan, and Pramod K. Mishra. "Understanding colour models: a review." ARPN Journal of Science and Technology 2, no. 3 (2012): 265-275.

[48] Chai, Douglas, and Abdesselam Bouzerdoum. "A Bayesian approach to skin colour classification in YCbCr colour space." TENCON 2000. Proceedings. Vol. 2. IEEE, 2000..

[49] Yang, Jie, Weier Lu, and Alex Waibel. "Skin-colour modeling and adaptation." Asian Conference on Computer Vision. Springer, Berlin, Heidelberg, 1998.

[50] Zarit, Benjamin D., Boaz J. Super, and Francis KH Quek. "Comparison of five colour models in skin pixel classification." Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems, 1999. Proceedings. International Workshop on. IEEE, 1999.

[51] Yoo, Tae-Woong, and Il-Seok Oh. "A fast algorithm for tracking human faces based on chromatic histograms." Pattern Recognition Letters 20.10 (1999): 967-978.

[52] Rish, Irina. "An empirical study of the naive Bayes classifier." IJCAI 2001 workshop on empirical methods in artificial intelligence. Vol. 3. No. 22. New York: IBM, 2001.

[53]

Table 2. The 15 colour components are arranged according to their region wise performance

Over all	Skin	Lip	Teeth	Dark	Tongue
HSV – H	Lab – a	nRGB – G	YCbCr - Cr	RGB – R	HSV – H
Lab – b	RGB – G	nRGB – B	HSV – H	YCbCr – Y	Lab – b
Lab – a	YCbCr – Y	YCbCr - Cb	nRGB - R	HSV – S	Lab – a
YCbCr – Cb	Lab – L	Lab – b	nRGB - B	Lab – L	RGB – G
RGB – G	Lab – b	HSV – H	YCbCr - Cb	RGB – G	YCbCr – Cb
nRGB – G	RGB – R	Lab – a	Lab – b	RGB – B	nRGB – G
YCbCr – Y	HSV – H	RGB – G	RGB – B	nRGB – G	YCbCr – Y
Lab – L	HSV – S	YCbCr - Cr	Lab – a	nRGB – B	Lab – L
nRGB – B	YCbCr – Cr	nRGB – R	RGB – G	HSV – V	YCbCr – Cr
YCbCr – Cr	RGB - B	YCbCr – Y	Lab – L	nRGB – R	nRGB – B
RGB – R	nRGB - G	Lab – L	YCbCr - Y	HSV – H	RGB – R
RGB – B	YCbCr - Cb	RGB – R	RGB – R	YCbCr - Cb	RGB – B
HSV – S	nRGB - B	RGB – B	nRGB - G	YCbCr - Cr	nRGB – R
nRGB – R	nRGB - R	HSV – V	HSV – S	Lab – a	HSV – S
HSV – V	HSV - V	HSV – S	HSV – V	Lab – b	HSV – V

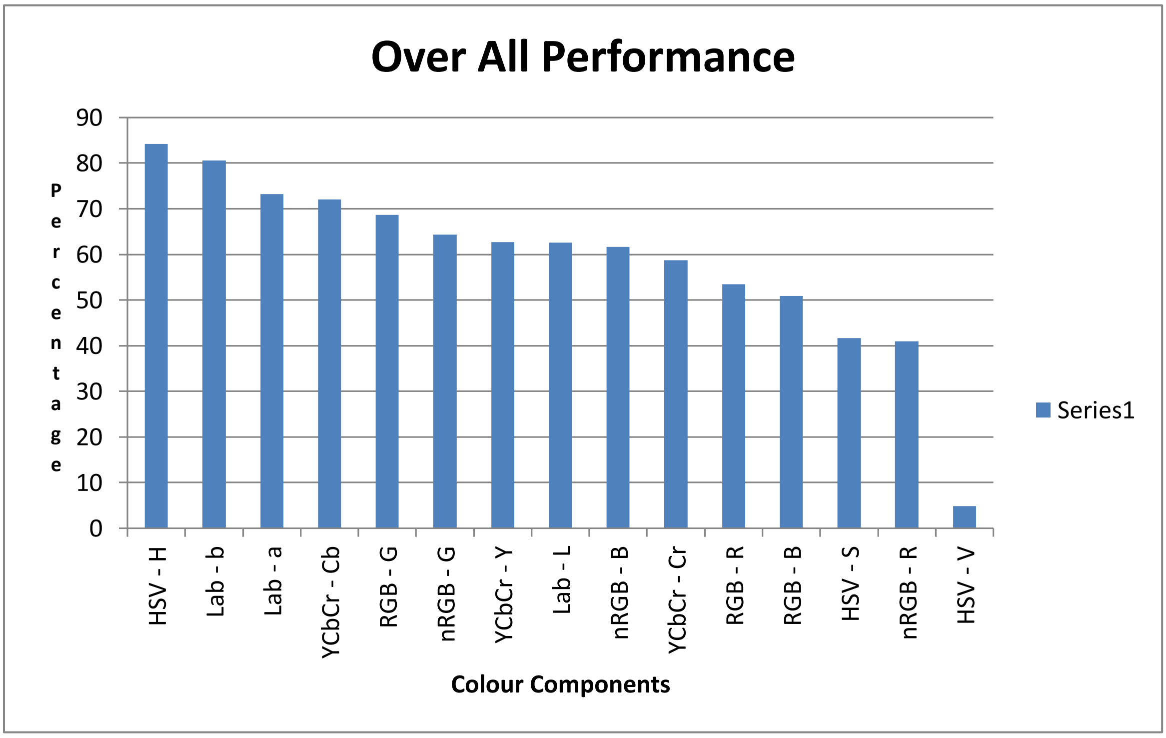

Fig. 6. Graphical representation of overall segmentation accuracy of different colour components.