Proceedings of the 28th IEEE
EMBS Annual International Conference
New York City, USA, Aug 30-Sept 3, 2006

Feature Extraction & Lips Posture Detection Oriented to the Treatment of CLP Children

A. Salazar, G. Daza S., L. Sánchez, F. Prieto, G. Castellanos, C. Quintero†
Universidad Nacional de Colombia Sede Manizales
† Universidad de Caldas
aesalazarj, gdazas, lgsanchezg, faprietoo, cgcastellanosd @**unal.edu.co**

***Abstract*— Adequate recognition of lips posture for speech articulation analysis requires of the measurement of several anthropometric mouth parameters. These are needed to estimate the position and contour of the lips and teeth and tongue positions, as well. Here, a method is proposed for lips contour detection under natural conditions without any extra hardware requirements for image acquisition. The purpose of the suggested process is to obtain the lips contour based on red hue fields detection. Afterward, geometrical features of lips are extracted from their detected contour. Image processing is divided into the following steps: Face and mouth search, lips contour detection and feature estimation from lips geometry. Due to high dimensionality of the initial feature space, it is very important to evaluate the performance of the lips features regarding their ability to discriminate pathological lip postures in the case of children with cleft lip and palate. In this paper, a proposed method for effective selection of the image feature set was developed using multivariate analysis techniques. Finally, the discriminant performance of the selected training sets was evaluated using bayesian estimators. Results of the comparison for different common testing algorithms show that the proposed processing method exposes better performance.**

I. Introduction

Surgical correction of cleft lip and palate (*CLP*) is often achieved after several interventions. Once the abnormality has been corrected, it is very important to determine the quality of the speech articulation. In this way, image processing techniques can be used for automated lip detection and tracking of the child's articulation. However, an adequate recognition of lip postures for speech articulation analysis requires the measurement of several anthropometric parameters of the mouth to estimate the position and contour of lips and teeth as well as tongue position.

Many approaches have been proposed for lip detection, some of them based on gray-level analysis [1], others on color analysis; template models founded on dynamic contours, active shape models [2], de-formable templates [3]. In [4], an algorithm for speaker's lip segmentation and feature extraction is presented, where a color videosequence of the speaker's face is acquired, mounting a special micro-camera, under natural lighting conditions and without any particular make-up. Nevertheless, all referenced works are oriented to lip segmentation, no considerations about lip contour detection are taken.

Here, a method is proposed for lip contour detection with no extra hardware requirements for image acquisition. The aim of this process is to obtain the lip contour. Then, lip geometrical features can be extracted.

Image processing is divided into the following stages: face and mouth search [5]; lips contour detection; and feature estimation from lips geometry. Lip posture analysis is provided to distinguish between normal and pathological phoneme articulations uttered by a *CLP* child. The analysis involves the 5 vowel phonemes of the Spanish language. As a result of the suggested algorithm for lip contour detection, 57 lip posture features are obtained for each phoneme. This situation leads to a large dimensionality feature space which is a regular issue linked to classification tasks. One approach to overcome this problem is effective feature selection, which reduces the dimensionality of the input feature space, such that the discriminant performance of the classifier can't be influenced. For this purpose, the use of multivariate analysis techniques lets assess the feature system reliability using several criteria like statistical independence or information redundancy, among others.

In this paper, a method for effective selection of the image feature set was also developed accomplishing multivariate analysis techniques. Finally, the discriminant performance of the selected training sets was evaluated using bayesian estimators.

II. Lips Detection

Lips processing starts detecting the face within the original image. A procedure called skin detector (SD) is used for such a purpose [6]. Once, the face coordinates have been defined, the lips hue is emphasized using the proposed technique in [3].

By means of connectivity analysis the mouth region is identified, therefore detection of external lip contour can be carried out. In this case, two different techniques are considered: lip hue enhancement described in [3], and red exclusion technique based on green and blue component analysis [7].

A. Finding the Mouth

Once, the face is located, the detection of mouth region is carried out according to the proposed procedure of *detection by prevalent regions analysis* (DPRA), as follows:

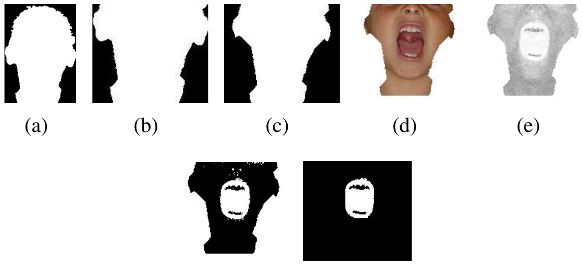
1) Input search space is constrained to the area related to FS (Fig. 1(a))
2) Upper image area of search space is clipped to a 1/3 of FS height (Fig. 1(b)).
3) Reduced picture is eroded to reduce any remaining background disturbance after segmentation procedure [8] (Fig. 1(c)).
4) Color information is retrieved, but only inside the FS space (Fig. 1(d)).
5) HSV transformation of a color image is followed by hue filtration according to [3] (Fig.1(e)).
6) A hard threshold procedure of hue-filtered image is executed, therefore, mouth area gets the stronger as possible contrast in comparison to neighbor face elements resulting in a binary picture. (Fig. 1(f)).
7) Selection of more prevalent region (mouth area) is carried out (Fig. 1(g)), by means of connectivity analysis.

The resulting mouth region describes a useful image space for feature extraction oriented to recognition of vowel phonemes.

III. Feature extraction

During treatment of *CLP* children, lips posture detection is accomplished in order to determine the quality of speech articulation. There is a concrete set of visual features known by therapists that describes lip postures for all Spanish vowel phonemes [9]. Because of non invasive diagnosis is performed, some of the features related to soft palate region are very hard to estimate by external image processing. So that, the present work only comprises a set of external features of lips and mouth contour.

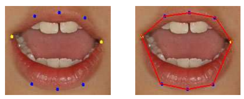
1-4244-0033-3/06/$20.00 ©2006 IEEE.



(a) (b) (c) (d) (e)

(f) (g)

Fig. 1. Mouth Detection.



(a) (b)

Fig. 2. 8-point Model for External Lips Contour.

A. Approximation of external lips contour

Suggested contour approximation is achieved based on the detection of 8-point polygonal set which is considered enough to describe the external border of lips shape. At the beginning, over a whole mouth region a couple of outermost extreme pixels (called vertices) is detected (yellow points in the Fig. 2(a)). Both extreme vertices became the initial reference for searching the remaining 6-point lips contour (blue points in Fig. 2(a)). After estimating an 8-point set of coordinates, each couple of consecutive points is connected drawing a line between them, as shown in the Fig. 2(b). The resulting polygon turns out to be the model for the feature extraction used for lips posture recognition.

B. Vertices Location

The points where both upper and lower lips edges concatenate are defined as vertices. To estimate vertices coordinates, two techniques are considered: Vertical Gradient (*VG*), proposed in [3] and a technique based on *Searching Space Reduction After Segmentation (SSRAS)* which is described as follows:

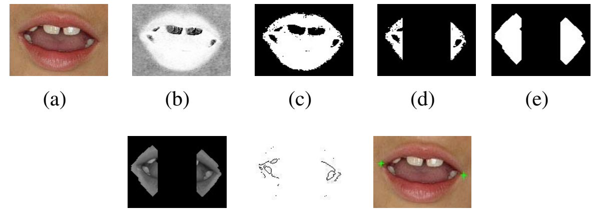
1) Input color image (Fig. 3(a)) is mapped to a HSV space, but filtering the hue component (Fig. 3(b)).
2) Grey scale image is mapped to a new binary picture having a suggested threshold of $240/255$ (Fig. 3(c)).
3) Analysis of spatial distribution of white pixels is performed in order to get estimates of mean (m_w) and standard deviation (σ_w).
4) Boundaries for searching space is defined as follows: ($\pm$ 1.5 – 2.8 σ_w) (Fig. 3(d)).
5) Resulting bounded area is expanded to increase the searching space [8] (Fig. 3(e)).
6) Original picture values are retrieved, but in grey scale within the bounded area for searching space (Fig. 3(f)).
7) Edge detection is achieved using a SUSAN operator [10] (Fig. 3(g)).
8) Vertical location values of vertices are obtained computing mean values of pixel distribution inside each of SUSAN image bands. Horizontal coordinate values for vertices are calculated from location of outermost extreme pixels inside each of the bands (Fig. 3(h)).

C. Finding Lips Geometry

Points of lips geometry are located according to procedures described in [3].

D. Initial Feature Set for Lips Posture Detection

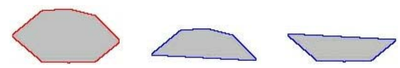
Based on the detection of lips geometry, a set of descriptors (features that are invariant to operations of rotation and scaling [8]) is proposed. Descriptors are calculated using a function $f(x, y)$ which is defined as the distance from the mass center of the given region to each one of the contour pixels. Estimation of descriptors is done for regions showed in Fig. 4: region 1 is the external contour 4(a), Region 2 is upper lip 4(b), finally region 3 is the lower lip 4(c).



(a) (b) (c) (d) (e)

(f) (g) (h)

Fig. 3. Detection of Mouth Vertices.



(a) Region 1. (b) Region 2. (c) Region 3.

Fig. 4. Regions for Descriptor Estimates.

Table I shows a set of suggested descriptors for lip postures detection where d stands for region density, m_{pq} are pq order bidimensional moments and ϕ_n are n invariant moments [8]. μ_f and σ_f are the mean value and standard deviation of $f(x, y)$ function values. As a result, a total set of 57 features is obtained for lip posture detection.

E. Effective Feature Selection

In order to reduce the computacional cost, and increment the system precision [11], it becomes necessary to eliminate the irrelevant, redundant and erratic features.

The presented method of effective feature selection is developed according to the following restrictions: *discriminant capacity, reliability and absence of correlation.*

The method is based on multivariate and univariate linear statistical techniques (named: MANOVA, PCA and ANOVA) and the main goal is to find an effective feature set that evaluated as a whole may lead to enough discriminant level between classes in order to obtain the suitable automatic classification. The smaller probability error is chosen as validation criteria [12], [13].

Proposed method involves the following steps (Fig. 5) [13]:

1) MANOVA: This technique implies extracting the initial features hyperspace as a subspace having a highly discriminant level as a whole, but not necessarily in independent way, as it happens if using ANOVA or the hypothesis test, among others. MANOVA

Notation	Feature	Notation	Feature	Notation	Feature
1	d	8	m_{03}	15	ϕ_5
2	m_{01}	9	m_{12}	16	ϕ_6
3	m_{10}	10	m_{21}	17	ϕ_7
4	m_{11}	11	ϕ_1	18	μ_f
5	m_{20}	12	ϕ_2	19	σ_f
6	m_{02}	13	ϕ_3		
7	m_{30}	14	ϕ_4		

TABLE I
FEATURE SET.

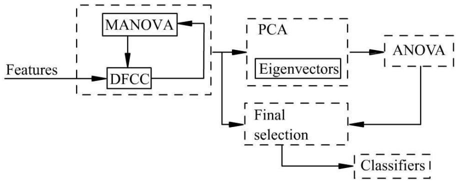


Fig. 5. Feature selection - Flow chart

procedure is accomplished according to *Wilks*' test transforming the Λ's into F distribution values (Λ to $F - statistics$) [14].

MANOVA provides just an index of the discriminant capacity for a given input hyperspace, but no information about which variables are relevant variables. This implies developing of a search procedure to find which of the feature subsets exposes the highest discriminant values.

Searching procedure is called *Decision by Flow of Increasing Set (DFIS)* algorithm:

1) Calculate the F-statistic (transformation of Wilks' Λ) for one-feature subsets. From these values, the feature with the largest F-statistic is chosen and its cumulative probability value is computed using the F distribution.
2) Construct 2-dimensional subsets, combining the feature previously chosen in step 1 with the remaining features. Each one of these subsets is evaluated through the Wilks' test, and its respective F-statistic is updated.
3) Select the 2-dimensional subset with the largest F-statistic. Calculate its respective cumulative probability using the F-distribution. This value must exceed the calculated value in step 1, to update the subset. Otherwise, terminate the search.
4) Construct the feature subset adding one feature to the updated subset. These new analysis groups correspond to the subset selected in step 3 and each one remaining feature.
5) Return to step 3 and update the subset using the same criteria. Continue to step 4 and repeat the updating process over and over. The algorithm stops as shown in step 3 when added features do not increases the cumulative probability. In this manner, we can select those features that joint are more discriminant.

Note that if during the calculations $\Lambda \to 0/0$, there is a linear dependency in the current subset. Then its F-statistic is forced to zero and so the evaluated subset is rejected.

2) PCA: This technique is oriented to minimization of the correlation between principal components. PCA technique is preferred among others reduction procedures, because of its implementation easiness [15].

3) ANOVA: The target of the univariate analysis is to rescue discriminant projections taking advantage of the property of statistical independence among the principal components.

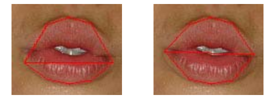
Although this method is provided to reduce a given PCA space, this procedure is achieved in a different hyperspace. So that, it is necessary to make a post-analysis of simplification using the resulting hyperspace after MANOVA analysis.

4) Final Feature Set: Because feature set obtained after ANOVA analysis does not match the original feature space, a suggested procedure for transformation to original feature space is described in [13].

IV. RESULTS

A. Data base

Image acquisition was restricted to a child stood in frontal position, pronouncing one vowel phoneme. Captured color images ($1320/phoneme$) have 2560×1920 dimension in JPEG standard.



(a) *VG*. (b) *SSRAS*.

Fig. 6. Vertices Location.

(a) */a/*. (b) */e/*. (c) */i/*. (d) */o/*. (e) */u/*.

Fig. 7. Contour detection for one *CLP* patient articulating the five vowel phonemes of Spanish language.

B. Mouth Location

DRPA method is evaluated by an expert having as performance measure the function $f_{DM}(x) = 1$ when the whole mouth is detected. Otherwise $f_{DM}(x) = 0$, being x is the resulting image. The algorithm successfully detected 6274 of 6600 images (95.06%) with an average processing time of 0.73716 sec per image.

C. Feature Extraction

Testing of the point location algorithm was carried out by the following techniques: location of vertices (Vt) using the VG procedure; location of Vt using the SSRAS procedure; location of the contour points using the Hue filtered Image; and location of contour points using algorithm of red exclusion.

Concerning to time process, comparison between *SSRAS* and VG procedures showed no big difference (less than 1%). However, this was not the trend in the case of detection of lips geometry. Fig 6 shows an example of vertices location for both procedures. The better performance of proposed the SSRAS procedure is evident.

Results of the comparison of the testing algorithms for point location showed that Hue filtered process exhibits better performance.

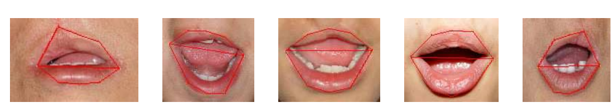
D. Tests on CLP patients after surgical repair

The paramount aim of this work, is the application in the treatment of children with *CLP*. According to this, test for detecting external lip contour were made over samples of children with this pathology. Figures. 7 and 8 depict some of the results.

The Images shown in 8 were taken under different light conditions, which leads to consider the algorithms could perform well disregarding the acquisition setup.

E. Feature selection Set

Table II shows obtained figures to test the proposed method of effective feature selection. Results were obtained for detection of lip postures of the 5 Spanish vowels. As it can be seen on table II resulting sets of selected features turned out to be different depending on pair given phoneme clases that were being compared (Subscript number indicates the region of which was extracted the feature, according to shown in the fig 4).



(a) (b) (c) (d) (e)

Fig. 8. Contour detection for several *LPHC* patients under different acquisition environments.

	/a/	/e/	/i/	/o/	/u/
/a/		11_1,11_2	11_2,12_3	2_2,18_2	9_3,18_3
/e/	11_1,11_2		6_1,12_2	3_2,18_2	7_2,18_2
/i/	11_2,12_3	6_1,12_2		3_2,18_2	9_2,18_2
/o/	2_2,18_2	3_2,18_2	3_2,18_2		3_3,11_2
/u/	9_3,18_3	7_2,18_2	9_2,18_2	11_2,3_3	

TABLE II

SET OF SELECTED FEATURES. NUMBER CORRESPOND TO THE NOTATION OF THE FEATURE SHOWED IN TABLE I AND SUBINDEX DENOTES THE REGIONS SHOWED IN FIG. 4.

The final set of effective selected features for recognition task as a whole was assembled by gathering all primary sets showed on table II. That means a total of 13 features were finally selected for lip postures recognition of Spanish vowels during treatment of CLP children. As a result a factor of 4.4 was obtained for dimensional reduction of initial feature space.

F. Classification Testing

Multiple classification task is accomplished through one-to-one algorithm. Performance of classifiers is evaluated using the criteria of minimal error probability [16].

Test database was divided into two groups for classification of all given classes of lips postures: first group of 291 samples (70%) was used for classifier training, while remaining 30% of samples was used for validation.

Classification performance for obtained set of effective features showed a global error of 32% in case of linear classifiers, and a global error of 30% for bayesian classifiers. Testing of initial feature set (described in section III-D) could not be achieved because of the lacking of computational requirements which a way far excedes the available memory resources.

V. CONCLUSIONS AND FUTURE WORK

A method is proposed for lip contour detection under natural conditions, and no extra hardware requirements for image acquisition. The purpose of the suggested process is to obtain lip contour based on red hue fields detection. Then, lip geometrical features are extracted from the detected lip contour. Results of comparison of different testing algorithms showed that proposed processing method reaches better performance.

A proposed effective feature selection based on the use of multivariate analysis techniques let assess the feature system reliability, using several criteria like statistical independence or informative redundancy, among others. This method allows to reduce computational cost without dropping the efficiency of classifiers. Although, the reached performance of proposed method is still very low, testing of initial feature set could not be achieved because of the computational requirements.

Future tests will be carried out for detecting morphological differences between normal and surgical repaired *CLP* children.

REFERENCES

[1] Mersereau R. Rao, R. Lip modelling for visual speech recognition. *28th Annual Asimular Conference on Signals, Systems, and Computer, IEEE Computer Society*, 2, 1994.

[2] J. Luettin. *Visual Speech and Speaker Recognition*. PhD thesis, Department of Computer Science. University of Sheffield, 1997.

[3] Torresani L. Caprile B. Coianiz, T. 2d deformable models for visual speech analysis. *Istituto per la Ricerca Scientifica e Tecnologica*, 1996.

[4] Luthon F. Liévin, M. Lip features automatic extraction. 1998.

[5] Venaktesh K. Stork D. Hennecke, M. Automatic speech recognition system using acoustic and visual signals.

[6] Abdel-Mottaleb M. Jain A. Hsu, R. Face detection in color images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24, 2002.

[7] Powers-D.M.W. Lewis, T.W. Audio visual speech recognition using red exclusion ans neural networks. *School of Informatics and Engineering*, 2003.

[8] Woods-R. González, R. *Tratamiento Digital de Imágenes.* Addison-Wesley Iberoamerica, S.A., 1996.

[9] T. Corredera. *Defectos en la Dicción Infantil. Procedimientos para su Corrección*. Editorial Kapelusz, 1949.

[10] S.M. Smith and J.M. Brady. SUSAN - a new approach to low level image processing. *Int. Journal of Computer Vision*, 23(1):45–78, May 1997.

[11] J. Carrasco. Reconocimiento de patrones. Technical report, Instituto Nacional de Astrofísica Óptica y Electrónica, México, 2003.

[12] M. A Carreira. A Review of Dimension Reduction Techniques. Technical Report CS–96–09, Dept. of Computer Science, University of Sheffield, January 1997.

[13] G. Daza L. G. Sánchez J. F. Suárez y G. Castellanos. Métodos multivariados para la selección efectiva de características en casos de clasificación múltiple. Technical report, Grupo de Procesamiento Digital de Señales. Universidad Nacional de Colombia, 2004.

[14] Johnson Thie. and Gary McDarby. Estadística. *School of Electrical Eng. and Telecommunications. Sydney Australia*, 2002.

[15] I. T. Jollieffe. *Principal Component Analysis*. Springer series in statistics, NYC, NY, 2002.

[16] R. O. Duda, P. E Hart, and D. G. Stork. *Pattern Classification*. Wiley-Interscience, New York, second edition, 2001.