Robust Glottal Gap tracking based on Temporal Intensity Variation

Gustavo Xavier Andrade Miranda
Juan Ignacio Godino Llorente
gxandrade@ics.upm.es

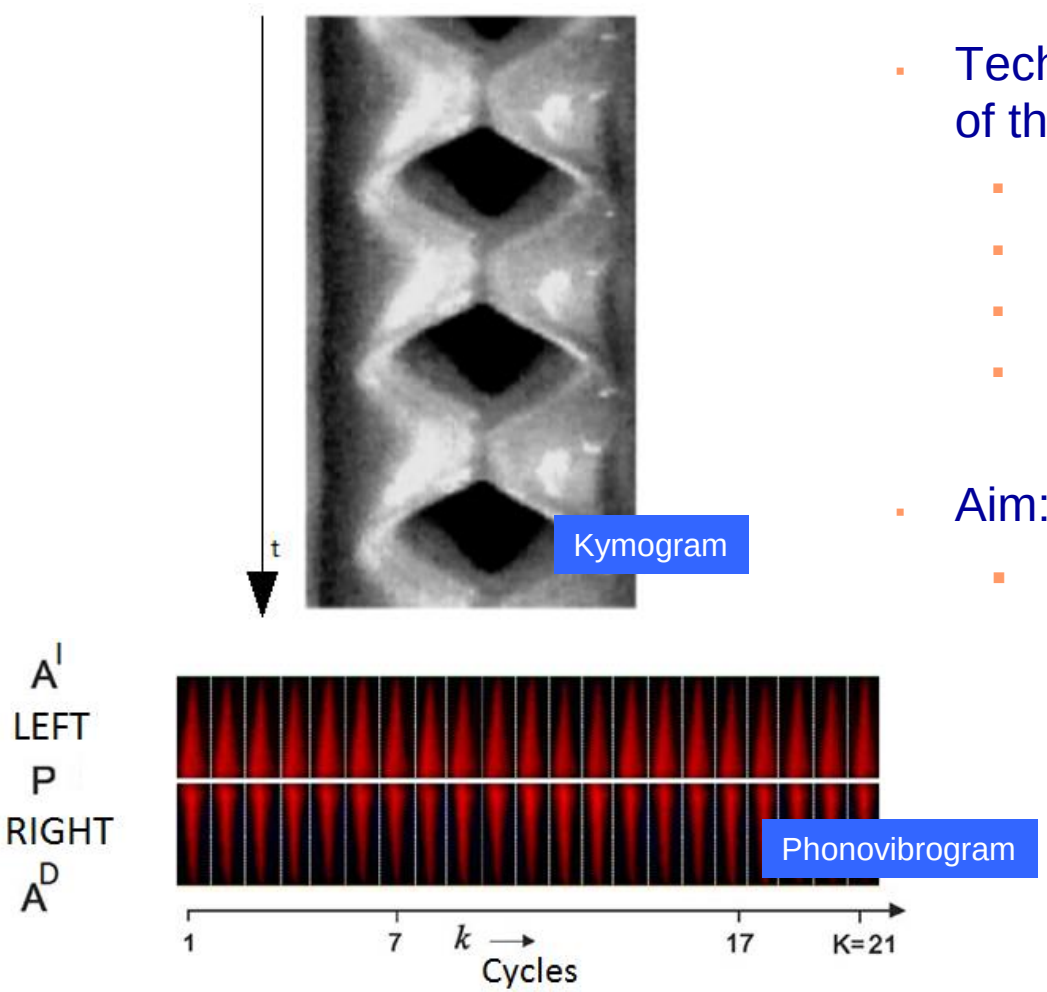
Dept. ICS, Universidad Politécnica de Madrid, Spain

Table of Content

Motivation (I)

- The detection of the glottal area is the first step and the most challenging in the assessment of the vocal fold vibration.
- It is possible to extract some simple measurements such as:
  - the ratio of vibratory amplitude
  - ratio of periods of vibration, etc.

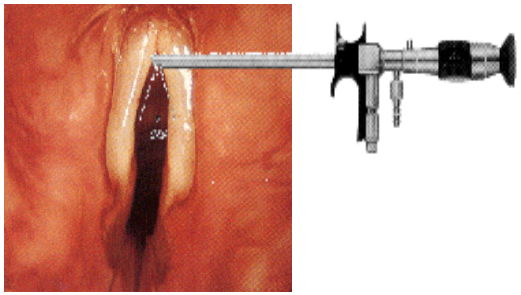
Motivation (II)



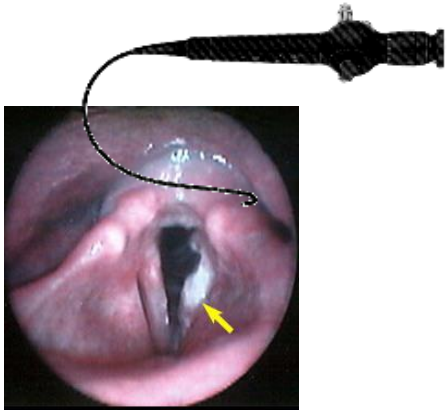
- Techniques used for a detailed analysis of the vibration of the vocal folds:
  - Glottal Vibration Profiles (GVP) ,
  - Glottal Area Waveforms (GAW),
  - Digital Kymograms (DKG),
  - Phonovibrograms (PVG).
- Aim:
  - Combining techniques based on temporal information with segmentation algorithm like active contours and watershed.

Methods for the larynx exploration

- Recording the movements of the vocal folds is impossible using a standard video cameras (25-30 frames/sec.)



- Slow Motion Stroboscopy (SMS)
  - Fibroscopy (or videofibroscopy).
  - Telelaringoscopy (or videotelelaringoscopy).



- **High-Speed Digital Images (HSDI)**
  - **High Speed Cinematography.**
  - Videokymography.

Segmentation Algorithm

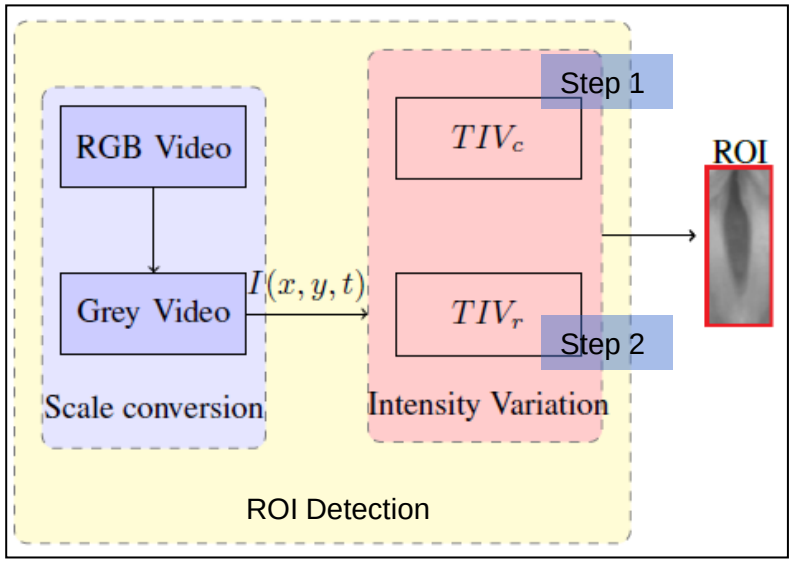
The literature reports different methods for segmenting the glottal gap:

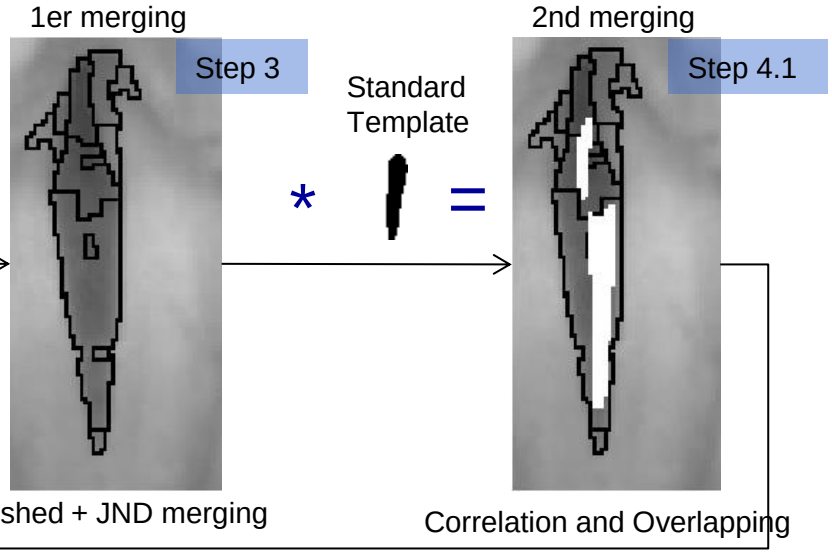
- Thresholding and Histograms [**mehta01**]
- Region growing [**Lohscheller07**]
- Watershed transform [**Osma-Ruíz08**]
- Actives contours (snakes) [**Marendic01**]

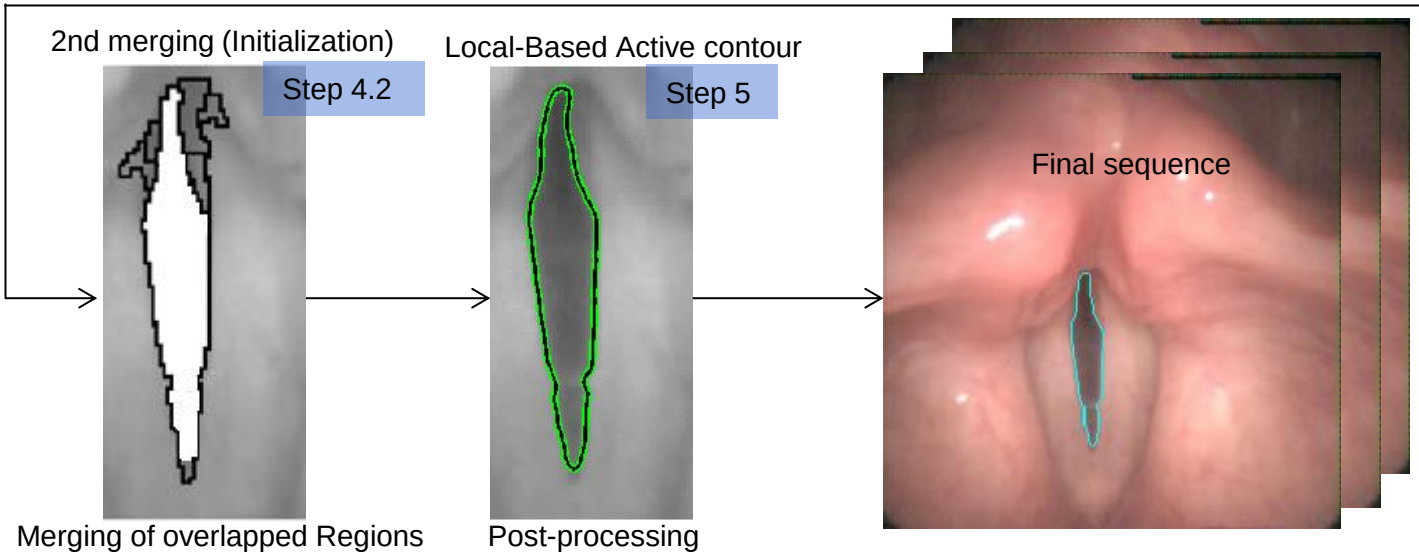
Most of the segmentation algorithms used in the literature **do not take into account the temporal dimension, so each frame is treated individually** leaving aside the information obtained from the previous frames

→ **Tracking**

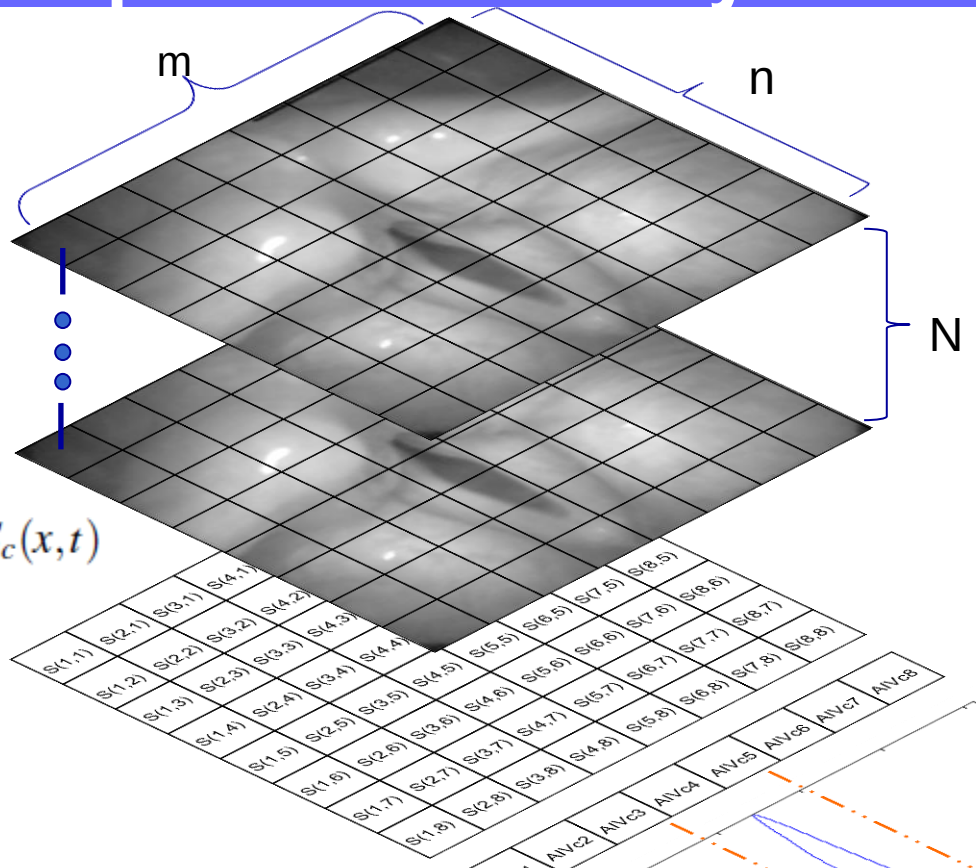
Scheme of the Proposed Method

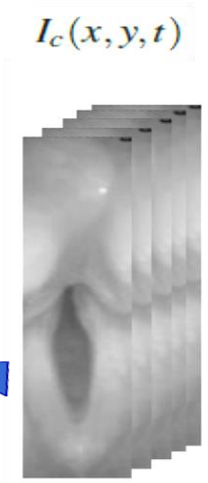






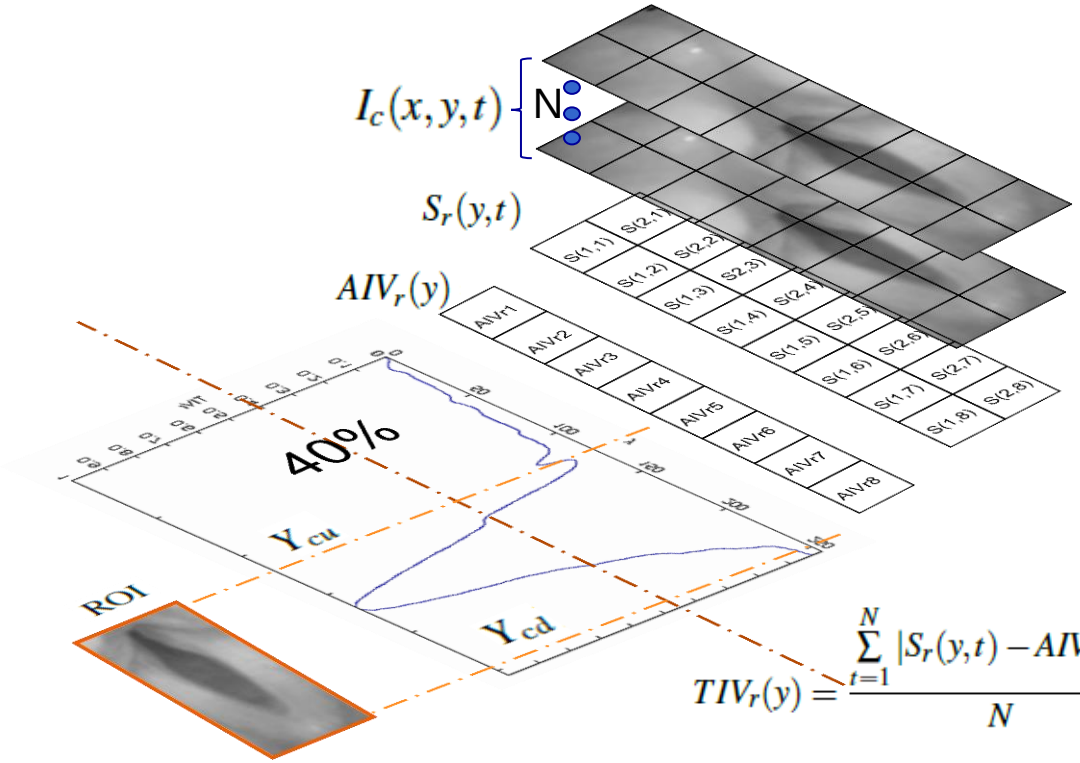
Step1:Total Intensity Variation in Columns (TIVc)





$$TIV_c(x) = \frac{\sum_{t=1}^{N} |S_c(x,t) - AIV_c(x)|}{N} \quad \forall x, 1 \le x \le m, \qquad |x_i - x_{i-1}| \le 0.009 \qquad \forall i \in x \mid TIV_c(i) < 0.4;$$

Step2:Total Intensity Variation in Rows (TIVr)

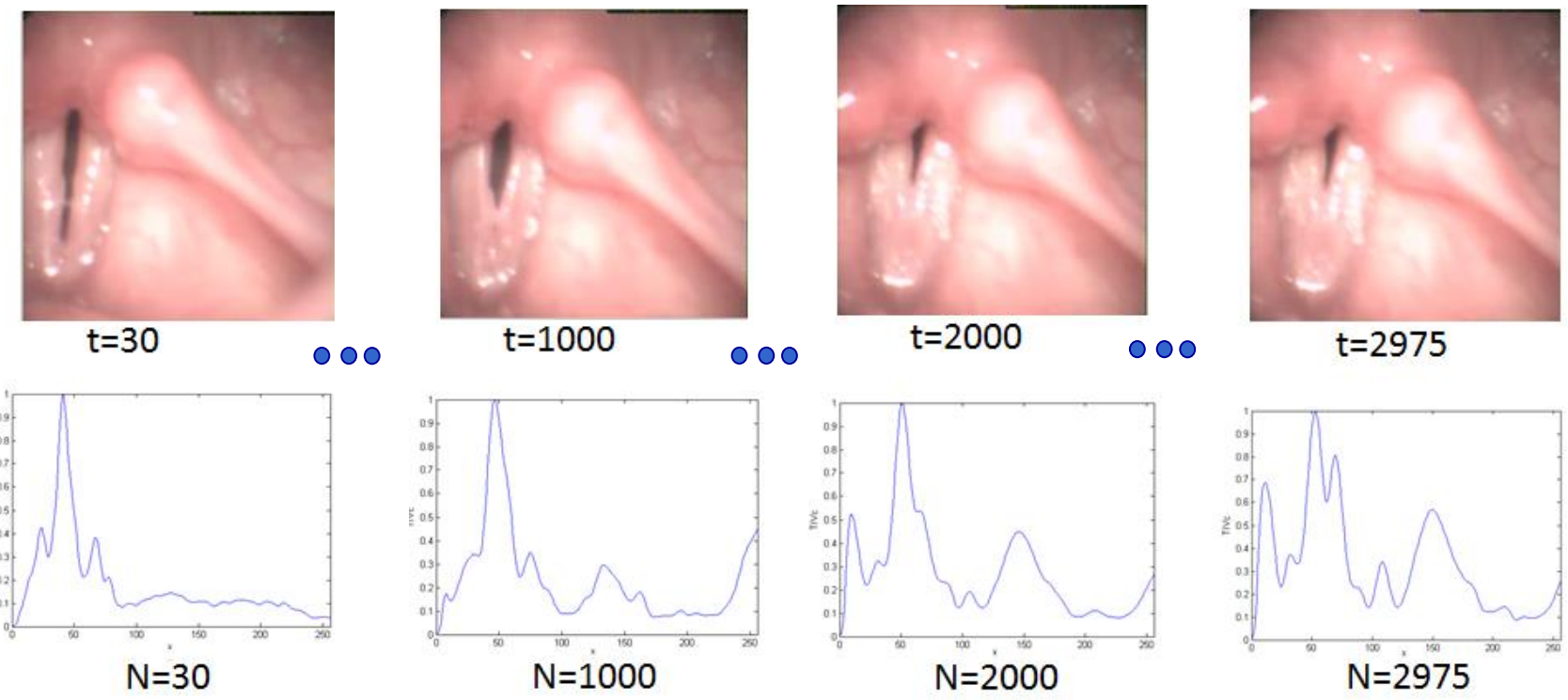


$$TIV_r(y) = \frac{\sum_{t=1}^{N} |S_r(y,t) - AIV_r(y)|}{N} \quad \forall y, 1 \le y \le n;$$

$$|y_i - y_{i-1}| \le 0.009 \qquad \forall i \in y \mid TIV_r(i) < 0.4;$$

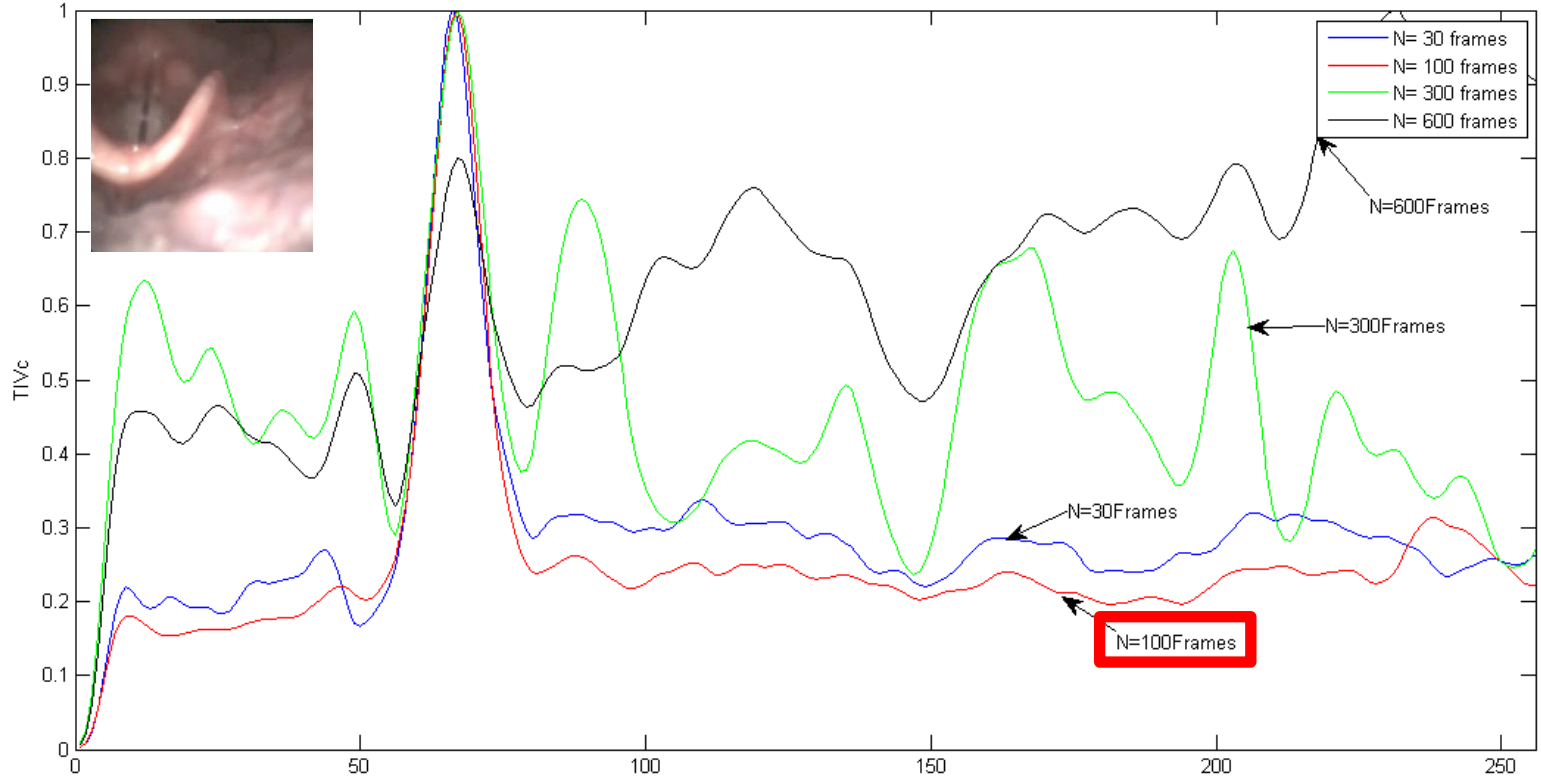
Determination of the optimal value of N (I)

- The selection of N is the most important for the correct ROI detection.
- The ROI has to tolerate the camera displacements



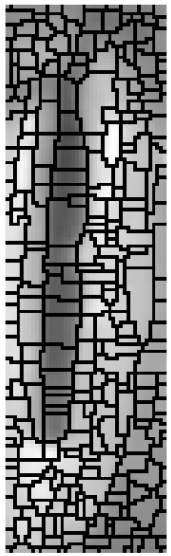
Determination of the optimal value of N (II)

- The choice of N was performed considering only the TIVc.
- N was chosen through experimentation observing the entire database (18 HSDI).
- The minimum requirement to achieve a robust ROI is choose a N that contain at least one complete glottal cycle.

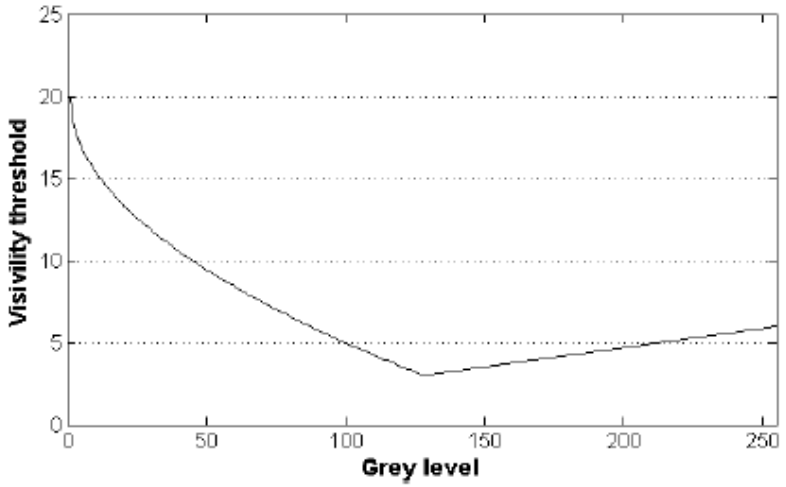


Step3: Watershed and First Region Merging

Watershed



$$JND(k) = \begin{cases} T_0 = [1-(k/127)^{0.5}]+3 & k \le 127 \\ \gamma(k-127)+3 & \text{otherwise} \end{cases}$$



Merging

$$F_c = [|mR_1 - mR_2| - MinJND(mR1, mR2) + 255]$$

[Osma-Ruíz08]

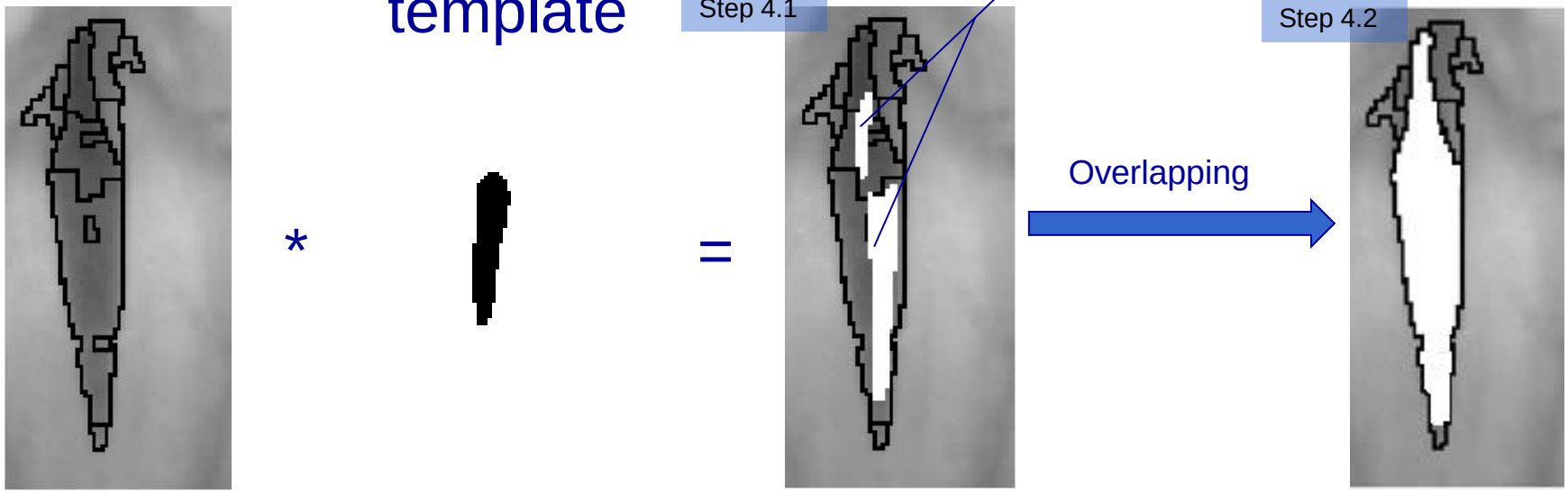
Step4: Correlation Regions Merging

- Obtained empirically based on manual segmentations.
- Composed by a white background and a black foreground.
- Size of the template is 12x42 pixels.
- Glottis-like shape.

1st Merging

Standard template

NCC Threshold matching = 0,45

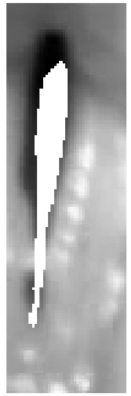


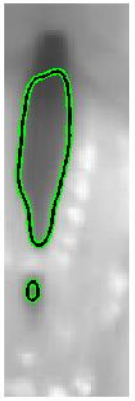
Step5: Localizing region-based active contours

$$E(\phi) = \int_{\Omega_x} \delta\phi(x) \left(\int_{\Omega_y} B(x,y) F(I(y), \phi(y)) dy \right) dx + \lambda \int_{\Omega_x} \delta\phi(x) \|\nabla\phi(x)\| dx$$ **[Lankton08]**

Initialization 100 iterations

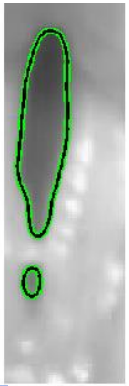
Image





Standard template

- Initialization based on standard template.
- Need more than 100 iterations to converge.
- Searching radio should be greater than 5 pixels.



Initialization using correlation

- Initialization based on correlation.
- Need less than 100 iterations to converge.
- Reliability when the glottis is divided.
- Searching radio is 5 pixels

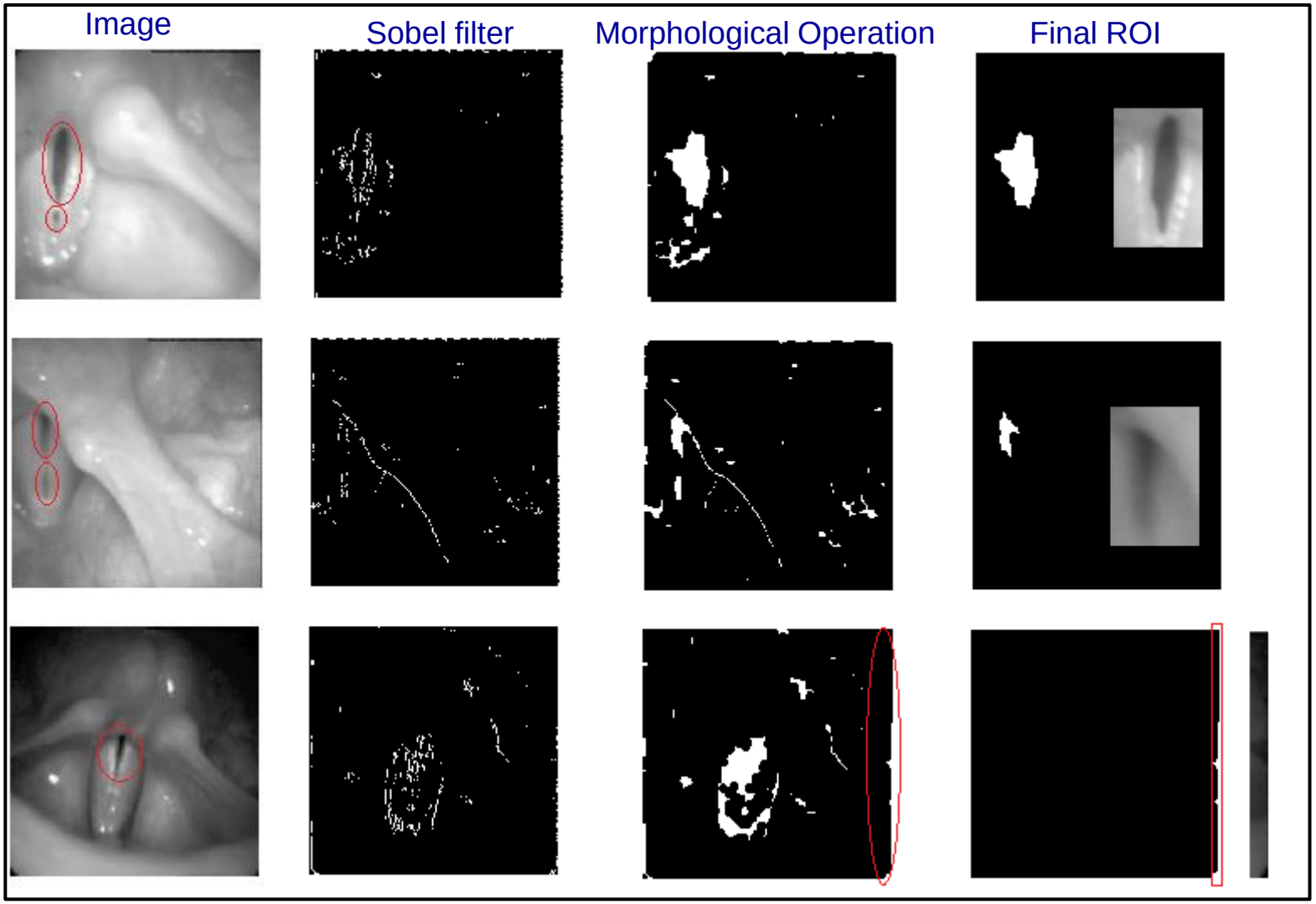
Results (I)

- The Database is composed by 18 HSDI sequences of resolution 256x256 pixels and the sampling rate is 4000 frames/seconds.

- All videos chosen have recorded under different conditions;
  - Different illumination levels
  - contrast problems
  - Presence of nodules
  - Partial occlusion of the glottis
  - Lateral displacements of the camera

- Three trials were develop:
  - Compare our ROI methodology with the proposed in **[Karak12]**.
  - Compare the results obtained by two manual segmentations with the method proposed using the Pratt index **[Pratt79]**.
  - A visual inspection of the PVG obtained by manuals and automatic segmentation.

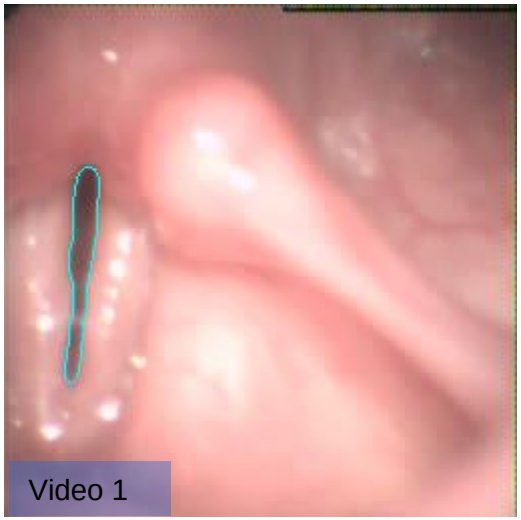
Results (II)

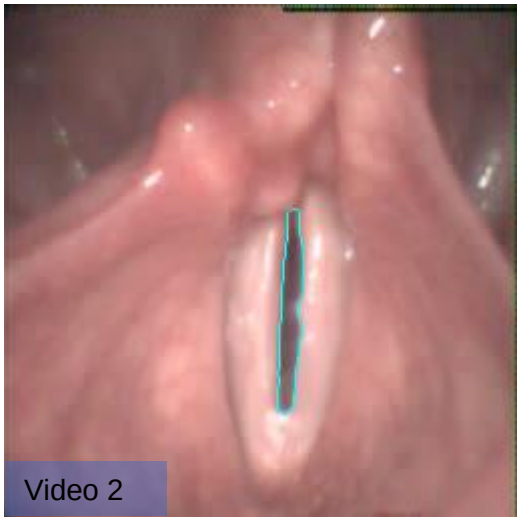


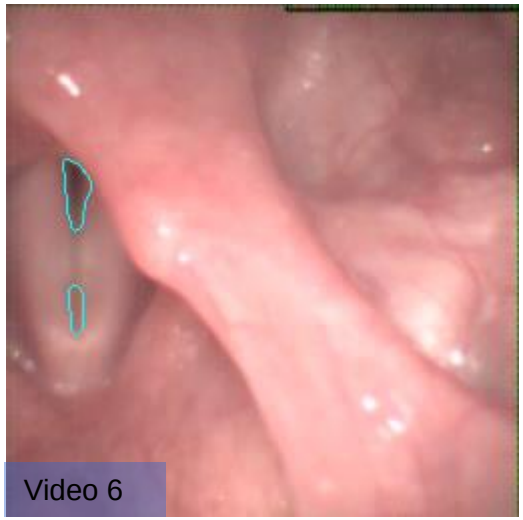
Method based
On TIVc and TIVr

Method proposed in [karak12] for ROI Detection

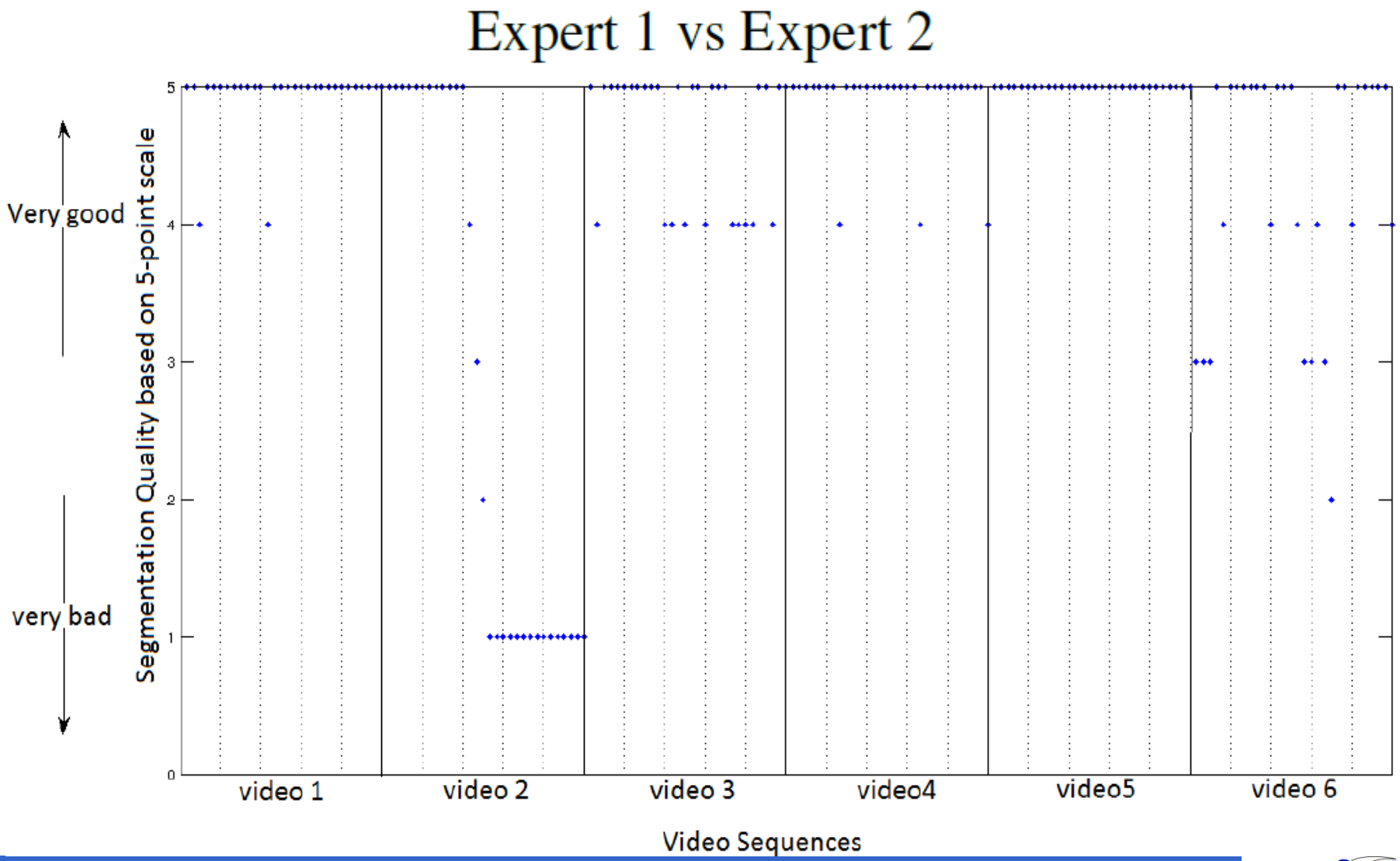
Results (III)



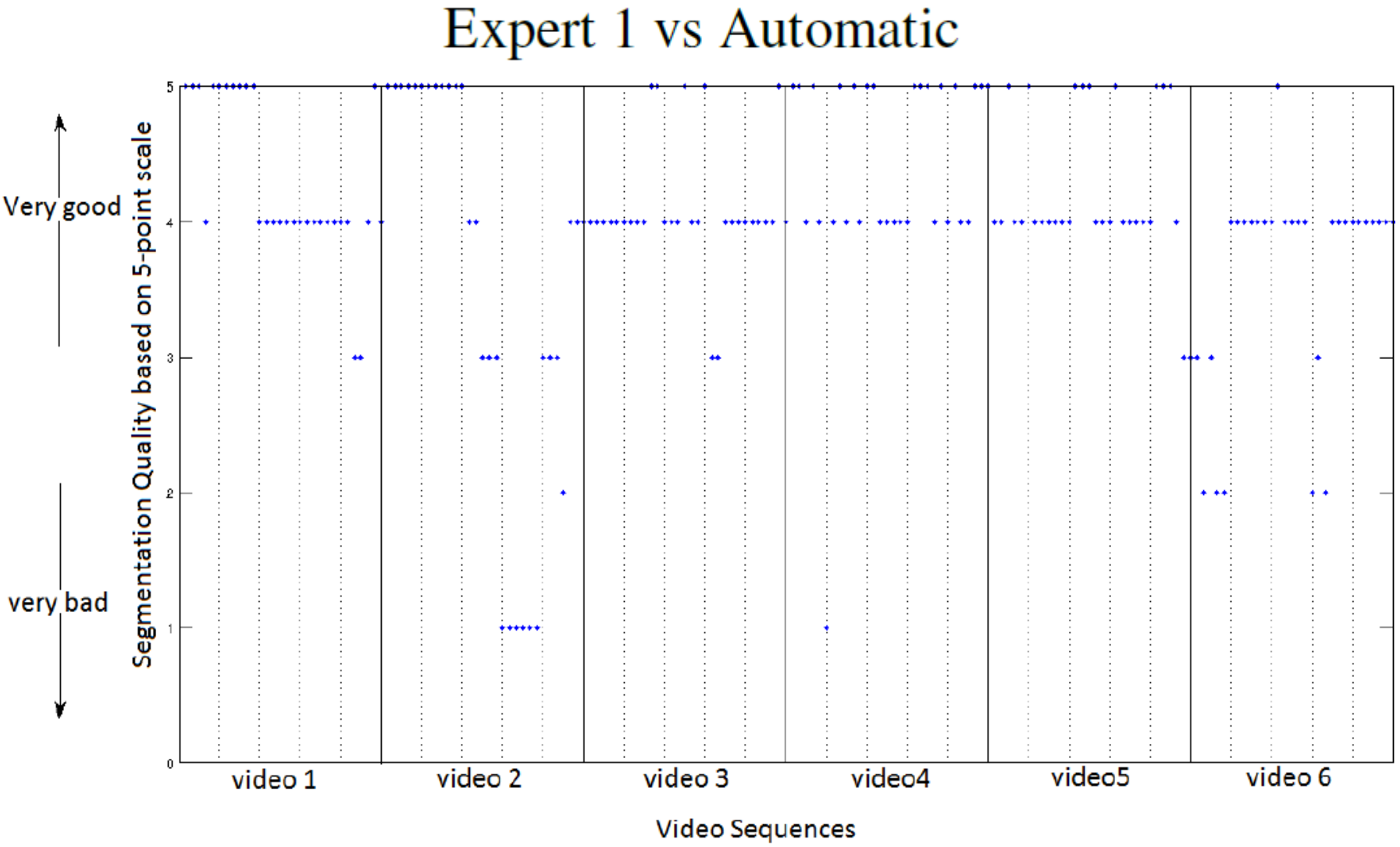




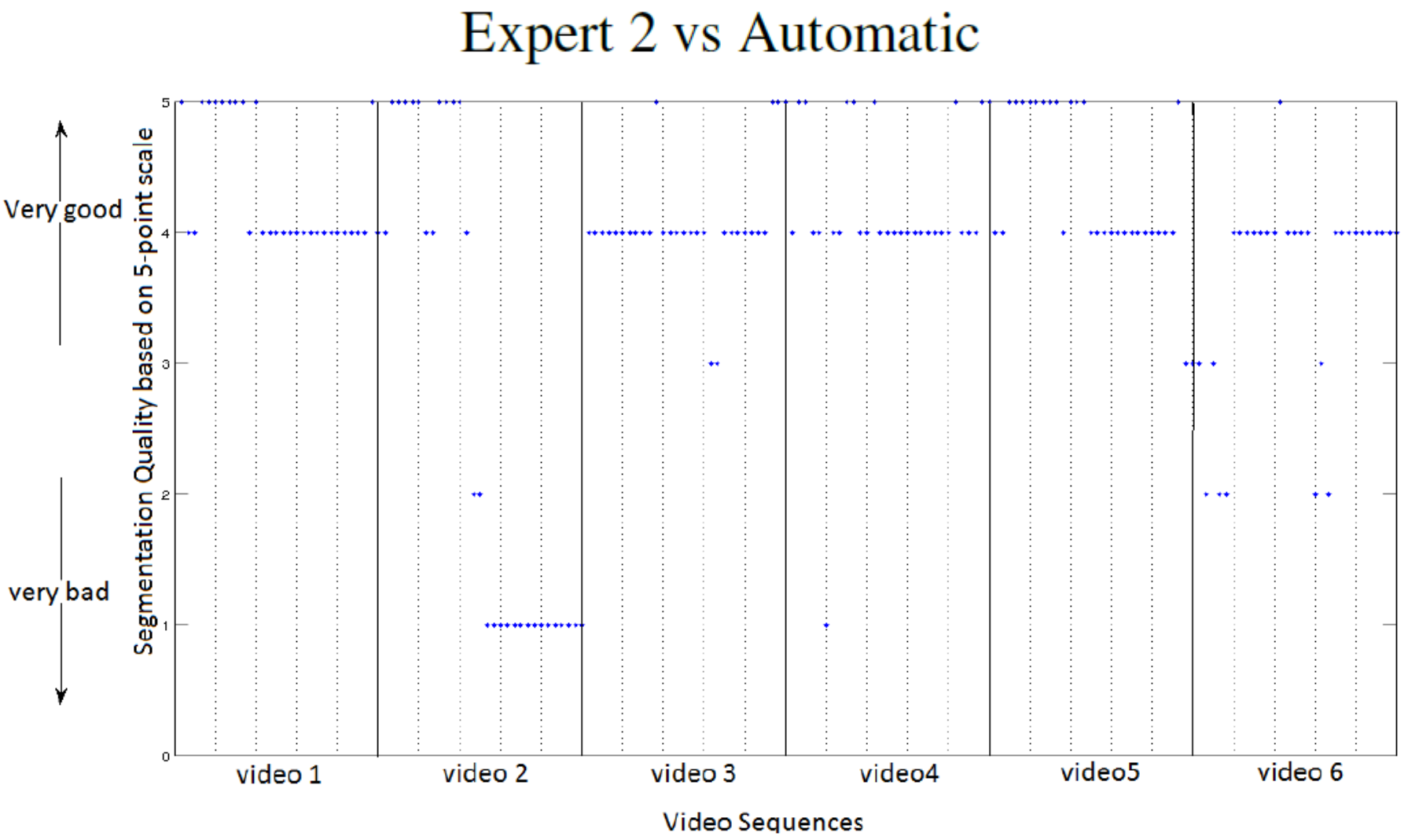
Results (III)


Expert 1 vs Expert 2
Very good
very bad
Segmentation Quality based on 5-point scale
5
4
3
2
1
0
video 1
video 2
video 3
video4
video5
video 6
Video Sequences

Results (III)

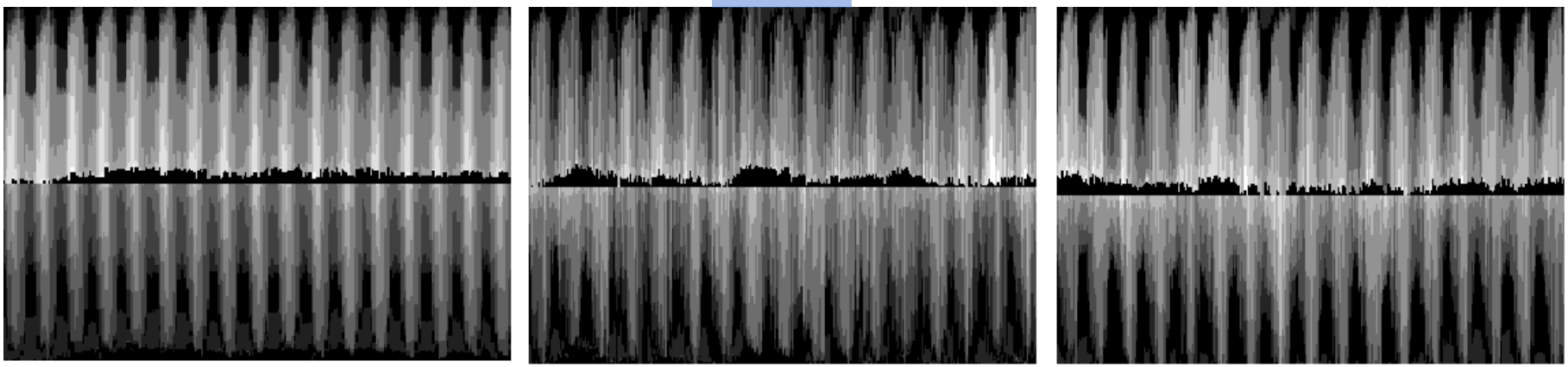

Expert 1 vs Automatic
Very good
very bad
Segmentation Quality based on 5-point scale
5
4
3
2
1
0
video 1
video 2
video 3
video4
video5
video 6
Video Sequences

Results (III)


Expert 2 vs Automatic
Very good
very bad
Segmentation Quality based on 5-point scale
5
4
3
2
1
0
video 1
video 2
video 3
video4
video5
video 6
Video Sequences

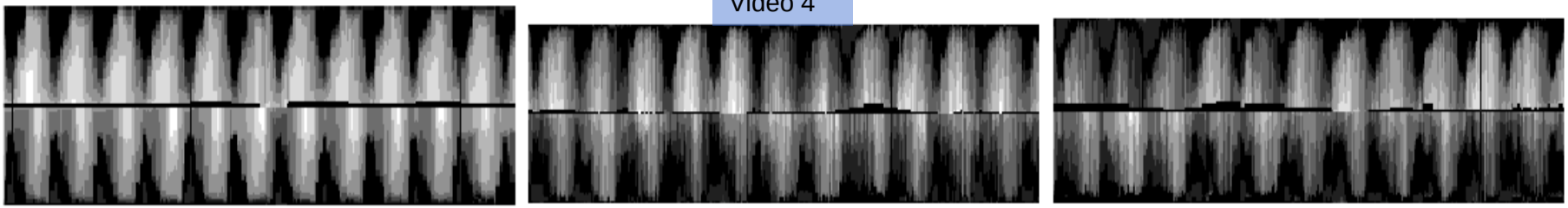
Results (IV)

Video 3



PVG Automatic PVG Expert 1 PVG Expert2

Video 4



PVG Automatic PVG Expert 1 PVG Expert2

Conclusions

- The main motivation of this paper was propose a complete framework for the assess the glottis tracking.
- The mechanism proposed for detect the ROI adjust better in the demanding cases like;
  - Glottis with partial occlusion
  - Camera movements
  - Depth differences between videos recording
  - Reliability against external artifacts
  - Glottis divide in two parts
- We have proposed a complete methodology, combining traditional and new techniques in image processing with temporal intensity variation for the ROI detection.
- Results obtained compared with manual segmentations are similar

References

- Mehta, D.D., Deliyski, D.D., Quatieri, T.F., Hillman, R.E.: Automated measurement of vocal fold vibratory asymmetry from high-speed videoendoscopy recordings. Speech, Language and Hearing Research 54(1), 47 – 54 (2011) **[Mehta01]**
- Lohscheller, J., Toy, H., Rosanowski, F., Eysholdt, U., Dollinger, M.: Clinically evaluated procedure for the reconstruction of vocal fold vibrations from endoscopic digital high-speed videos. Medical Image Analysis 11(4), 400 – 413 (2007) **[Lohscheller07]**
- Osma-Ruíz, V., Godino-Llorente, J.I., Sáenz-Lechón, N., Fraile, R.: Segmentation of the glottal space from laryngeal images using the watershed transform. Computerized Medical Imaging and Graphics 32(3), 193 – 201 (2008) **[Osma-Ruíz08]**
- Marendic, B., Galatsanos, N., Bless, D.: New active contour algorithm for tracking vibrating vocal folds. In: Image Processing, 2001. Proceedings. 2001 International Conference on, vol. 1, pp. 397–400 (2001) **[Marendic01]**
- I. E. Adbou, W. K. Pratt ¨*Quantitative design and evaluation of enhancement/thresholding edge detectors*¨ Proceedings of the IEEE, vol.67, pp. 753-763, 1979. **[Pratt79]**
- S. Lankton, A. Tannenbaum. ¨*A localizing region-based active contours*.¨ IEEE Trans. on Image Processing, pp. 2029-2039, 2008. **[Lankton08]**
- S.-Z. Karakozoglou, N. Henrich, C. D'Alessandro, Y. Stylianou, "Automatic glottal segmentation using localbased active contours and application to glottovibrography," Speech Communication, vol. 54, no. 5, pp. 641–654, 2012. **[Karak12]**

Thank you for your attention

Grants: **TEC2009-14123-C04** and **TEC2012-38630-C04-01** from the Spanish Ministry of Economy.

Thanks to Dr. Erkki Bianco and Gilles Degottex for the database provided.

Gustavo Xavier Andrade Miranda
Juan Ignacio Godino Llorente
gxandrade@ics.upm.es

Dept. ICS, Universidad Politécnica de Madrid, Spain