WRAPPING SNAKES FOR IMPROVED LIP SEGMENTATION

Matthew Ramage, Euan Lindsay

Dept. of Mechanical Engineering
Curtin University of Technology
GPO Box U1987, Perth, WA 6845, Australia

ABSTRACT

A key step in the process of lip-reading is determining the shape of the speaker's lips. This has previously been achieved through an energy method known as "snakes", however this approach has some limitations. This paper presents an adapted approach called *wrapping snakes*, where the image forces are modified based on the snake's location and orientation. This modification encourages wrapping snakes to continue along features they have already partially found, overcoming one of the problems of traditional snakes. The use of wrapping snakes allows for more accurate and robust lip segmentation, as well as increasing the speed of the segmentation.

***Index Terms*—** speech recognition, image segmentation, active contours

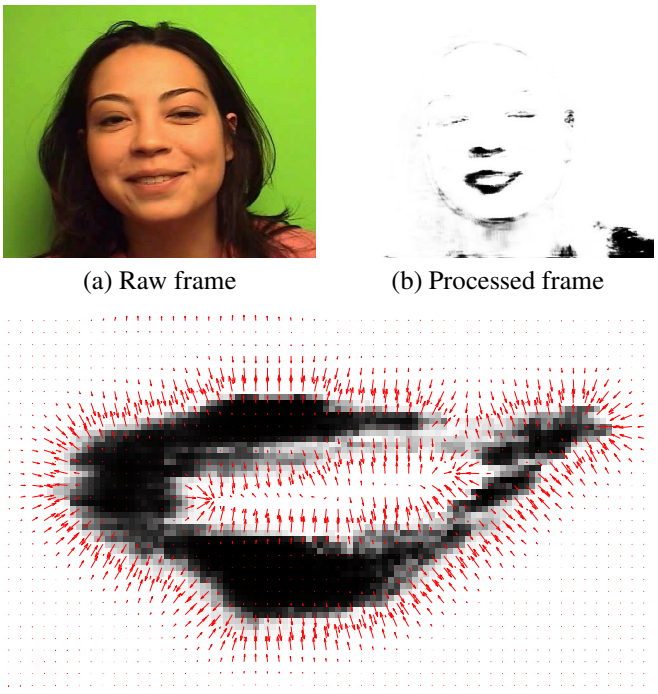
1. INTRODUCTION

Obtaining a transcript of what is spoken is desirable in many unscripted situations, such as live interviews and surveillance operations. Traditional audio-only speech recognition is not always suitable, as the recognition accuracy falls dramatically as the audio signal is degraded [1], and audio is not always available. In these situations, an alternative approach is to use visual speech recognition which uses video footage of a person speaking to identify the words they are saying.

The underlying principle of visual speech recognition is that the mouth shape is based on the underlying structure of the word being spoken [2]. This paper presents an adapted technique, known as wrapping snakes, for finding the shape of the lips. By accurately finding the lip shape, it will be possible to determine the words that are being spoken, and in turn a transcript can be obtained without using any audio information.

The lip segmenter uses images that have been preprocessed to identify areas that are a similar colour to lips. An example frame can be seen in Figure 1, using an image from the CUAVE image set [3].

A good lip segmenter mush be able to find the shapes quickly, be robust, and be easy to use. A technique known as "snakes" is reasonably well suited to this task, but there are still many areas for improvement. Wrapping snakes is based on this technique, but is modified to overcome some of the original limitations.



(a) Raw frame (b) Processed frame

(c) Image forces in the lip region

Fig. 1. Example of a raw frame, processed frame, and image forces in the lip region.

2. TRADITIONAL SNAKES

Much of the recent work in lip segmentation has focused on deformable models [4, 5, 6]. These mathematical models use an energy function to fit a parameterised model to the image. One of the more promising approaches has been the use of snakes.

Snakes are a series of connected points that are controlled by a mathematical model. They are a form of active contour model, which use an energy minimising spline that is guided

by internal and external forces [7]. The internal forces are due to the rigidity and tension of the spline, while the external forces are chosen to track the desired features.

As demonstrated in [7], snakes can be used to track the outer lip boundary in video sequences. Snakes are used in visual speech recognition to find the shape of the lip boundary, allowing the system to parameterise the lip shape. The sequence of lip shapes can then be used to perform the actual speech recognition.

The snake energy is minimised using an iterative procedure, where the next position of the snake is calculated as

$$\begin{aligned} \mathbf{x}_t &= (\mathbf{A} + \gamma\mathbf{I})^{-1}(\mathbf{x}_{t-1} - \mathbf{f}_\mathbf{x}(x_{t-1}, y_{t-1})) \\ \mathbf{y}_t &= (\mathbf{A} + \gamma\mathbf{I})^{-1}(\mathbf{y}_{t-1} - \mathbf{f}_\mathbf{y}(x_{t-1}, y_{t-1})) \end{aligned} \quad (1)$$

where $\mathbf{A}$ is a coefficient matrix as described in [7]. In Equation 1, $\mathbf{f_x}$ and $\mathbf{f_y}$ are the image forces in the x and y directions respectively.

3. TRADITIONAL SNAKE BEHAVIOUR

Figure 2 illustrates a snake successfully finding an outer lip boundary. The snake is initialised (blue line), several iterations are performed (green lines), and finally the snake converges to the lips (red line). This is an ideal situation, as there is no noise (false positives) in the image, the snake was initialised relatively close to the lip boundary and was touching the lip boundary in some places. As a result, the majority of the snake found the lip boundary on the first iteration.

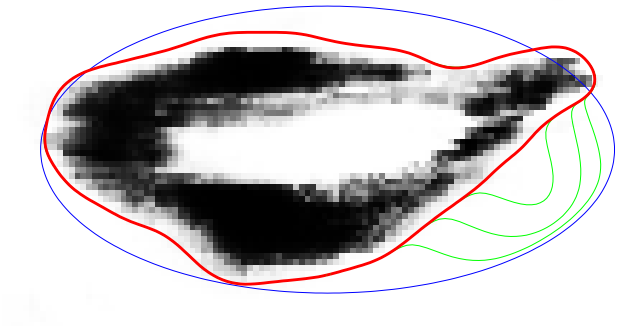


Fig. 2. Traditional snake finding the lip boundary under ideal conditions.

While this shows that snakes can be used to find the lip boundary, it is more important to analyse the snake's performance for less than ideal conditions. This includes poor initialisation and noisy images.

In situations where the snake is initialised in a position close to noise in the image, the traditional snake cannot always overcome the image forces due to the noise to successfully find the desired feature. This is especially true if the noise is strong and the desired feature is relatively weak. As can be seen in Figure 3, the nose produced a much stronger feature than the corner of the mouth, resulting in the increased tension pulling the snake away from the wrong feature.

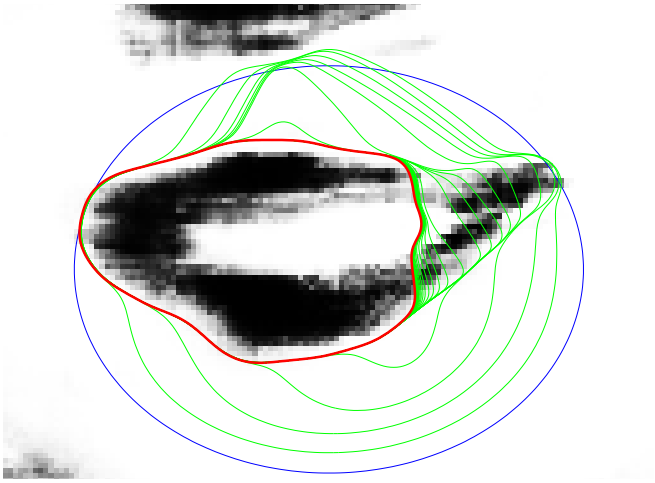


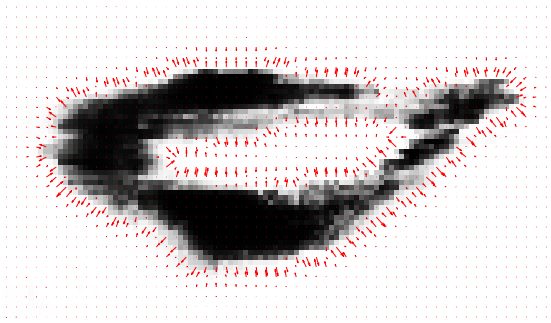
Fig. 3. Traditional Snake with strong noise and a weak target feature.

With traditional snakes, the image force is always perpendicular to the edge feature, which does not encourage the snake to continue along an edge it has partially found. As a result, if different parts of the snake are attached to separate edge features, it will not successfully find either complete feature.

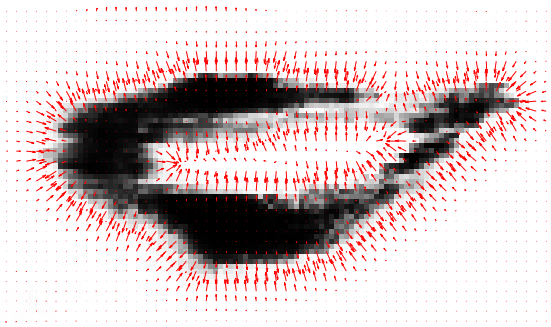
With traditional snakes, not only is there nothing encouraging the snake to continue along features it has found, but there is also nothing discouraging the snake from retreating along these same features. As the direction of the image forces is always perpendicular to the feature, the only force parallel to the feature is the tension force. In sections of a traditional snake that are curving away from the feature, the tension of the snake will cause it to be pulled back along the feature.

A previous attempt to solve this problem took the approach of creating an image force that has a longer reach that just a simple gradient function. This approach used gradient vector flow (GVF) fields [8], which apply a small Gaussian kernel many times to the lip mask image. As shown in Figure 4, the overall effect is smoother image forces that reach much further than before, which can help pull the snake into concave areas [8]. This technique increases the reach of the image forces and can pull the snake into small concave areas, but does not help with large concave features, or two close but independent features. The problem with GVF image forces is still that they don't encourage a snake to continue along a feature it has already partially found.

What is needed is a force that pushes a snake along a feature it has already partially found. This would allow the snake to wrap around the lip feature once it partially finds it. This wrapping force would be a substitute for the image force in Equation 1.

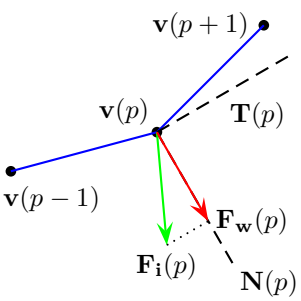


(a) Gradient forces

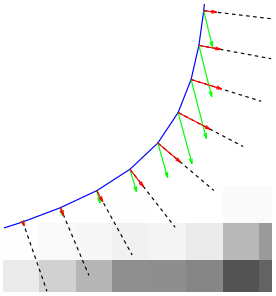


(b) Gradient Vector Flow (GVF) forces

Fig. 4. Comparison of different image forces.



(a) Determining wrapping force.



(b) Example forces.

Fig. 5. Comparison of wrapping forces (red) and image forces (green).

4. WRAPPING SNAKES

To overcome the problems with traditional snakes, a wrapping force is introduced as a substitute for the image force in Equation 1. The wrapping force is based on the image force, but is modified by the snakes shape and location at each iteration. The wrapping force is simply the component of the image force that is in the direction of the normal of the snake at that point.

As can be seen in Equation 2, the wrapping force, $\mathbf{F_w}$, can be calculated as the dot product of the image force, $\mathbf{F_i}$, and the unit normal, $\hat{\mathbf{N}}$, multiplied by the negative unit normal.

$$\mathbf{F_w} = -(\hat{\mathbf{N}} \bullet \mathbf{F_i})\hat{\mathbf{N}} \tag{2}$$

The original pair of equations for calculating the position of the snake at the next iteration (see Equation 1) are also used for wrapping snakes. In the original equations, $\mathbf{f_x}$ and $\mathbf{f_y}$ represent the x and y components of the image forces, whereas for wrapping snakes they represent the x and y components of the wrapping forces.

The direction of the unit normal at a given point can be calculated from the locations of the current and adjacent points on the snake. Let the location of the point be defined as

$$\mathbf{v}(p) = (x(p), y(p)) \tag{3}$$

Let $\mathbf{N}(p)$ be the normal vector at point p. The direction of the normal vector will be halfway between the directions of the vector from point $\mathbf{v}(p)$ to $\mathbf{v}(p-1)$ and the vector from point $\mathbf{v}(p)$ to $\mathbf{v}(p+1)$ (see Figure 5). The tangent at this point is $\mathbf{T}(p)$.

The direction of $\mathbf{N}(p)$ is perpendicular to the tangent $\mathbf{T}(p)$. The tangent can be calculated as

$$\mathbf{T}(p) = \frac{\mathbf{v}(p) - \mathbf{v}(p-1)}{|\mathbf{v}(p) - \mathbf{v}(p-1)|} + \frac{\mathbf{v}(p+1) - \mathbf{v}(p)}{|\mathbf{v}(p+1) - \mathbf{v}(p)|} \tag{4}$$

Since $\mathbf{N}(p)$ is perpendicular to $\mathbf{T}(p)$, it can be calculated by performing a 90° rotation (see Figure 5 (a)). The direction of the rotation is not important, as the dot product in Equation 2 will correct for it.

Since this new wrapping forces is always perpendicular to the snake, it will encourage the snake to be pulled along features the snake is curving away from (see Figure 5 (b)). This encourages the snake to continue along features it has already found, overcoming one of the shortcomings of traditional snakes.

When the snake is perpendicular to the image forces, the wrapping forces will be equal to the image forces, as the image force is already in the same direction as the normal. This means that the snake will behave in a similar way to traditional snakes when no wrapping is required. One advantage of using wrapping snakes, even when traditional snakes work, is reduction in iterations required to successfully locate the lip boundary. This is illustrated in Figure 6, where only one iteration is required for the wrapping snake compared to four iterations for the traditional snake (see Figure 2).

As the wrapping force encourages the snake to continue along a feature it has already found, it allows the snake to correctly find a weak target feature even when there is strong noise nearby (Figure 7). When this is compared to the traditional snake (Figure 3), it is clear that the wrapping and internal forces allow the snake to disregard the shadow under the nose, without falling off the side of the mouth.

Even with very poor initialisation, wrapping snakes can still successfully locate the lip boundary. Figure 8 shows a snake with an initial position is along the shadow under the

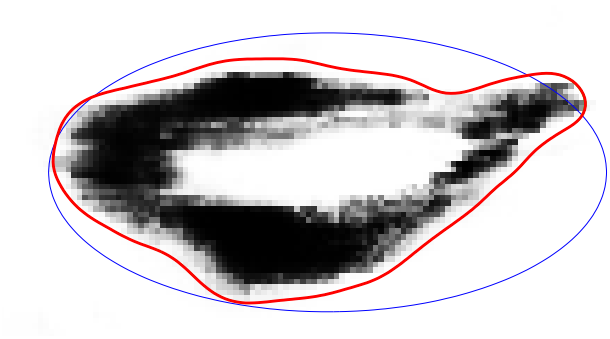


Fig. 6. Wrapping snake finding the lip boundary under ideal conditions.

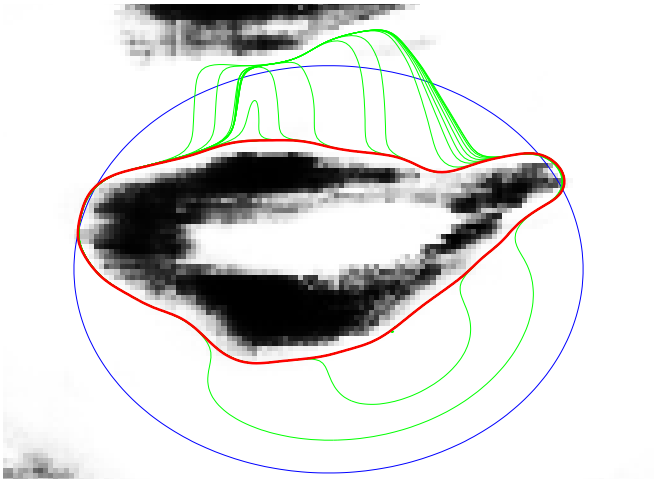


Fig. 7. Wrapping snake with weak target and strong noise features.

jaw line and also passes through the shadow under the nose. In this situation, the wrapping snake is still able to successfully locate the lip boundary given enough iterations, whereas the traditional snake will never succeed.

5. CONCLUSION

By modifying the the image force based on the snakes location and orientation, the wrapping force encourages the snake to continue along features it has partially found. It has been shown that wrapping snakes allow the lip boundary to be found more accurately than with traditional snakes, and are more robust to noise and poor initialisation.

As visual speech recognition requires an accurate representation of the lip shape, the use of wrapping snakes can improve the performance of these systems. With the requirements of a good lip segmenter being fast operation, robust to noise and poor initialisation, and be easy to use, wrapping snakes are well suited for this purpose.

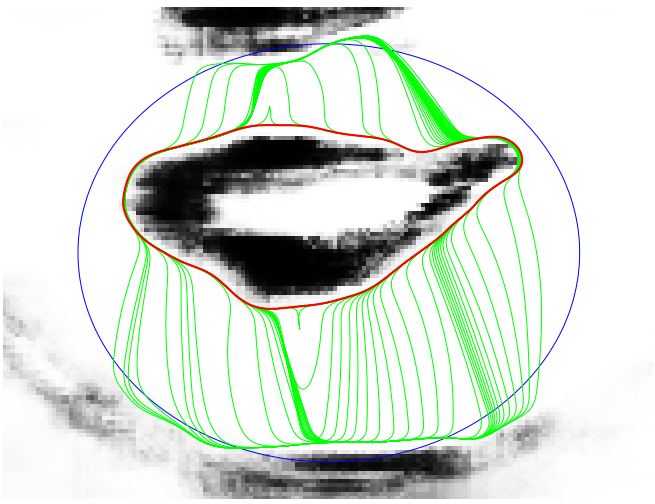


Fig. 8. Wrapping snake successfully finding the lips, even with very poor initialisation.

6. REFERENCES

[1] S. Dupont, "Audio-visual speech modeling for continuous speech recognition," *IEEE transactions on multimedia*, vol. 2, no. 3, pp. 141, 2000, 1520-9210.

[2] E. J. Holden and R. Owens, "Visual speech recognition using cepstral images," in *Proceedings of IASTED International conference on Signal and Image Processing*, 2000, p. 331336.

[3] E. Patterson, S. Gurbuz, Z. Tufekci, and J. N. Gowdy, "Cuave: a new audio-visual database for multimodal human-computer interface research," 2002, vol. 2, pp. 2017–2020.

[4] Tsuhan Chen and R. R. Rao, "Audio-visual integration in multimodal communication," *Proceedings of the IEEE*, vol. 86, no. 5, pp. 837–852, 1998, 0018-9219.

[5] N. Eveno, A. Caplier, and P. Y. Coulon, "Automatic and accurate lip tracking," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 5, pp. 706–715, 2004.

[6] Zhilin Wu, P. S. Aleksic, and A. K. Katsaggelos, "Lip tracking for mpeg-4 facial animation," in *Fourth IEEE International Conference on Multimodal Interfaces, 2002. Proceedings.*, 2002, pp. 293–298.

[7] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *International Journal of Computer Vision*, vol. 1, no. 4, pp. 321–331, 1988.

[8] Xu Chenyang and J. L. Prince, "Snakes, shapes, and gradient vector flow," *Image Processing, IEEE Transactions on*, vol. 7, no. 3, pp. 359–369, 1998, 1057-7149.