Deformation Analysis of the Vibrational Patterns of the Vocal Folds

Abdul Karim Saadah and Nikolas P. Galatsanos

Illinois Institute of Technology
Department of Electrical and Computer Engineering
Chicago, Illinois 60616
tel. (312)-567-5259
e-mail: npg@ece.iit.edu

Diane Bless and Anie Ramos

Department of Communicative Disorders
University of Wisconsin-Madison
Madison, WI 53706

Abstract- Videostroboscopy is an examination which yields a permanent record of the moving vocal folds. Thus, it allows the diagnosis of abnormalities which contribute to voice disorders. In this paper, in order to find and quantify the deformation of the vocal folds in videostroboscopic recordings, an active contours (snakes) based approach is used to delineate the vocal folds in each frame of the videostroboscopic image sequence. After this delineation, a new elastic registration algorithm is used to register the vocal fold contours between adjacent frames of the video sequence. This algorithm is based on the regularization principle and is very effective for our application. A least-squares approach is used to fit an affine model to the displacement vectors found by elastic registration. The parameters of this model, rotation, translation, and deformation along two principle axes, quantify the deformation and allow the succinct characterization of the videostroboscopic recordings based on the deformations that occurred. Experiments are shown with synthetic and real videostroboscopic data that demonstrate the value of the proposed approach.

Keywords: vocal folds, deformation analysis, elastic registration, affine transformation modeling.

1 Introduction

Videostroboscopy is a method for the clinical examination of the vibrational characteristics of the vocal folds [1, 2]. This method since it provides a permanent image record of the moving vocal folds, is used for diagnosis of vibrational abnormalities which contribute to voice disorders that cannot be detected by other clinical methods [1, 2]. Since the vocal folds vibrate too fast to be recorded on video, stroboscopic flashes which are slightly "of phase" with the frequency of the vocal folds are used to yield an image sequence in which the vocal folds appear to move in "slow motion"[1]. It is of great interest to voice clinicians both for diagnostic and research purposes to measure and to quantify the deformation of the vocal folds during phonation [2]. Up to this point voice clinicians have been visually inspecting videostroboscopic recordings for diagnostic purposes. However, it is clear that the observer bias and the inability of humans to absorb and quantify the large amounts of information in video sequences compromise the effectiveness of this examination. The goal of the work in this paper is to measure and quantify the deformations that occur in videostroboscopic recordings of the vocal folds.

For this purpose the vocal fold contours are first detected using active contour models (snakes) [3, 8]. Then, the deformation of the vocal folds is found by elastically registering the contours of the vocal folds from frame-to-frame. The collection of all the vectors that describes the displacement of the contours of the vocal folds in a point-by-point manner from frame-to-frame is called the displacement vector field (DVF). Once the DVF is found, the need to model it quantitatively arises. Although the DVF captures the vocal fold deformations, it does not provide a succinct description of them. Thus, a model that describes them is necessary. A simple model that is easy to compute and also captures the deformations of the vocal folds is an *affine transformation* (AT) model [7]. In this paper, a least squares procedure is developed to fit the AT model to the DVF between successive frames. The obtained AT is then decomposed to a rotation and a deformation along two principle axes. The time evolution of the rotation and deformation parameters characterizes succinctly the deformation of the vocal folds in the videostroboscopic sequence.

The rest of this paper is organized as follows. In section 2, we present the new elastic registration algorithm that was used for the registration of the vocal fold contours from frame-to-frame. In section 3, we present the modeling of the vocal fold deformations using the DVF and the AT model. In section 4, we present our experimental results. Finally, in section 5, we present our conclusions from this work and our future research directions.

2 Regularization Based Elastic Registration of the Vocal Folds

Once the vocal fold contours have been delineated, they have to be matched elastically from frame-to-frame. In elastic contour matching the two contours, say the smaller 1 and the larger 2, are matched/registered by finding the correspondence between points in contours 1 and 2. Assume that the vocal fold contours 1 and 2 in two successive frames are defined by points ($x_1(m)$, $y_1(m)$) and ($x_2(k)$, $y_2(k)$), for $m = 1,\ 2,\ ..\ L_1$ and $k = 1,\ 2,\ ..\ L_2$. Matching the two contours corresponds to finding the pairs of indices $(m(i), k(i))$ for $i = 1,\ 2,\ ..\ L_2$ where L_2 the larger contour, that describes a correspondence of points between the two contours. The pair of points $(m(i), k(i))$ defines the i^{th} displacement vector, and the collection of all these points for $i = 1,\ 2,\ ..\ L_2$ vectors gives the DVF. The DVF describes the elastic deformation of the two contours.

An elastic contour registration method based on dynamic programming (DP) and auto-regressive (AR) modeling was proposed in [4]. The cost function to be minimized using DP is chosen to be a function of the estimated (using AR models) displacement vector and the actual displacement vector found. Another algorithm that also matches deformed contours was proposed in [5]. Unlike the approach in [4], this approach is based entirely on DP. However, instead of choosing as cost function the distance between two points (size of the displacement vectors), the weighted sum of the difference between two successive displacement vectors and the size of the displacement vectors themselves is used.

For our application, we used a new elastic contour matching algorithm [6] that ameliorates the difficulties of the algorithms in [4, 5]. The proposed algorithm is based on a cost function that is minimized using *simulated annealing* (SA) [11]. Simulated annealing is an approach to solve combinatorial optimization problems. Simulated annealing associates a " temperature " T, with a model and searches for the energetically optimal configuration by repeatedly trying to alter the state of the system. A simulated annealing algorithm is completely defined upon specifying a cooling schedule. This cooling schedule consists of the initial temperature of the system T_0, a rule for determining the number of transitions at each temperature, a rule for determining the lowering of the temperature, and a termination criterion.

A cost function based on the *regularization principle* is defined [9, 10]. This cost function is minimized with respect to choice of the displacement vectors $(m(i), k(i))$ for $i = 1, 2, \cdots L_2$ and is of the form

$$\begin{aligned} C = \ & (1-\lambda) \sum_{i=1}^{L_2} [\ (x_1(m(i)) - x_2(k(i)))^2 + (y_1(m(i)) - y_2(k(i)))^2\ ] \\ & + \lambda \sum_{i=1}^{L_2} [|\ \phi(m(i), k(i)) - \phi(m(i-1), k(i-1))\ |]^2 , \end{aligned} \tag{1}$$

where $\phi(m(i), k(i))$ is the angle of the i^{th} displacement vector which is defined by the points $m(i)$ and $k(i)$ of the contours 1 and 2, respectively. The first part of the function C captures the requirement that the total distance between points that are matched should be small. The second part of the function imposes the smoothness constraint to the displacement vector field by constraining adjacent motion vectors to have similar directions. The use of the regularization principle in the registration problem is especially helpful when the deformations are severe and the contours are not smooth. In this case, the data alone are not sufficient to provide a meaningful DVF.

3 Affine Transform Modeling of the Deformation

After elastic registration, the DVF between adjacent frames of the vocal folds is available. The DVF provides a better visualization of the deformation than the raw video-recordings. However, the DVF alone does not provide a *quantitative* description of the deformation. Therefore, in this section we present the model that is used to quantify the deformation from the DVF.

As explained previously the deformation at the location i between two successive frames for example 1 and 2 is represented as a vector which is given by two pairs of coordinates. This vector is called the displacement vector (DV). Let ($x_1(i)$, $y_1(i)$) and ($x_2(i)$, $y_2(i)$) denote the origin and the end-point of the DV at location i. Assuming that

L such points are given, then the DVF between frames 1 and 2 is given by the collection of all available DVs. The DVF is fully described by the set of points

$$(\ x_l(i),\ y_l(i)\ )\ \text{for}\ l=1,2\ \text{and}\ i=1,2,\cdots,L.$$

A simple model for the DVF that is very appropriate because it can capture the deformation during the motion of the vocal folds is based on the *affine transformation* (AT) [7, 12]. This model is described by the following set of linear equations

$$\begin{bmatrix} x_2(i) \\ y_2(i) \end{bmatrix} \approx \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x_1(i) \\ y_1(i) \end{bmatrix} + \begin{bmatrix} T_x \\ T_y \end{bmatrix} \text{ for } i=1,2,\cdots,L. \tag{2}$$

The $\approx$ sign is used because fitting an AT model to more than 3 motion vectors is an overdetermined problem which has to be solved approximately. In what follows, we will elaborate on this point. The matrix A defined by the coefficients $a_{11}, a_{12}, a_{21}, a_{22}$ describes *deformation* and *rotation* [7]. While T_x, T_y describe *translation* in the x and y directions, respectively. More specifically, the matrix A can be decomposed into the product of R and D [13], where D is a symmetric positive definite matrix that represents deformation and R is an orthonormal matrix. More specifically

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = R \cdot D,\ R = \begin{bmatrix} cos\theta & -sin\theta \\ sin\theta & cos\theta \end{bmatrix} \text{ and } D = \begin{bmatrix} a & c \\ c & b \end{bmatrix}. \tag{3}$$

Because of the symmetry of D we can also write

$$D = \begin{bmatrix} a & c \\ c & b \end{bmatrix} = \begin{bmatrix} \vec{\mathbf{e}}_1^T \\ \vec{\mathbf{e}}_2^T \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \begin{bmatrix} \vec{\mathbf{e}_1}, \vec{\mathbf{e}_2} \end{bmatrix}, \tag{4}$$

where λ_i, $\vec{\mathbf{e}_i}$, for $i=1,2$ are the eigenvalues and the corresponding orthogonal eigenvectors, respectively, of matrix D. To find D and R we observe that

$$A^T A = (RD)^T (RD) = D^T R^T RD, \tag{5}$$

but, since R is orthonormal matrix then,

$$R^T R = I, \tag{6}$$

where I is the identity matrix. Therefore,

$$A^T A = D^T D. \tag{7}$$

But since D is symmetric, there exists a unique decomposition of $A^T A$ such that D is symmetric and positive definite.

Let λ_α^2, $\vec{\mathbf{e}_\alpha}$, for $\alpha = 1,2$, be the eigenvalues and eigenvectors, respectively of the symmetric matrix $A^T A$, i.e.

$$A^T A = \lambda_1^2 \vec{\mathbf{e}_1}\vec{\mathbf{e}_1}^T + \lambda_2^2 \vec{\mathbf{e}_2}\vec{\mathbf{e}_2}^T. \tag{8}$$

Then

$$D = \lambda_1 \vec{\mathbf{e}_1}\vec{\mathbf{e}_1}^T + \lambda_2 \vec{\mathbf{e}_2}\vec{\mathbf{e}_2}^T. \tag{9}$$

Once D is estimated, R can be estimated uniquely as $R = AD^{-1}$. The eigenvectors of the D matrix give the perpendicular directions of the maximum and minimum deformation [7]. These directions are given by the angles

$$\phi_1 = tan^{-1}(\frac{\mathbf{e}_1(y)}{\mathbf{e}_1(x)}), \text{ and } \phi_2 = tan^{-1}(\frac{\mathbf{e}_2(y)}{\mathbf{e}_2(x)}) , \tag{10}$$

where $\mathbf{e}_j(x)$, $\mathbf{e}_j(y)$ are the x and y components of the vector $\vec{\mathbf{e}_j}$. The corresponding eigenvalues give the changes of the magnitude of the deformation along these directions. Therefore, the AT model of the DVF provides a succinct description of the changes that occurred in the shapes of the contours.

The $\approx$ sign in Eq.(2) is used because in general it is impossible to find a unique set of $(a_{11}, a_{12}, a_{21}, a_{22}, T_x, T_y)$ parameters that will satisfy Eq.(2) exactly for all $i = 1, 2, \cdots, L$. Thus, a set of such parameters is found that "on the average" is optimal for $i = 1, 2, \cdots, L$. For this problem the *least-squares* (LS) approach is used [14]. According to this approach a model is found which minimizes the sum of the squared error over the entire set of range data that is fitted with our model. Rearranging Eq.(2) for $i = 1, 2, \cdots L$ we can write the following system of linear equations $Bx = b$ with

$$B = \begin{bmatrix} x_1(1) & y_1(1) & 0 & 0 & 1 & 0 \\ 0 & 0 & x_1(1) & y_1(1) & 0 & 1 \\ x_1(2) & y_1(2) & 0 & 0 & 1 & 0 \\ 0 & 0 & x_1(2) & y_1(2) & 0 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_1(L) & y_1(L) & 0 & 0 & 1 & 0 \\ 0 & 0 & x_1(L) & y_1(L) & 0 & 1 \end{bmatrix}, \tag{11}$$

$$x = \begin{bmatrix} a11, & a12, & a21, & a22, & T_x, & T_y \end{bmatrix}^T, \tag{12}$$

$$b = \begin{bmatrix} x_2(1), & y_2(1), & x_2(2), & y_2(2), & \cdots & x_2(L), & y_2(L) \end{bmatrix}^T, \tag{13}$$

where B and b are a $2L \times 6$ matrix and $2L \times 1$ vector, respectively, of the data. Vector x is 6×1 and contains the parameters that are fitted to the data. For $L > 6$ clearly this is an overdetermined system of equations so we resort to a least-squares solution.

Minimizing $\|Bx - b\|_2$, where $\|\cdot\|_2$ is the l_2-norm, gives the following equations,

$$B^T B \hat{x}_{LS} = B^T b ,$$

from which we can solve for $\hat{x}_{LS}$ and get the affine model parameters. The matrix $B^T B$ is a 6×6 symmetric matrix given by

$$\begin{bmatrix} \sum_{i=1}^L x_1^2(i) & \sum_{i=1}^L x_1(i)y_1(i) & 0 & 0 & \sum_{i=1}^L x_1(i) & 0 \\ \sum_{i=1}^L x_1(i)y_1(i) & \sum_{i=1}^L y_1^2(i) & 0 & 0 & \sum_{i=1}^L y_1(i) & 0 \\ 0 & 0 & \sum_{i=1}^L x_1^2(i) & \sum_{i=1}^L x_1(i)y_1(i) & 0 & \sum_{i=1}^L x_1(i) \\ 0 & 0 & \sum_{i=1}^L x_1(i)y_1(i) & \sum_{i=1}^L y_1^2(i) & 0 & \sum_{i=1}^L y_1(i) \\ \sum_{i=1}^L x_1(i) & \sum_{i=1}^L y_1(i) & 0 & 0 & L & 0 \\ 0 & 0 & \sum_{i=1}^L x_1(i) & \sum_{i=1}^L y_1(i) & 0 & L \end{bmatrix}.$$

Apart from the trivial straight line case where $y_1(i) = \alpha x_1(i)$ for $i = 1, 2, \cdots, L$, where α is a constant, $B^T B$ is positive definite and thus $(B^T B)^{-1}$ exists. Therefore, the computation of $\hat{x}_{LS}$ is easy.

4 Experimental Results

In this section we present experiments that test the proposed algorithms on real data of normal and abnormal videostroboscopic image sequences of the larynx. In the first experiment, we used a sixteen frame videostroboscopic image sequence showing a complete vibrational cycle of normal vocal folds. All our real data was provided by the last two authors. In Fig.(1), the sequence is shown running from top left to bottom right starting from frame 1 and ending at frame 16. The bright contours delineate the boundaries of the vocal folds in the successive frames and were produced using the snake algorithm [8]. It can be seen that the positions and changes in the shape of the vocal folds are correctly extracted using the snakes. The DVFs of all contours for the successive frames are shown in Fig.(2).

Fig.(3) (a) and (b) show the relative deformation and the directions of deformation, respectively, for the vocal folds from frame-to-frame. The relative deformation from frame-to-frame is given by the eigenvalues of matrix D *i.e.* λ_1 and λ_2, while the corresponding directions of deformation are given by the angles ϕ_1 and ϕ_2. As seen, λ_2 which represents the relative deformation along the ϕ_2 direction is close to one, i.e. $\lambda_2 \approx 1$. This implies that the deformation along the vertical direction is very small. For this sequence, a larger deformation occurs along the horizontal direction, i.e. $\phi_1 \approx 90$. The time intervals for which $\lambda_1 > 1$ defines the opening cycle of the vocal folds. From Fig.(3)(a) we observe that the opening cycle lasts until frame 8 after which the vocal folds start closing ($\lambda_1 < 1$). Also from Fig.(3)(a) we observe that the maximum relative opening takes place between frames 3 and 4 and the maximum relative closing takes place between frames 14 an 15. These observations cannot be made neither from the raw images in Fig.(1) nor from the DVF in Fig.(2). For this sequence, as shown in Fig.(3)(c), the translational motion along the x and y directions is very small since it represents the small translational motion of the centroid of the vocal folds. The vocal folds do not undergo rotational motion also and this can be observed in Fig.(3)(d).

Another sixteen frame normal videostroboscopic image sequence was analyzed. Due to space constraints only the deformation analysis is shown in Fig.(4). From Fig.(4)(a) and as in the previous sequence, the opening cycle ($\lambda_1 > 1$) lasts until frame 8, after which the vocal folds start closing ($\lambda_1 < 1$). The translational and rotational components of the deformation are again very small, see Fig(4)(c), and (d).

In another experiment, we used sixteen frames from an abnormal videostroboscopic image sequence of a patient with polyps [2]. In Fig.(5) the sequence is shown running from top left to bottom right. The bright contours delineate the boundaries of the vocal folds in the successive frames. The translational and rotational components of the deformation are again very small, see Fig(6)(c), and (d).

For this sequence also there is almost no deformation along the ϕ_2 direction for which $\lambda_2 \approx 1$, see Fig.(6)(a). From Fig.(6)(a) we can again observe that the dominant deformation is along the ϕ_1 direction. From this figure we can also tell that the vocal folds kept opening till frame 6 ($\lambda_1 > 1$) then an unexpected closing occurred between frames 7 through 9, then they opened between frames 9 and 10 and after that the closing cycle began.

In the last experiment, we used twenty frames from an abnormal videostroboscopic image sequence of a patient with a cyst [2]. Due to space constraints only the deformation analysis is shown in Fig.(7). The translational and rotational components of the deformation are again very small, see Fig(7)(c), and (d). From this figure we can also tell that the vocal folds kept opening till frame 7 ($\lambda_1 > 1$) then an unexpected closing occurred between frames 8 and 9, then they continued to open till frame 11.

Because of the continuity of the opening and closing portions of the vocal fold cycle, for a normal vibrational pattern, the function $\lambda_1^{'}(t) = 1 - \lambda_1(t)$ where t indicates the time evolution (frame #), has to have only one zero-crossing per cycle. This is required so that the contour of the vocal fold expands in the horizontal direction. Similarly, during the closing $\lambda_1(t) < 1$. This is observed in Fig.(3)(a) and Fig.(4)(a). Sequences with abnormalities have an odd number (larger than one) zero-crossings per cycle for the function $\lambda_1^{'}(t)$, see for example, Fig.(6)(a) and Fig.(7)(a). Thus, the number of zero-crossings of the function $\lambda_1^{'}(t)$ can be used as a criterion for detecting abnormalities in videostroboscopic recordings of the vocal folds.

5 Conclusions

In this paper the problem of modeling the deformations of the vocal folds from videostroboscopic recordings was addressed. For this purpose a new system was developed. This system delineates the contours of the vocal folds using snakes, elastically registers the contours using a new regularization-based algorithm, thus, the DVF between adjacent frames is obtained, and finally fits an affine transform model to the available DVF.

We found that regularization can be very beneficial for this elastic registration problem especially in cases were severe deformations are present. We also observed that many important features of the deformation of the vocal folds can be captured very effectively by the time evolution of only a few affine transform parameters of our model. The most significant parameter is λ_1 which represents the deformation along the horizontal direction. In the future, based on this parameter we plan to try to explore metrics that quantify the degree of abnormality of the vibrational cycle of the vocal folds.

References

[1] D. Bless, M. Hirano, and R. Feder, "Videostroboscopic evaluation of the larynx," *Ear Nose & Throat Journal*, vol. 66, No. 7, July 1987.

[2] M. Hirano, and D. Bless, *Videostroboscopic Examination of the Larynx,* Singular Publishing Group Inc., 1993.

[3] Kass, A. Witkin and D. Terzopoulos "Snakes: active contour models," *Proc. 1st Int Conf. on Comp. Vision,* pp. 259-269, 1987.

[4] H. Maître and Y. Wu, "Dynamic programming algorithm for elastic registration of distorted pictures based on autoregressive model," *IEEE Trans. Acoust., Speech, and Signal Processing,* vol. 37, No. 2, pp. 288-297, 1989.

[5] D. Geiger, Alok Gupta, L. A. Costa, and J. A. Vlontzos, "Dynamic programming for detecting, tracking, and matching deformable contours," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* pp. 294-302, vol. 17, No. 3, March 1995.

[6] A. K. Saadah, N. Galatsanos and D. Bless, "A new elastic registration algorithm for videostroboscopic images of the larynx," *Proceedings of IASTED*, International Conference Signal and Image Processing-SIP-95, Las Vegas, pp. 93-96, Nevada 1995.

[7] E. A. Cemal, *Continuum Physics,* Academic Press, 1976.

[8] Y. L. Fok, J. C. K. Chan, and R. T. Chin, "Automated analysis of nerve-cell images using active contour models," *IEEE Trans. on Medical Imaging,* vol. 15, no. 3, pp. 353-367, July 1996.

[9] N. P. Galatsanos and A. K. Katsaggelos, "Methods for choosing the regularization parameter and estimating the noise variance in image restoration and their relation," *IEEE Trans. Image Processing*, vol. 1, no. 3, pp. 322-336, July 1992.

[10] D. Terzopoulos, "Regularization of inverse visual problems involving discontinuities," *IEEE Trans. Pattern Analysis Mach. Intell.*, vol. 8, no. 4, pp. 413-424, July 1986.

[11] E. Aarts and J. Korst, *Simulated Annealing and Boltzman Machines,* John Wiley, 1990.

[12] Petra. A. van den Elsen, Evert-Jan D. Pol, and Max A. Viergever, "Medical image matching - a review with classification," *IEEE Engineering in Medicine and Biology*, vol. 12, no. 1, pp. 26-39, March 1993.

[13] T. Y. Young and S. Gunasekaran, "A regional approach to tracking 3D motion in an image sequence," *Advances in Computer Vision and Image Processing*, T. S. Huang, Ed., pp. 63-99, Greenwich, Conn.: JAI Press, 1988.

[14] G. H. Golub and C. F. Van Loan, *Matrix Computations,* The Johns Hopkins University Press, 1989.

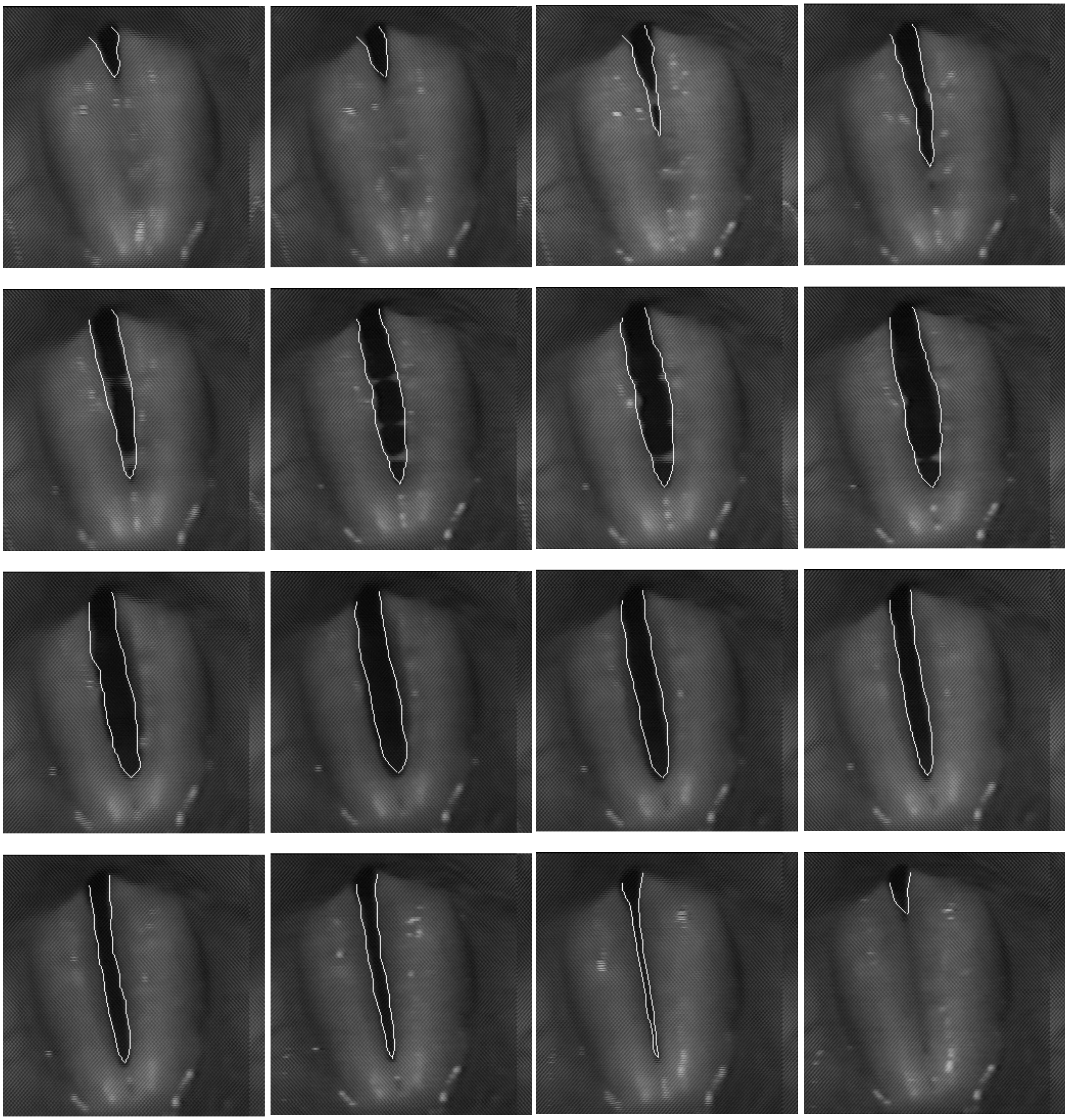


Figure 1: The bright closed contours delineate the boundaries of the vocal folds in the successive frames of a normal sequence run from top left to the bottom right.

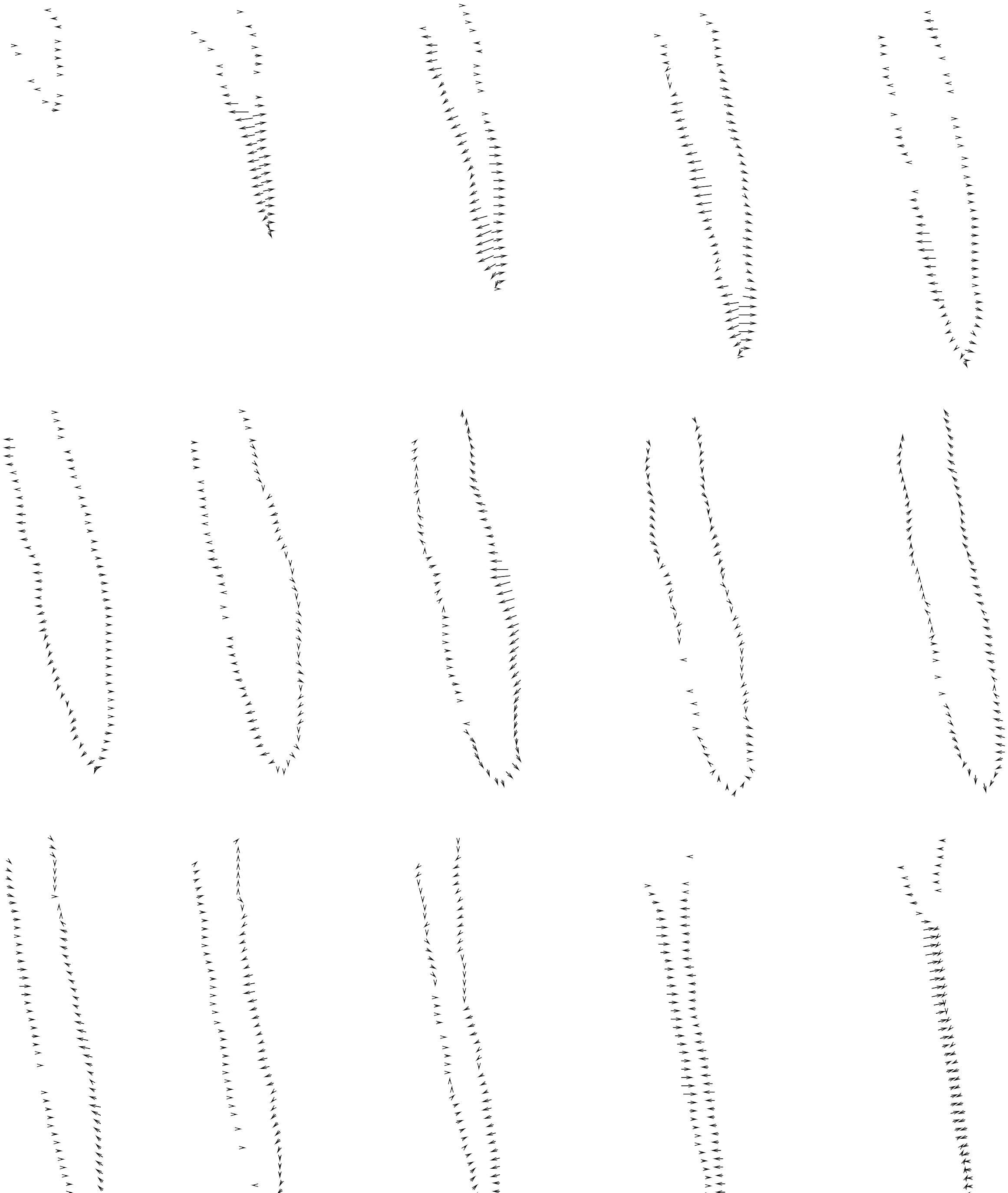


Figure 2: The registration of the normal sequence of videostroboscopic images using the proposed SA based approach runs from top left to the bottom right, the DVFs are sampled and magnified.

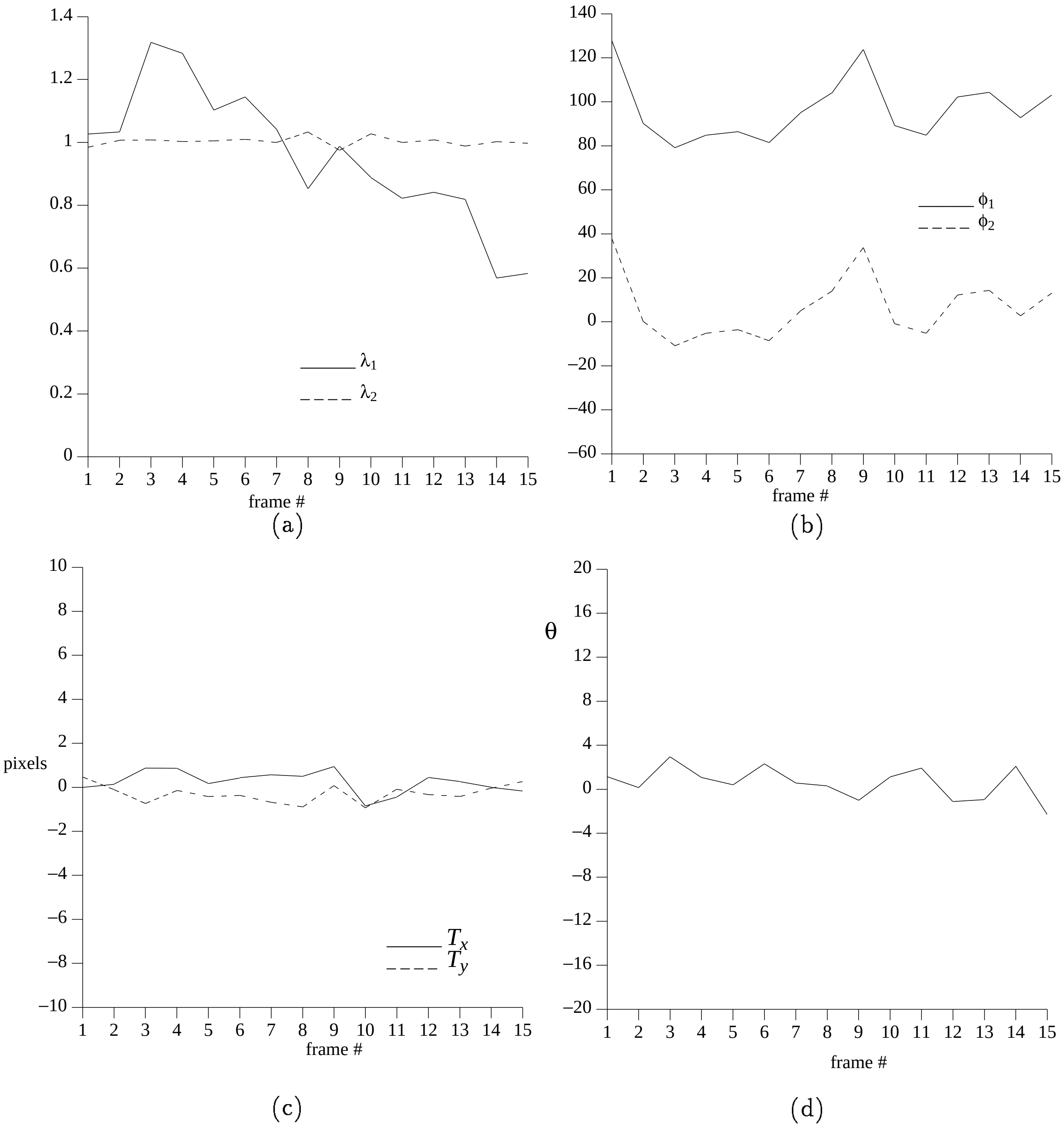
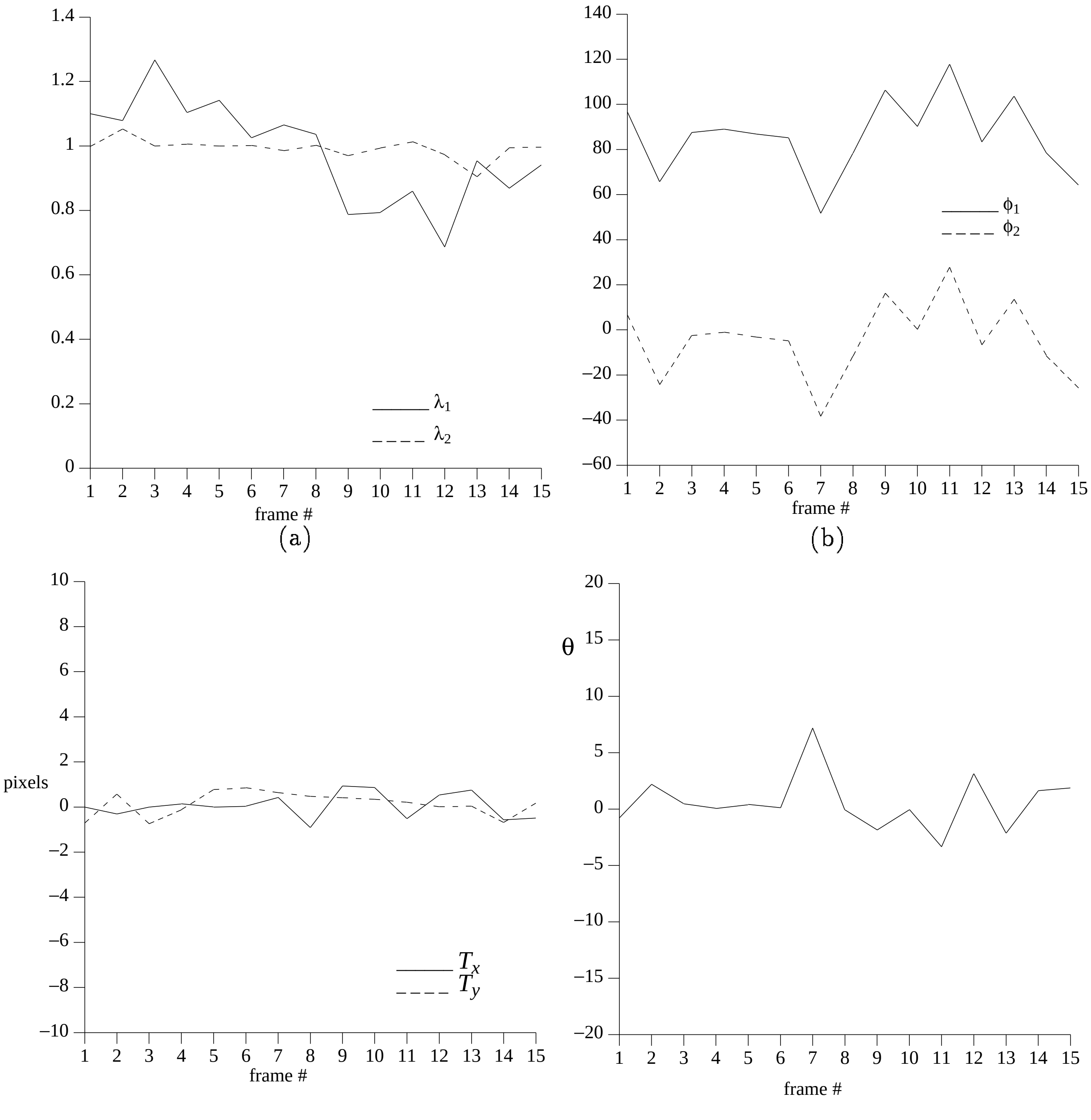


Figure 3: The figures in (a), (b), (c) and (d) represent time evolution of the eigenvalues, principal directions of deformation, the translation along the x and y axes and the rotation angle for the frames of the first normal videostroboscopic image sequence respectively.



(c) (d)

Figure 4: The figures in (a), (b), (c) and (d) represent time evolution of the eigenvalues, principal directions of deformation, the translation along the x and y axes and the rotation angle for the frames of the second normal videostroboscopic image sequence respectively.

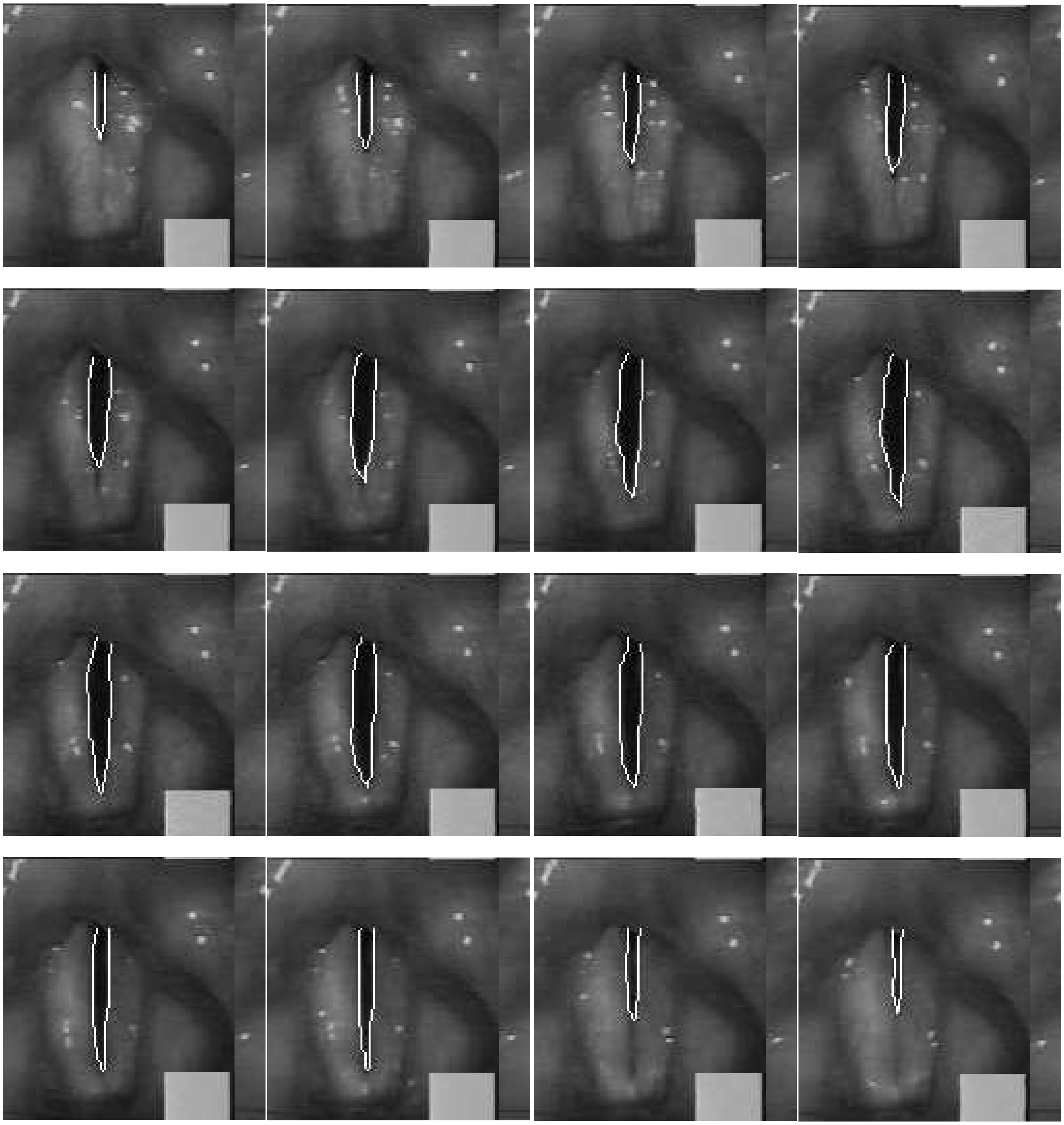


Figure 5: The bright closed contours delineate the boundaries of the vocal folds in the successive frames of the first abnormal sequence for patient with polyps run from top left to the bottom right.

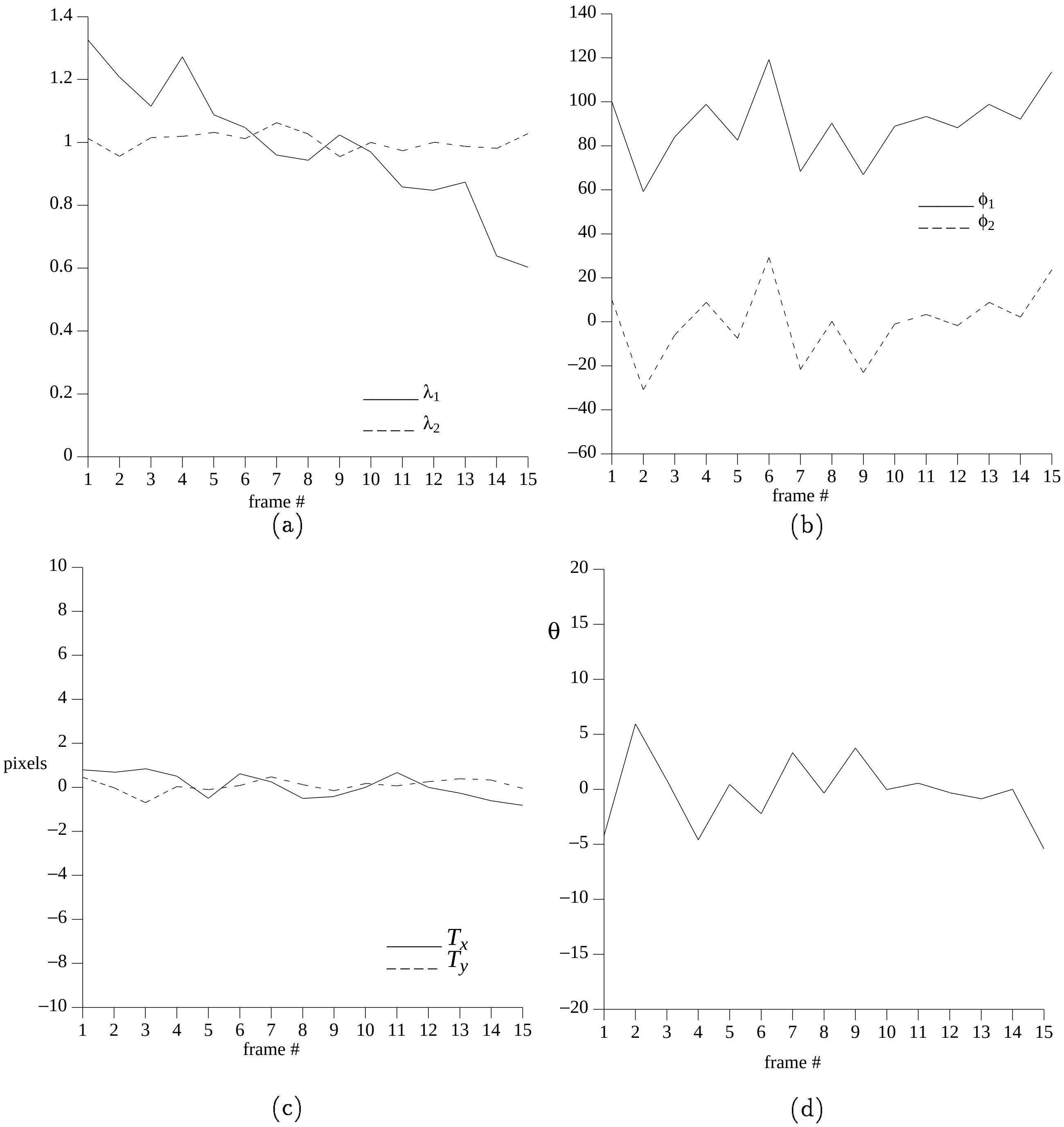


Figure 6: The figures in (a), (b), (c) and (d) represent time evolution of the eigenvalues, principal directions of deformation, the translation along the x and y axes and the rotation angle for the frames of the first abnormal videostroboscopic image sequence respectively.

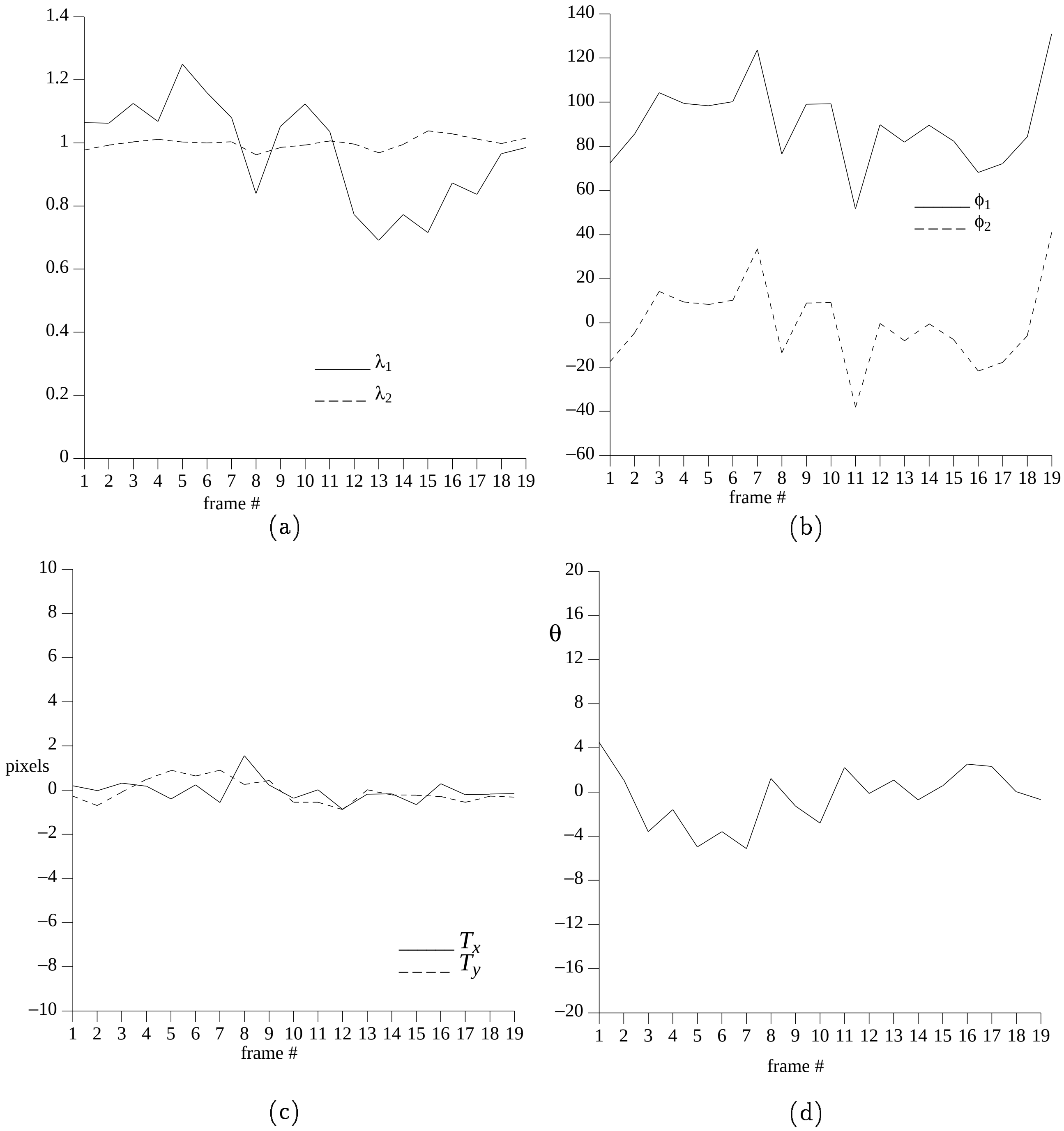


Figure 7: The figures in (a), (b), (c) and (d) represent time evolution of the eigenvalues, principal directions of deformation, the translation along the x and y axes and the rotation angle for the frames of the second abnormal videostroboscopic image sequence respectively.