Deformation analysis of the vocal folds from videostroboscopic image sequences of the larynx*

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

and
Carmen Ana Ramos
St. Michael's Hospital
Voice Disorders Clinic
Toronto, ON, Canada

Received:

*This work was supported by the Whitaker foundation.
†corresponding author.

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 when large deformations are present. 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] The method provides a permanent image record of the moving vocal folds, and is thus 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 "off 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] Voice clinicians routinely inspect videostroboscopic recordings visually 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.

The need for this measurement and quantification stems from the simple fact that computers are much better than humans at sorting out information in large amounts of data. The unaided human observer, no matter how trained, cannot capture and quantify all the available information about the motion of the vocal folds that is available in a videostroboscopic image sequence. For example, is is impossible for the human observer to measure accurately quantities such as the direction of motion and the change of the area of the vocal fold opening. *Small changes* in these quantities that could indicate the initial stages of a disease or a trend during treatment could escape undetected from a human observer. In addition, it is widely known that during videostroboscopic examinations as performed today, where trained humans observe the vibratory pattern, the *bias* of the observer is a serious problem limiting the usefulness of this examination.[2] It is impossible just by visual inspection to classify always *objectively*, let alone quantify and measure, properties of the motion pattern of the vocal folds. Motivated by the above limitations of this examination, the objective of this work is to propose a system that will help ameliorate some of these difficulties.

The vocal folds are non-rigid objects. Non rigid objects when in motion change shape over

time. Finding the motion of non-rigid objects is a hard problem which has been studied in the past for a number of different medical imaging applications in mind, see for example Refs. 3,4,5,6,7,8 and the references within. There exist three main methodologies to find the motion of non-rigid objects. For the first, specific anatomical features of the object (tokens) are employed to find the motion from frame-to-frame. For the second, the mechanical properties of the underlying elastic body are modeled and laws of mechanics are used to describe the motion. For the third, no information about the shape or the mechnical properties of the underlying objects is used. Thus, the motion is determined using only a set of rules that describe the desired properties of the resulting motion field. Most of the approaches in the above mentioned references are guided by the features of the application in mind and use a combination of these methodologies.

For this work the vocal folds are first detected using active contour models (snakes).[9] Then, the deformation of the vocal folds is found by elastically registering the contours of the vocal folds from frame-to-frame. The main characteristics of this elastic registration problem are: (1) There are no distinct features in the images that can be used as tokens to help track the motion from frame-to-frame. (2) The mechanical properties of the vocal folds are not yet very well understood, thus it is hard to write down equations that describe their motion. Therefore, the third methodology has to be used for finding non-rigid motion for this problem. (3) The registration algorithm which will be used must be able to accommodate large deformations that appear in the data due to the severe temporal sub-sampling that occurs in videostroboscopy.

Recently, combinatorial optimization algorithms have been employed to find the motion when only the third methodology is used. A method based on dynamic programming and auto-regressive modeling was proposed in Ref. 3 This approach was modified and used

successfully in Ref. 4 for the problem of motion estimation of skeletonized angiographic images. Another approach was also proposed in Ref. 5. Unlike the approach in Ref. 3 the approach in Ref. 5 is based entirely on dynamic programing. We found that the algorithm in Ref. 4 and our implementation of the algorithm in Ref. 5 experience difficulties when applied to contours that have been severely deformed. For this purpose we proposed a new elastic contour matching algorithm[10] that ameliorates the difficulties that we encountered when applying the algorithms in Refs. 4,5. The proposed algorithm is based on a cost function which is minimized using *simulated annealing*.[11] Simulated annealing is an optimization algorithm that is able to obtain near-optimal solutions for a wide variety of combinatorial optimization problems. The definition of the cost function is based on the principle of regularization, see for example Refs. 12,13, and thus trades-off between two requirements. First, smoothness of the displacement vector field and second that the overall displacement is small.

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 Once the displacement vector field is found, the need to model it quantitatively arises. Although the displacement vector field 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* model.[14] In this paper, a least squares procedure is developed to fit the affine transformation model to the displacement vector field between successive frames. The obtained affine transformation 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.

It is well known that the vocal fold vibration consists of the inferior portion (in a coronal plane) of a vocal fold moving approximately 50 degrees out of phase with the superior portion. This vertical phase difference means that during part of the closing phase of a glottal cycle, the bottom part of the vocal folds are closer together than the top. Problems with adequate illumination make reliable visualization and measurement difficult for the motion of the inferior portion of the vocal fold. Consequently for the purposes of this study measurements were limited to the *superior margin of the glottal surface* that could be easily visualized.

The rest of this paper is organized as follows. In section 2, we present the snakes implementation that we used to delineate the vocal folds in the videostroboscopic data. In section 3, we present the new elastic registration algorithm that was used for the registration of the vocal fold contours from frame-to-frame. In section 4, we present the modeling of the vocal fold deformations using the displacement vector field and the affine transformation model. In section 5, we present our experimental results. Finally, in section 6, we present our conclusions from this work and our future research directions.

2 BOUNDARY DETECTION OF VOCAL FOLDS USING SNAKES

The delineation of the vocal folds can be viewed as the combination of two tasks. 1. finding the locations of abrupt intensity changes, edges, that signify the presence of the folds; 2. ordering the edges to form connected curves or edge linking. There is an enormous amount of research in edge detection and edge linking, see for example Refs. 15,16. The approaches that exist to handle these tasks can be classified into two broad categories. 1. *data-driven* approaches,[17] where an edge detector operates directly on image intensities. The

problem with these approaches is that errors made in edge detection propagate to edge linking without any opportunity for correction. 2. *model-driven* approaches,[18] where a parametric expression for the curve is assumed. The main advantage of this technique is that it is insensitive to noise and gaps in curves. However, the quantization of the image and the parameter space used influence the final outcome. Furthermore, this method does not use any information about the local structure of the image.

Active contour models (snakes) is a method that combines the advantages of both previous approaches.[19,20] The main difficulty in applying snakes to different problems is in selecting the initial positions for them. For our application, this initialization problem can be solved by user interaction in the first frame. For the following frames, the snake position found in the preceding frame can be used as the initial position for the current frame.

A snake is an energy-minimizing spline guided by external forces and is influenced by image forces that pull it towards features such as lines and edges. Snakes are active contour models which lock on to nearby edges localizing them accurately. *Snakes* were originally proposed in Ref. 9 in which an energy functional was defined for the contour and the minimum energy contour was determined using variational calculus. A snake, is represented by a vector $\mathbf{v}$ whose elements are functions of two spatial coordinates x and y, i.e. $v(i) = (x(i), y(i))$ are the locations on the snake and are called snake elements, or *snaxels*. A snake is *closed* if $\mathbf{v}$ forms a closed curve, otherwise it is *open*. The total snake energy, E_{snake}, consists of a sum of two energy terms:

$$E_{snake}(\mathbf{v}) = \sum_{i=1}^{n} E_{snake}(i) = \sum_{i=1}^{n} (\lambda E_{int}(i) + (1-\lambda) E_{ext}(i)) \ , \qquad (1)$$

where $E_{int}(i)$ and $E_{ext}(i)$ are the internal and external energies respectively at snaxel $v(i)$. This definition of the energy function complies with the principle of regularization. According to this principle fidelity to the data E_{ext} and prior knowledge E_{int} are combined to resolve

the ambiguities of imperfect data in estimation problems, see for example Refs. 12,13. The parameter $\lambda \in [0,1]$ is the *regularization parameter* that determines the relative trade-off between data fidelity and prior knowledge which for this problem are captured by the terms E_{ext} and E_{int}, respectively.

We seek the optimum locations $v(i), i = 1, 2, \cdots, n$, that minimizes the snake energy in Eq.(1). The internal energy E_{int} imposes a model of smooth curves which restricts the solutions to the class of controlled continuity splines; the external energy E_{ext} causes the snake to attach itself to salient features in the image. The combined effect of this two-energy term yields a model that allows itself to deform in conformation to the nearest salient features. In general, E_{int} consists of a continuity term E_{cont} and a curvature term E_{curv}. E_{cont} encourages snaxels to be evenly spaced and eliminates the tendency of shrinkage and expansion. E_{curv} imposes the smoothness constraint on the snake. For the external energy, both the magnitude and direction of intensity gradient are used in the formulation of E_{ext}.

We use similar definitions as in Refs. 19,20; the energy functional thus yields the following $E_{snake}(i)$:

$$\begin{aligned} E_{snake}(i) &= \lambda E_{int}(i) + (1-\lambda) E_{ext}(i) \\ &= \lambda(\gamma E_{cont}(i) + (1-\gamma) E_{curv}(i)) + (1-\lambda) E_{ext}(i) \ , \end{aligned} \tag{2}$$

where λ, $\gamma \in [0,1]$.

The regularization parameter λ weighs the importance of the internal and the external energy terms. For the values of $\lambda \gg (1-\lambda)$, the E_{int} is dominant. This restricts the snake to concur more with the *model-driven* approach, while for the values of $\lambda \ll (1-\lambda)$, the E_{ext} is dominant. This causes the snaxels to settle at locations of high intensity gradient, causing the snake to behave like an edge linking algorithm. This concurs more with the

data-driven approach. In the continuum of $\lambda \in [0,1]$, the snake exhibits an intermediate behaviour between the *model-driven* and *data-driven* approaches.

3 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 pair of indices $(m(i), k(i))$ for $i = 1,\ 2,\ ..\ L_2$ where L_2 is the larger contour, that describe 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 displacement vector field. The displacement vector field describes the elastic deformation of the two contours.

An elastic contour registration method based on dynamic programming [21] and auto-regressive modeling was proposed in Refs. 3,4. The cost function to be minimized using dynamic programming is chosen to be a function of the estimated (using auto-regressive models) displacement vector and the actual displacement vector found. The advantage of this algorithm is that the complexity of the deformations does not affect the computational cost. However, the main limitations of this approach for our application are the following: First, it does not guarantee that all pixels of the smaller contour will correspond to points in the larger contour. This will affect the motion estimation in the next stages. Sec-

ond, there is a possibility of crossovers between adjacent displacement vectors when severe shrinking/expansion is present.

Another algorithm that also matches deformed contours was proposed in Ref.5. Unlike the approach in Refs. 3,4, this approach is based entirely on dynamic programming. However, instead of choosing as the 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. In order to keep the implementation cost at a reasonable level, dynamic programming is applied to a sub-graph which is based on a neighborhood system which is defined around every point of the largest contour. The main drawback of this approach for our application is the choice of the cost function which makes this method unsuitable for matching contours which are different in size. This is exactly the case when severe shrinking/expansion is present from contour-to-contour. In other words, one-to-one matching leaves unmatched points when the two contours are not of equal sizes. Finally, this approach requires that both contours are described as an ordered list of points.

In this section, we present a original elastic contour matching algorithm[10] that alleviates the difficulties of the algorithms in Refs. 4,5 for our application. The proposed algorithm is based on a cost function that is minimized using *simulated annealing*.[11] The simulated annealing algorithm is a general optimization technique for solving combinatorial optimization problems. The algorithm is based on *randomization techniques*. However, it also incorporates a number of aspects related to *local search* techniques. In contrast to local search techniques the simulated annealing algorithm finds high-quality solutions which do not strongly depend on the choice of the initial solution, *i.e.* the algorithm is *effective* and *robust*. Furthermore, the simulated annealing algorithm can escape from local minima while it still exhibits the

favorable features of local search algorithms, *i.e.* simplicity and general applicability. Due to the highly non-convex nature of the function that is minimized an optimization algorithm which is based on a random search would be appropriate for this problem. Simulated annealing is such an algorithm.

Finding the displacement vector field is an *ill-posed* problem and it is common in all ill-posed problems to impose simultaneously a small square error and a smoothness constraint. This is the very well known *regularization principle*.[12,13] A cost function based on the regularization principle is defined. This cost function is minimized with respect to the choice of the displacement vectors $(m(i), k(i))$ for $i = 1, 2, \cdots L_2$ and is of the form

$$\begin{aligned} C = \quad & (1-\lambda) \quad \sum_{i=1}^{L_2} [\ (x_1(m(i)) - x_2(k(i)))^2 + (y_1(m(i)) - y_2(k(i)))^2\ ] \\ & +\lambda \quad \sum_{i=1}^{L_2} [|\ \phi(m(i), k(i)) - \phi(m(i-1), k(i-1))\ |]^2\ , \end{aligned} \tag{3}$$

where the parameter $\lambda \in [0, 1]$ is the regularization parameter and $\phi(m(i), k(i))$ 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 displacement vector field.

3.1 Simulated annealing

Simulated annealing has been applied in the past to a number of non-convex combinatorial optimization problems in image processing, see for example Ref. 22. The name simulated annealing originates from the analogy with the physical annealing process of solids.

Simulated annealing associates a " temperature " T, with a model and searches for the optimal configuration based on a defined energy by repeatedly trying to alter the state of the system. At each *iteration*, a transition is first generated and then either accepted or rejected. A rejected transition is regarded as a transition from the current state to itself. By appropriately selecting the probability of generating and accepting state j from state i, the simulated annealing algorithm can be shown to converge,[23] for any value of the temperature $T > 0$, after a (possibly large) number of transitions. The temperature is then lowered repeatedly, and after enough iterations the system approaches equilibrium. This process is terminated using a stopping rule when the change of the expected cost is very small compared to the expected cost at T_0, see for example Refs. 10,11.

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.

We used the rule for lowering the temperature described in Ref. 11, the temperature T_{k+1} is given by

$$T_{k+1} = \frac{T_k}{1 + \frac{T_k \cdot ln(1+\delta)}{3\sigma_{T_k}}}, \qquad k = 0, 1, \cdots \quad , \tag{4}$$

where δ is a small positive number. Small δ-values lead to small decrements of the temper-

ature. σ_{T_k} is the standard deviation of the cost function at T_k where

$$\sigma_{T_k} = (1/L \sum_{i=1}^{L_{T_k}} (C_{T_k}(i) - \overline{C_{T_k}})^2)^{1/2}, \text{ and } \overline{C_{T_k}} = 1/L \sum_{i=1}^{L_{T_k}} C_{T_k}(i) \ ,$$

$C_{T_k}(i)$ is the value of the cost function at state i when the temperature is T_k, the bar denotes averaging over all values of the cost function $C_{T_k}(i)$, for $i = 1, 2, \cdots, L_{T_k}$ generated at a given temperature T_k.

The algorithm terminates when, for some k, we have $\Delta\overline{C}_{T_k}$ is small compared to $\overline{C}_{T_0}$,[11] *i.e.*

$$\frac{T_k}{\overline{C}_\infty} \frac{\partial \overline{C}_T}{\partial T} |_{T=T_k} < \varepsilon_s \ , \tag{5}$$

where ε_s is some small positive number called the *stop parameter*, $\overline{C}_\infty$ is the expected cost at $T \to \infty$ and $\frac{\partial \overline{C}_T}{\partial T}|_{T=T_k}$ is the partial derivative of the expected cost at $T = T_k$.

Let T_k denote the value of the temperature, L_{T_k} the length of the Markov chain when the temperature is T_k, i_0 the initial state (configuration). Then, our registration algorithm can be described in pseudo code as follows:

begin

INITIALIZE(i_0, T_0)

$k := 0$;

$i := i_0$;

repeat

for $l := 1$ **to** L_{T_k} **do**

begin

GENERATE (new state j);

if $C(j) \leq C(i)$ **then** $i := j$

else

if $\exp(\frac{C(i)-C(j)}{T_k}) >$ random[0, 1) **then** $i := j$

end;

$k := k + 1$;

CALCULATE (T_k) Eq.(4);

until stop criterion Eq.(5)

end;

We used $L_{T_k} = L_1$ for $k = 1, 2, \cdots$. The initial state i_0 is selected randomly. We start at an initial value of the temperature T_0 evaluated using the same approach as in Ref. 11. The **GENERATE** (new state j) picks at random at state i two adjacent points from the larger contour, which are the origins of two displacement vectors. Then the new state j is generated by exchanging the displacement vectors of the two points previously selected.

4 AFFINE TRANSFORM MODELING OF THE DEFORMATION

After elastic registration, the displacement vector field between adjacent frames of the vocal folds is available. The displacement vector field provides a better visualization of the deformation than the raw video-recordings. However, the displacement vector field 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 displacement vector field.

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 displacement vector field between frames 1 and 2 is given by the collection of all available displacement vectors. The displacement vector field 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 displacement vector field that is very appropriate because it can capture the deformation during the motion of the vocal folds is based on the *affine transformation*.[14,24] This model is described by the following set of linear equations

$$\begin{bmatrix} x_2(i) \\ y_2(i) \end{bmatrix} \approx \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x_1(i) \\ y_1(i) \end{bmatrix} + \begin{bmatrix} T_x \\ T_y \end{bmatrix} \text{ for } i = 1, 2, \cdots, L. \tag{6}$$

The $\approx$ sign is used because fitting an affine transformation 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*,[14] 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,[25] where D is a symmetric positive definite matrix that represents deformation and R is an orthonormal matrix. More specifically

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = R \cdot D, \ R = \begin{bmatrix} cos\theta & -sin\theta \\ sin\theta & cos\theta \end{bmatrix} \text{ and } D = \begin{bmatrix} a & c \\ c & b \end{bmatrix} . \tag{7}$$

Because of the symmetry of D we can also write

$$D = \begin{bmatrix} a & c \\ c & b \end{bmatrix} = \begin{bmatrix} \vec{\mathbf{e}}_1^T \\ \vec{\mathbf{e}}_2^T \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \begin{bmatrix} \vec{\mathbf{e}_1}, \vec{\mathbf{e}_2} \end{bmatrix} , \tag{8}$$

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

$$A^T A = (RD)^T (RD) = D^T R^T R D, \tag{9}$$

but, since R is orthonormal matrix then,

$$R^T R = I, \tag{10}$$

where I is the identity matrix. Therefore,

$$A^T A = D^T D. \tag{11}$$

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.

$$A^T A = \lambda_1^2 \vec{\mathbf{e}_1} \vec{\mathbf{e}_1}^T + \lambda_2^2 \vec{\mathbf{e}_2} \vec{\mathbf{e}_2}^T. \tag{12}$$

Then

$$D = \lambda_1 \vec{\mathbf{e}_1} \vec{\mathbf{e}_1}^T + \lambda_2 \vec{\mathbf{e}_2} \vec{\mathbf{e}_2}^T. \tag{13}$$

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.[14] These directions are given by the angles

$$\phi_1 = tan^{-1}(\frac{\mathbf{e}_1(y)}{\mathbf{e}_1(x)}), \text{ and } \phi_2 = tan^{-1}(\frac{\mathbf{e}_2(y)}{\mathbf{e}_2(x)}) , \tag{14}$$

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 affine transformation model of the displacement vector field provides a succinct description of the changes that occurred in the shapes of the contours.

The $\approx$ sign in Eq.(6) 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.(6) 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* approach is used.[26] 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.(6) for $i = 1, 2, \cdots L$ we can write the following system of linear equations $Bx = b$ with

$$B = \begin{bmatrix} x_1(1) & y_1(1) & 0 & 0 & 1 & 0 \\ 0 & 0 & x_1(1) & y_1(1) & 0 & 1 \\ x_1(2) & y_1(2) & 0 & 0 & 1 & 0 \\ 0 & 0 & x_1(2) & y_1(2) & 0 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_1(L) & y_1(L) & 0 & 0 & 1 & 0 \\ 0 & 0 & x_1(L) & y_1(L) & 0 & 1 \end{bmatrix}, \tag{15}$$

$$x = \begin{bmatrix} a_{11}, & a_{12}, & a_{21}, & a_{22}, & T_x, & T_y \end{bmatrix}^T, \tag{16}$$

$$b = \begin{bmatrix} x_2(1), & y_2(1), & x_2(2), & y_2(2), & \cdots & x_2(L), & y_2(L) \end{bmatrix}^T, \tag{17}$$

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.

5 EXPERIMENTAL RESULTS

In this section we present two categories of experiments. In the first, the performance of the new simulated annealing-based elastic registration approach that we propose is compared to previous elastic registration algorithms that are based on the same methodology. To the best of our knowledge two such algorithms exist, more specifically, the algorithms in Refs. 4 and 5. In the second, the proposed system (contour delineation, elastic registration, deformation modeling) is tested on videostroboscopic recordings of normal larynges, a larynx with a polyp, and a larynx with a cyst.

For the first category, experiments are presented to test the new registration algorithm on contours produced by synthetic deformations since in order to evaluate the algorithms' performances quantitatively, the true displacement vector field must be known. More specifically, the three registration algorithms from Refs. 4,5 and section 3 were compared. For the algorithm in Ref. 4 the authors of the paper provided us with their code. The algorithm in Ref. 5 was modified to match all the points of the larger contour. For our modification, the algorithm was reapplied in the regions were points were left unmatched. All three algorithms were applied to contours that have been deformed in a controlled manner by an affine transformation, given by

$$D = \text{diag}\{d, d\} \quad d < 1.0 \ , \tag{18}$$

and a θ° rotation, see Eq.(7). This affine transformation when applied to the outer contour

of Fig.(1) results in the deformed inner contours, see Fig.(1). Notice that D "squeezes" the initial contour in both the x and y directions.

For the simulated annealing that was used in our elastic registration algorithm,[10] we chose a cooling schedule defined by three parameters, i.e. the initial acceptance ratio χ_0, the δ parameter, and the stop parameter ε_s. In the previous, and all of the following experiments we used $\chi_0 = 0.98$, $\delta = \frac{T_0}{500}$, and $\varepsilon_s = 10^{-8}$. Choosing $\chi_0 \approx 1$ sets the initial temperature T_0 of the system at very high value, while, δ which controls the decrease rate of the temperature in Eq.(4) is relatively small, and finally the stop parameter ε_s is chosen to be very small.

A mean square error metric, E_{MSE}, in pixels for the simulated elastic matching problem was used to objectively evaluate the results. E_{MSE} is defined by

$$E_{MSE} = 1/N \sum_{i=1}^{N} [(x_2(i) - \hat{x}_2(k(i)))^2 + (y_2(i) - \hat{y}_2(k(i)))^2] \ , \tag{19}$$

where $(x_2(i), y_2(i))$ are the true positions computed from Eq.(6) and $(\hat{x}_2(k(i)), \hat{y}_2(k(i)))$ were found from the elastic matching of contours 1 and 2.

We applied the three algorithms for different deformations. Fig.(1) shows some of these deformations of the outer contour. Applying the three algorithms for $d = 0.5, \theta = 0^\circ$ yields the displacement vector fields of Fig.(2). Registration errors of the first two algorithms can be seen where crossings of motion vectors occurred in Fig.(2)(a) and Fig.(2)(b). In Fig.(2)(c) the resulting displacement vector field from the proposed approach is shown. Using the definition in Eq.(19) with $N = 303$ the E_{MSE} for the three algorithm in Refs. 4,5 and for the proposed approach for different deformations are shown in Table(1).

The previous experiments and for $d = 0.5, \theta = 0^\circ$ were run on a SUN/SPARC-10 workstation, the processing time for the algorithm in Ref. 5 was 50 seconds, for the algorithm

in Ref. 4 was 55 seconds and for the proposed approach it took 390,870 iterations in 69 seconds. For the other experiments similar numbers in terms of times and iterations were observed. These simple experiments indicate that the proposed approach has the ability to track better elastic contours during deformation and computationally it is not much more expensive.

In the second category, experiments are presented that test the proposed algorithms on real data of videostroboscopic image sequences of a normal and abnormal larynges. 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.(3), 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 described in section 2. It can be seen that the positions and changes in the shape of the vocal folds are correctly extracted using the snakes. In locating the contours of the vocal folds we used a small number (30-40) of snaxels, then the complete boundary was obtained by applying cubic spline interpolation[27] to the snaxels at the right and the left side of the vocal folds separately. For our application we chose the value of $\gamma = 0.5$ and the value of $\lambda = 0.8$ this yields more of a model-driven solution and is robust to noise. The displacement vector fields of all contours for the successive frames are shown in Fig.(4). From this figure and in frames 7-10 it is observed that some points on the vocal fold contour move together while others move oppositely. This seems to happen only around the time of maximum opening and the "turn-around" toward closure.

Because of the frame rate limitations of videostroboscopy, as explained in the introduction, the frames in an apparent videostroboscopic cycle originate from a number depending on

the frequency of phonation. For example fundamental frequency of 140 Hz at 33 frames/sec would sample every 4^{rth} cycle. Therefore, there are instances where the visible anterior closure of the vocal folds shows a large descrepancy, large lateral distances between adjacent frames, as seen for example in Fig.(4) (displacement vector fields of frames $2 \rightarrow 3$, $3 \rightarrow 4$, $4 \rightarrow 5$ and $15 \rightarrow 16$); therefore the line connecting the most anterior aspects of the openings between adjacent frames was used as the reference for reconstructing the glotal contour for the subsequent motion analysis.

This correction is based on the observation that the structural motion is medial-lateral (horizontal in the images) rather than anterior-posterior and is consistent with high-speed photography descriptions of the vocal folds vibratory patterns.[28,29] In order to have anterior-posterior motion there would need to be concomitant shortening at the ventricular folds and other visible structures which was not the case. Therefore, it is safe to assume in these images that all the missing motion was medial-lateral and that the line correction enhanced the motion interpretation analysis of the data. This is not to discount vertical movement. However, while the vertical position is continuously changing, over the one second representing a cycle the motion is negligible as evidensed by no change in the size of the true vocal folds or the position of the supraglotic structure.

In order to evaluate the performance of the affine transformation parameters in modeling the displacement vector and characterizing the deformation of the vocal folds, a prediction mean-square-error metric was used, the E_{pmse} is defined by:

$$E_{pmse} = 1/N \sum_{i=1}^{N} \|d(i) - \hat{d}(i)\|^2 \tag{20}$$

where N is the number of displacement vectors, $d(i) = (x_2(i) - x_1(i), y_2(i) - y_1(i))$ is the displacement vector found from the elastic registration of contours 1 and 2, and $\hat{d}(i) =$

$(\hat{x_2}(i) - x_1(i), \hat{y_2}(i) - y_1(i))$ is the displacement vector predicted by

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

where $(a_{11}, a_{12}, a_{21}, a_{22}, T_x, T_y)$ are the affine transformation parameters obtained for this displacement vector field. Using the definition in Eq.(20), the E_{pmse} for the displacement vector fields in Fig.(8) is shown as an example in Table(2) for the larynx with a polyp. The values of the E_{pmse} are small which indicates the ability of the affine transformation model to capture the deformation of the vocal folds. This sequence was chosen as an example to demonstrate the ability of the affine transformation model to capture the motion information even when it is irregular.

Fig.(5) (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 x-axis, is very small. For this sequence, a larger deformation occurs along the horizontal direction (y-axis), i.e. $\phi_1 \approx 90$. The time intervals for which $\lambda_1 > 1$ defines the opening cycle of the vocal folds. From Fig.(5)(a) we observe that the opening cycle lasts until frame 8 after which the vocal folds start the closing cycle ($\lambda_1 < 1$). Also from Fig.(5)(a) we observe that the maximum relative opening movement takes place between frames 3 and 4 and the maximum relative closing movement takes place between frames 14 an 15. These observations cannot be made from the raw images in Fig.(3) or from the displacement vector field in Fig.(4). For this sequence, as shown in Fig.(5)(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. By observing the images and the motion vectors, which are not shown in this paper due to space limitations, we concluded

that the sudden change in the curves of the angles ϕ_1 and ϕ_2 in Fig. (5) (b) around frames 8-10 is probably due to patient or camera motion.

Another sixteen frame videostroboscopic image sequence of a normal larynx was analyzed and the deformations are shown in Fig.(6). From Fig.(6)(a) and as in the previous sequence, the opening cycle ($\lambda_1 > 1$) lasts until frame 8, after which the vocal folds start the closing cycle ($\lambda_1 < 1$). Similar comments as for the previous sequence can be made here about the curves for ϕ_1 and ϕ_2.

In another experiment, we used sixteen frames from a videostroboscopic image sequence of a patient with polyps.[2] In Fig.(7) 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. Fig.(8) shows the displacement vector fields between each two successive contours of the vocal folds using the new algorithm described in section 3. The translational and rotational components of the deformation are again very small, see Fig(9)(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.(9)(a). From Fig.(9)(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 the opening movement until frame 6 ($\lambda_1 > 1$) then an unexpected closing movement 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 a videostroboscopic image sequence of a patient with a cyst.[2] The translational and rotational components of the deformation are again very small, see Fig(10)(c), and (d). From this figure we can also tell that the vocal folds kept the opening movement until frame 7 ($\lambda_1 > 1$) then an unexpected closing

movement occurred between frames 8 and 9, then they continued to open until frame 11.

6 CONCLUSIONS

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

We found that the affine transformation describes well the deformations of the vocal folds. The proposed affine transformation model is quite general and can be applied in a number of different ways. For example, when the symmetry of the vocal fold motion is examined then, two affine transformation models one for the left and one for the right side of the vocal folds should be used. In the experiments we show we 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. However, there exist scenarios where the other parameters of the affine transformation might be significant. One such scenario is when global motion is present.

Although videostroboscopy is considered to be one of the most valuable clinical tools available for assesing vocal fold vibrations it is not without problems. It produces an averaged signal and does not provide detail about any single cycle. It is based on assumptions that the vocal fold vibration is fairly regular which frequently is not the case.

Recently high-speed digital video asystem that has neither of these limitations has been used to investigate non-periodic vibration characteristics of the vocal folds. With such systems recordings are made at 1000 frames/sec or faster so that in contrast to recordings made at 30 frames/sec, many video frames are captured during one true rather than one apparent glottal cycle removing the periodicity limitation of stroboscopy.[30] However, the methods described in this paper could be applied to high-speed video images thus providing more powerful interpretations of vocal fold vibrations.

Acknowledgements: This work was supported by the Whitaker foundation. The authors of this paper are grateful to the authors of Ref. 4 for kindly providing them with the code for their registration algorithm and to the authors of Ref. 20 for providing them with a preprint of their paper.

7 References

[1]D. Bless, M. Hirano, and R. Feder, "Videostroboscopic evaluation of the larynx," Ear Nose & Throat Journal, **66**(7) (1987).

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

[3]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, **37**(2), 288-297 (1989).

[4]B. Tom, S. N. Efstratiadis, and A. K. Katsaggelos, "Motion estimation of skeletonized angiographic images using elastic registration," IEEE Transactions on Medical Imaging, **13**, 450-460 (1994).

[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, **17**(3), 294-302 (1995).

[6]J. S. Duncan, R. L. Owen, L. H. Staib and P. Anandan, "Measurement of non-rigid motion using contour shape descriptors," Computer Vision and Pattern Recognition, 318-324 (1991).

[7]P. Shi, G. Robinson, R. T. Constable, A. Sinusas and J. Duncan, "A Model-based integrated approach to track myocardial deformation using displacement and velocity constraints," International Conference on Computer Vision, 687-692 (1995).

[8]C. Nastar and N. Ayache, "Non-rigid motion analysis in medical images: a physically based approach," 13th International Conference, Information Processing in Medical Imaging, 17-32 (1993).

[9]Kass, A. Witkin and D. Terzopoulos, "Snakes: active contour models," Proc. 1st Int Conf.

on Comp. Vision, 259-269 (1987).

[10]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, Nevada, 93-96 (1995).

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

[12]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, **1**(3), 322-336 (1992).

[13]D. Terzopoulos, "Regularization of inverse visual problems involving discontinuities," IEEE Trans. Pattern Analysis Mach. Intell., **8**(4), 413-424 (1986).

[14]E. A. Cemal, Continuum Physics, (Academic Press, 1976).

[15]J. Canny, "A computational approach to edge detection," IEEE Trans. Pat. Anal. Mach. Intell., **8**, 679-698 (1986).

[16]A. Martelli, "Edge detection using heuristic search methods," Computer Graphics Image Processing, **1**, 169-182 (1972).

[17]D. Marr and H. K. Nishihara, "Visual information processing: artificial intelligence and the sensorium of sight," Technology Reviews, **81**(1), 2-23 (1978).

[18]R. O. Duda and P. E. Hart, "Use of the Hough transformation to detect lines and curves in pictures," Communications of the Association for Computing Machinery, **15**, 11-15 (1972).

[19]Kok F. Lai and R. T. Chin, "On regularization, formulation and initialization of the active contour models (Snakes)," Asian Conference of Computer Vision, 542-545 (1993).

[20]Kok F. Lai and R. T. Chin, "On modeling, extraction, detection and classification of deformable contours from noisy images," Journal of Image and Vision Computing, in review.

[21]R. E. Bellman, "Applied Dynamic Programming," (Princenton University Press, 1962).

[22]S. Geman and D. Geman, "Stochastic relaxation, Gibbs distribution, and Bayesian restoration of images," IEEE Trans. Pattern Analysis Mach. Intell., **6**(6), 721-741 (1984).

[23]N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller and E. Teller, "Equation of state calculations by fast computer machines," Journal of Chemical Physics, **21**, 1087-1092 (1953).

[24]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, **12**(1), 26-39 (1993).

[25]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., 63-99, Greenwich, Conn.: JAI Press (1988).

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

[27]B. P. Flannery, W. H. Press, S. A. Teukolsky and W. T. Vetterling, Numerical Recipes in C, (Cambridge University Press, 1988).

[28]R. Timcke, H. vonLeden and P. Moore, "Laryngeal vibrations:II. Physiologic vibrations. Archives of Otolaryngology," **69**, 438-444 (1959).

[29]H. vonLeden, P. Moore and R. Timcke, "Laryngeal vibrations: III. The pathologic larynx. Archives of Otology," **71**, 1232-1250 (1960).

[30]T. Wittenberg, M. Moser, M. Tigges and U. Eysholds, "Recording, processing and analysis of digital high-speed sequences in glottography," Machine Vision and Applications, **8**, 399-404 (1995).

d	θ	E_{MSE} for the algorithm		
		in Ref. 7	in Ref. 5	using SA
0.85	0°	2.67	3.62	1.42
0.8	0°	3.85	4.59	1.49
0.7	0°	6.14	11.49	1.53
0.5	0°	7.18	14.27	1.70

Table I: The E_{MSE} for the three algorithms for the different deformations in Fig.(1).

frames	E_{pmse}	*frames*	E_{pmse}
$1 \rightarrow 2$	0.68	$2 \rightarrow 3$	1.70
$3 \rightarrow 4$	0.94	$4 \rightarrow 5$	0.76
$5 \rightarrow 6$	1.10	$6 \rightarrow 7$	1.25
$7 \rightarrow 8$	1.20	$8 \rightarrow 9$	0.84
$9 \rightarrow 10$	1.37	$10 \rightarrow 11$	0.25
$11 \rightarrow 12$	0.36	$12 \rightarrow 13$	0.14
$13 \rightarrow 14$	0.22	$14 \rightarrow 15$	1.26
$15 \rightarrow 16$	0.72		

Table II: The E_{pmse} of the affine transformation model based on prediction of the displacement vector field of the sequence with the polyps in Fig.(7).

8 Figure Captions

Figure 1: The inner contour is obtained by deforming the outer contour using the matrix D in Eq.(18) and a rotation $\theta = 0°$, see Eq.(7) for (a) $d = 0.8$, (b) $d = 0.7$, (c) $d = 0.5$.

Figure 2: The DVF registration for $d = 0.5, \theta = 0°$ using the (a) algorithm in Ref. 7, (b) algorithm in Ref. 5, (c) proposed regularization based approach, (d) true displacement vector field.

Figure 3: The bright closed contours delineate the boundaries of the vocal folds in successive frames of a normal larynx run from top left to the bottom right.

Figure 4: The registration of the successive boundaries of the sequence of videostroboscopic images of a normal larynx using the proposed simulated annealing based approach runs from top left to the bottom right, the displacement vector fields are sampled and magnified.

Figure 5: 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 larynx videostroboscopic image sequence respectively. The units of the vertical axes in (b) and (d) are degrees.

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 second normal larynx videostroboscopic image sequence respectively. The units of the vertical axes in (b) and (d) are degrees.

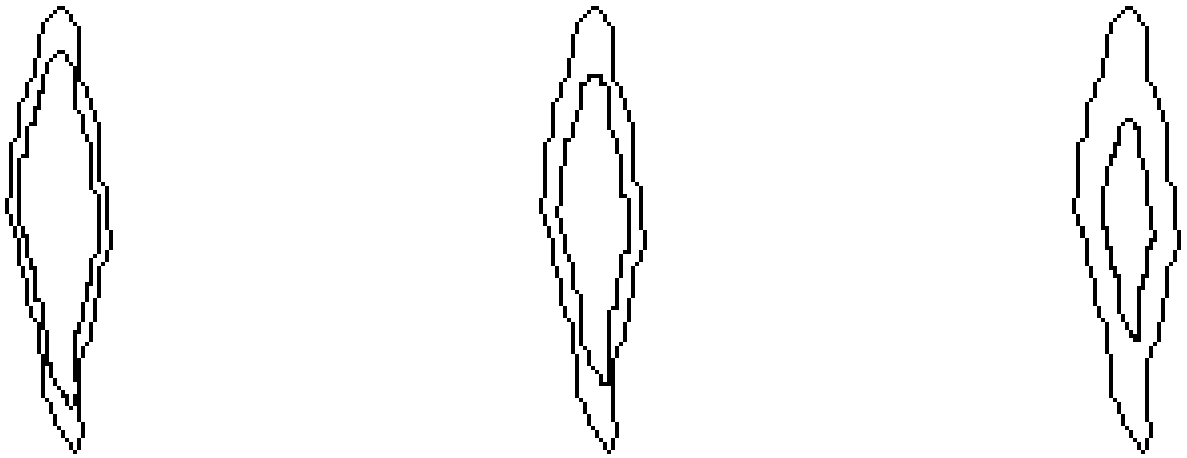
Figure 7: The bright closed contours delineate the boundaries of the vocal folds in the successive frames of the first abnormal larynx sequence for a patient with polyps run from top left to the bottom right.

Figure 8: The registration of the successive boundaries of the first abnormal larynx sequence of videostroboscopic images using the proposed SA based approach runs from top left to the bottom right, the displacement vector fields are sampled and magnified.

Figure 9: 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 larynx videostroboscopic image sequence respectively. The units of the vertical axes in (b) and (d) are degrees.

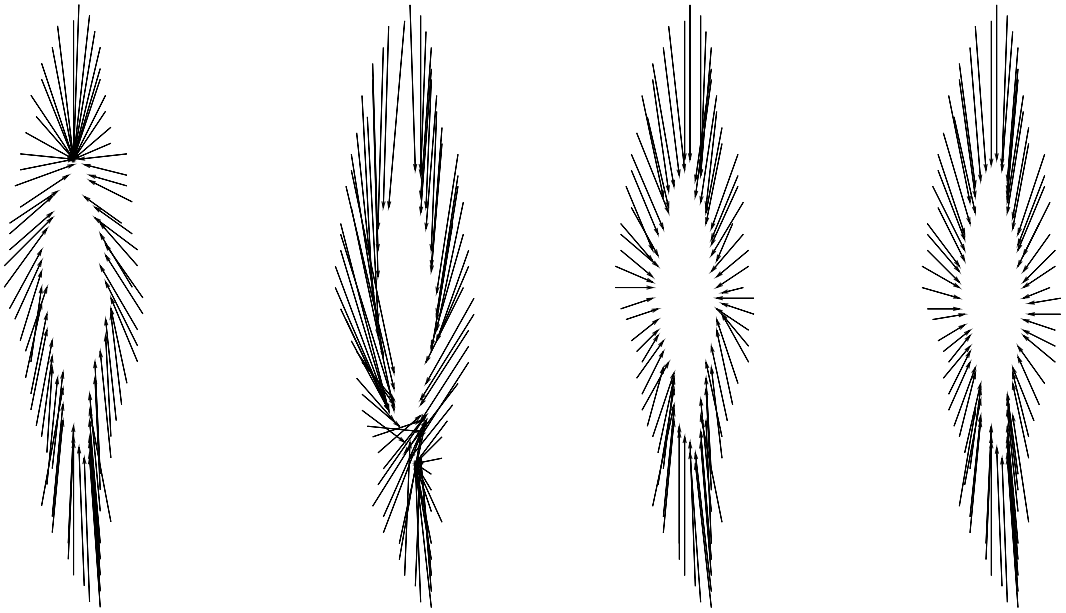
Figure 10: The figures in (a), (b), (c) and (d) represent time evolution of the eigenvalues,

principal directions of deformation, the translation along the x and y axes and the rotation angle for the frames of the second abnormal larynx videostroboscopic image sequence respectively. The units of the vertical axes in (b) and (d) are degrees.



(a) (b) (c)

Figure 1: The inner contour is obtained by deforming the outer contour using the matrix D in Eq.(18) and a rotation $\theta = 0°$, see Eq.(7) for (a) $d = 0.8$, (b) $d = 0.7$, (c) $d = 0.5$.



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

Figure 2: The DVF registration for $d = 0.5, \theta = 0°$ using the (a) algorithm in Ref. 7, (b) algorithm in Ref. 5, (c) proposed regularization based approach, (d) true displacement vector field.

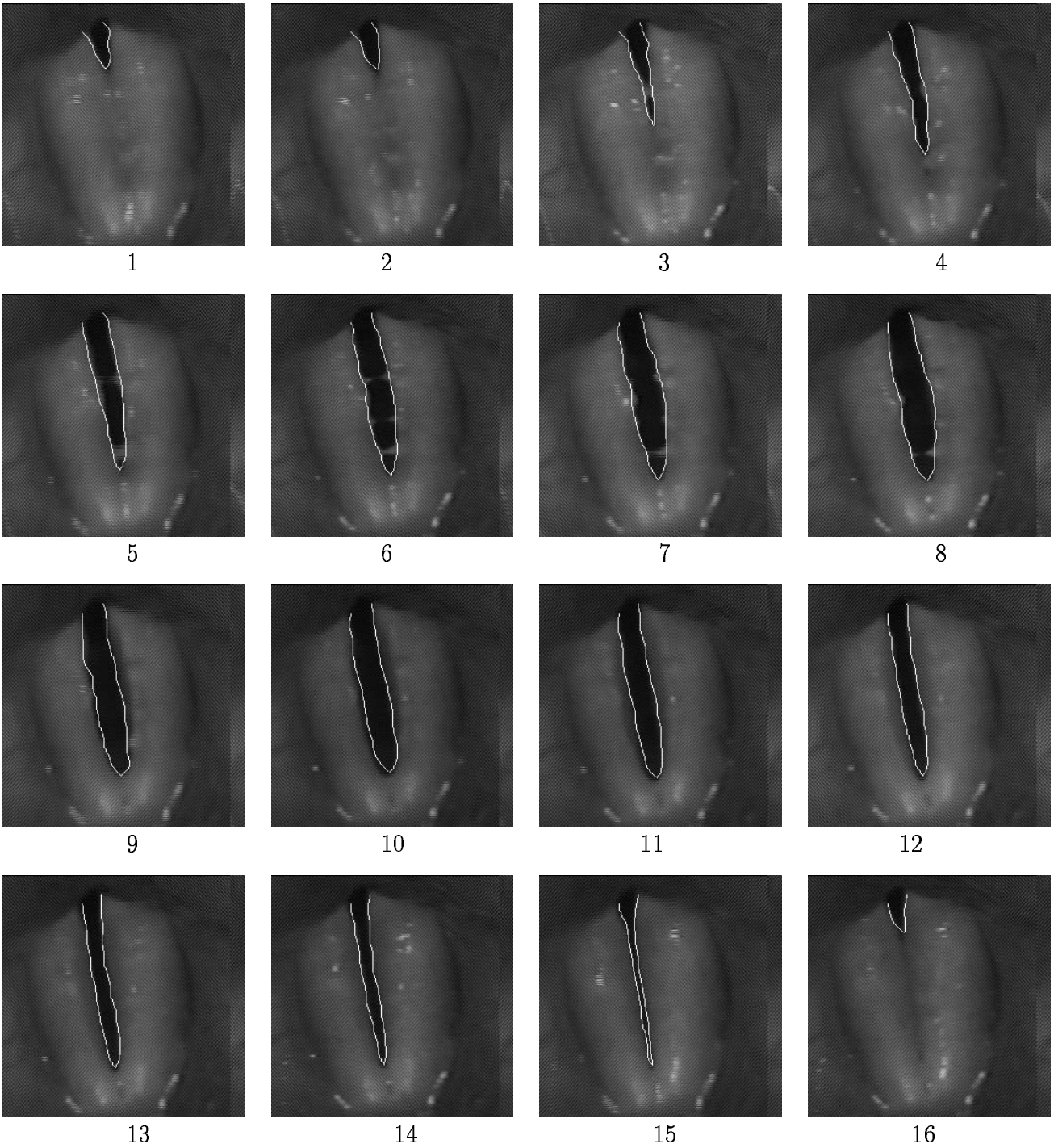


Figure 3: The bright closed contours delineate the boundaries of the vocal folds in the successive frames of a normal larynx run from top left to the bottom right.

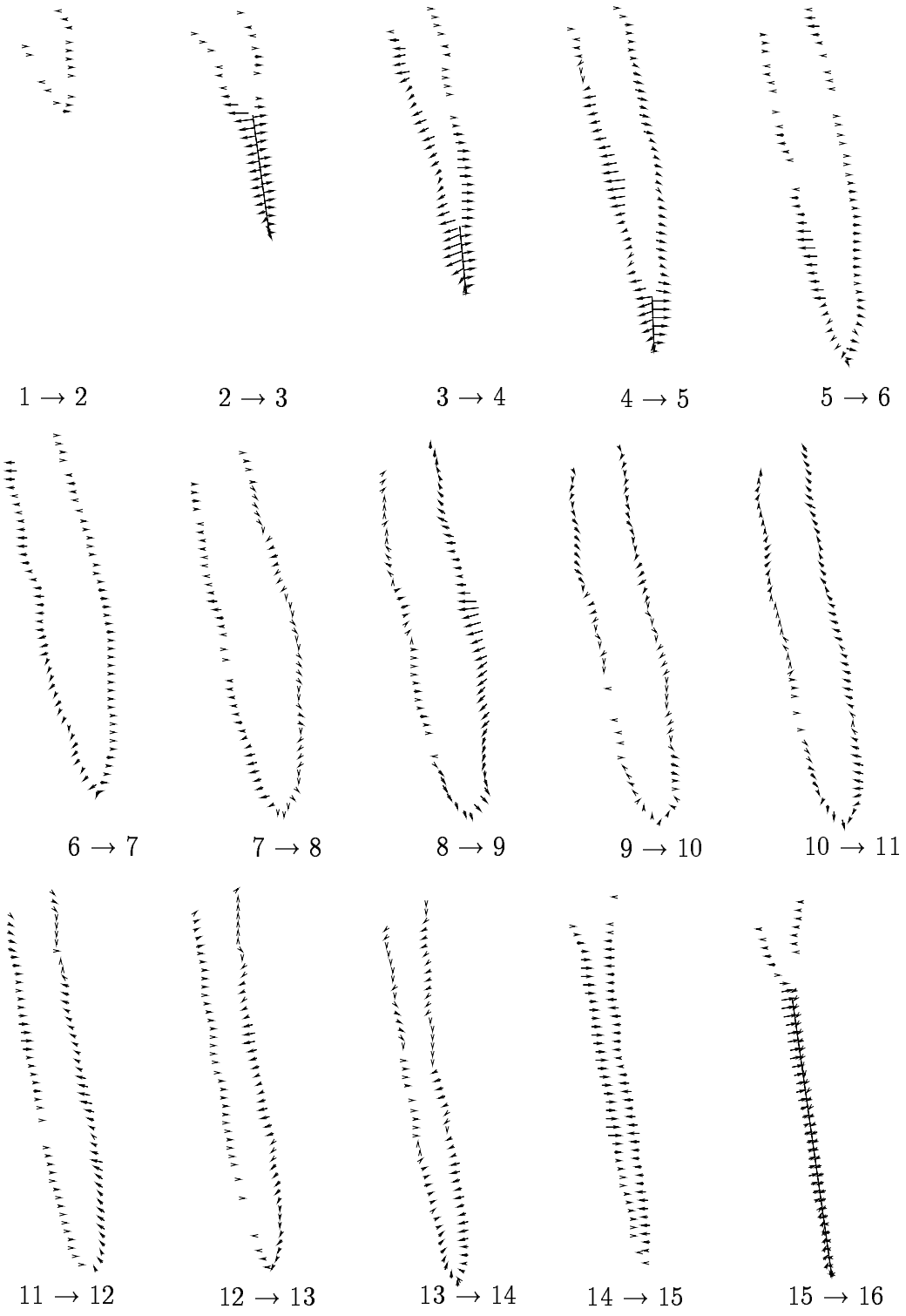


Figure 4: The registration of the successive boundaries of the normal sequence of videostroboscopic images using the proposed simulated annealing based approach runs from top left to the bottom right, the displacement vector field are sampled and magnified.

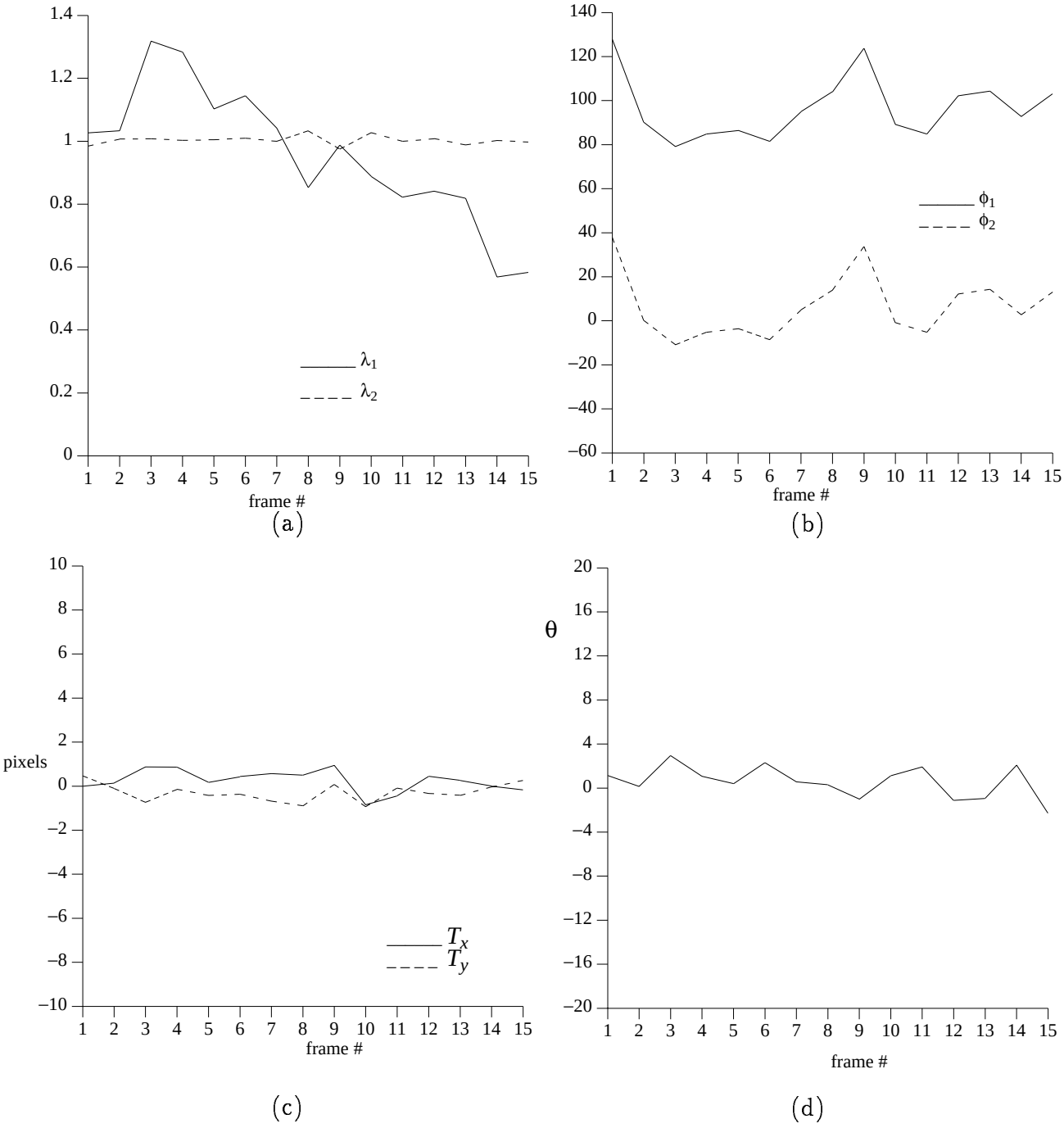


Figure 5: The figures in (a), (b), (c) and (d) represent time evolution of the eigenvalues, principal directions of deformation, the translation along the x and y axes and the rotation angle for the frames of the first videostroboscopic image sequence, respectively. The units of the vertical axes in (b) and (d) are degrees.

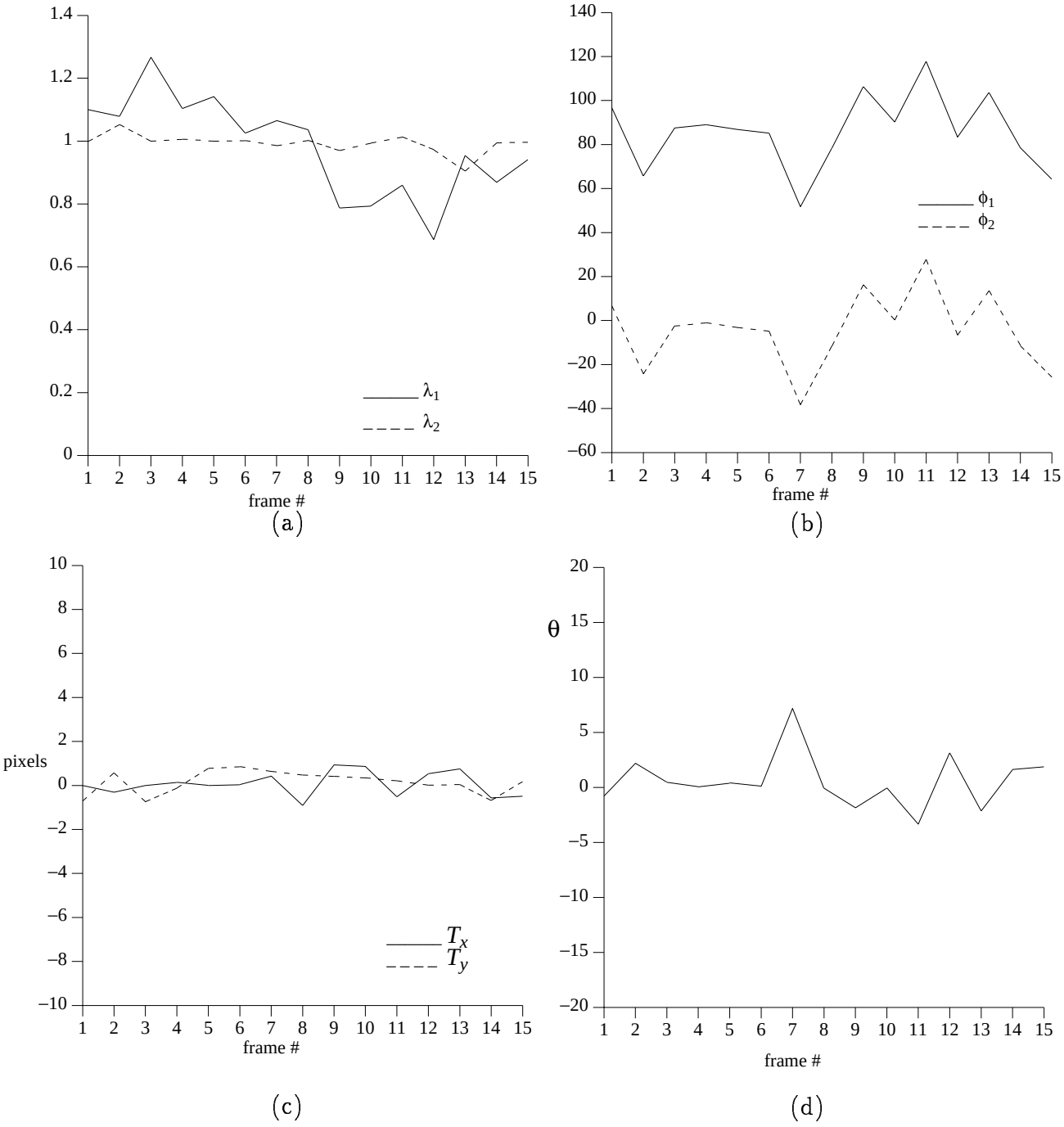


Figure 6: The figures in (a), (b), (c) and (d) represent time evolution of the eigenvalues, principal directions of deformation, the translation along the x and y axes and the rotation angle for the frames of the second videostroboscopic image sequence, respectively. The units of the vertical axes in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

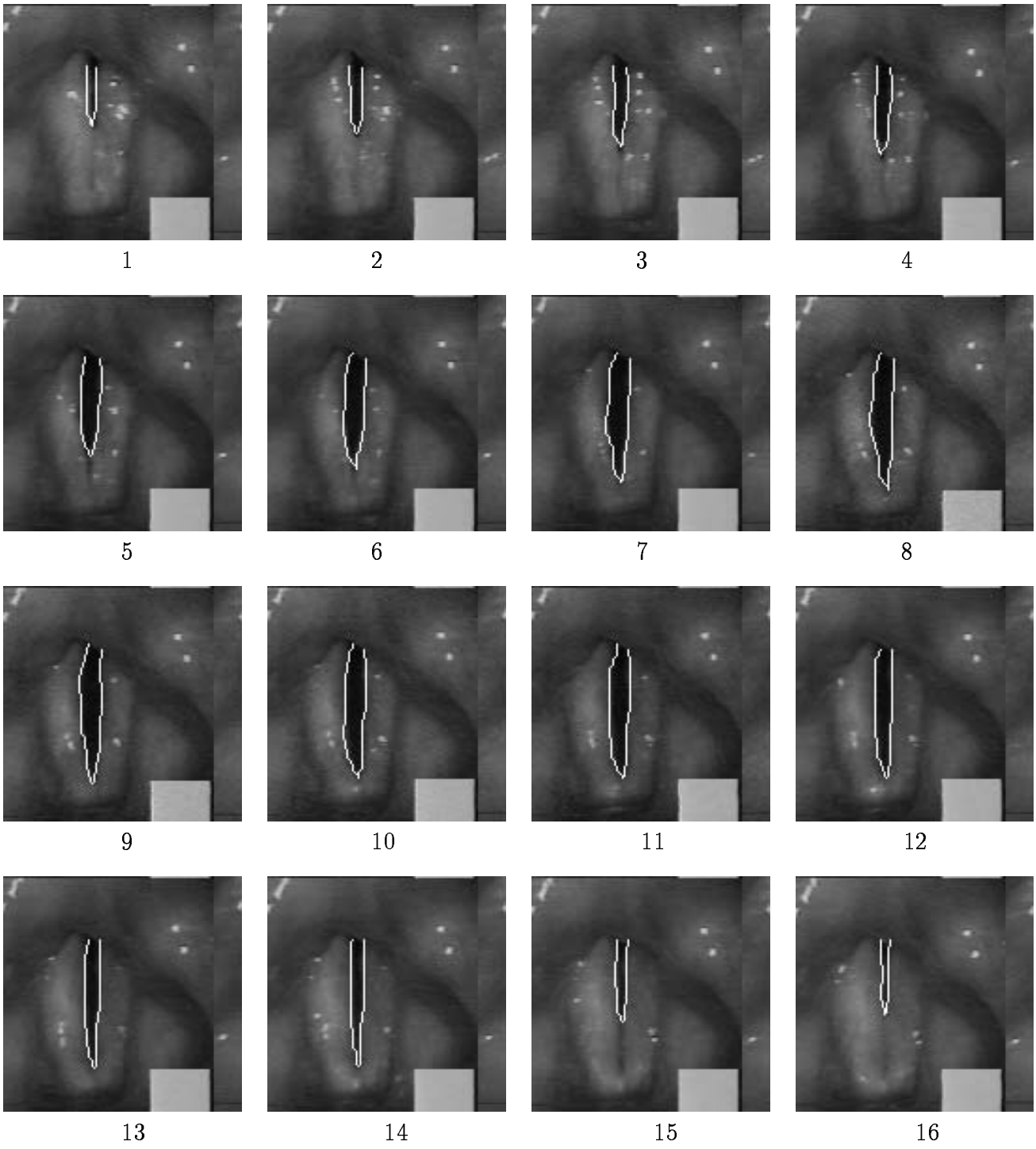


Figure 7: The bright closed contours delineate the boundaries of the vocal folds in the successive frames of the sequence for a patient with polyps run from top left to the bottom right.

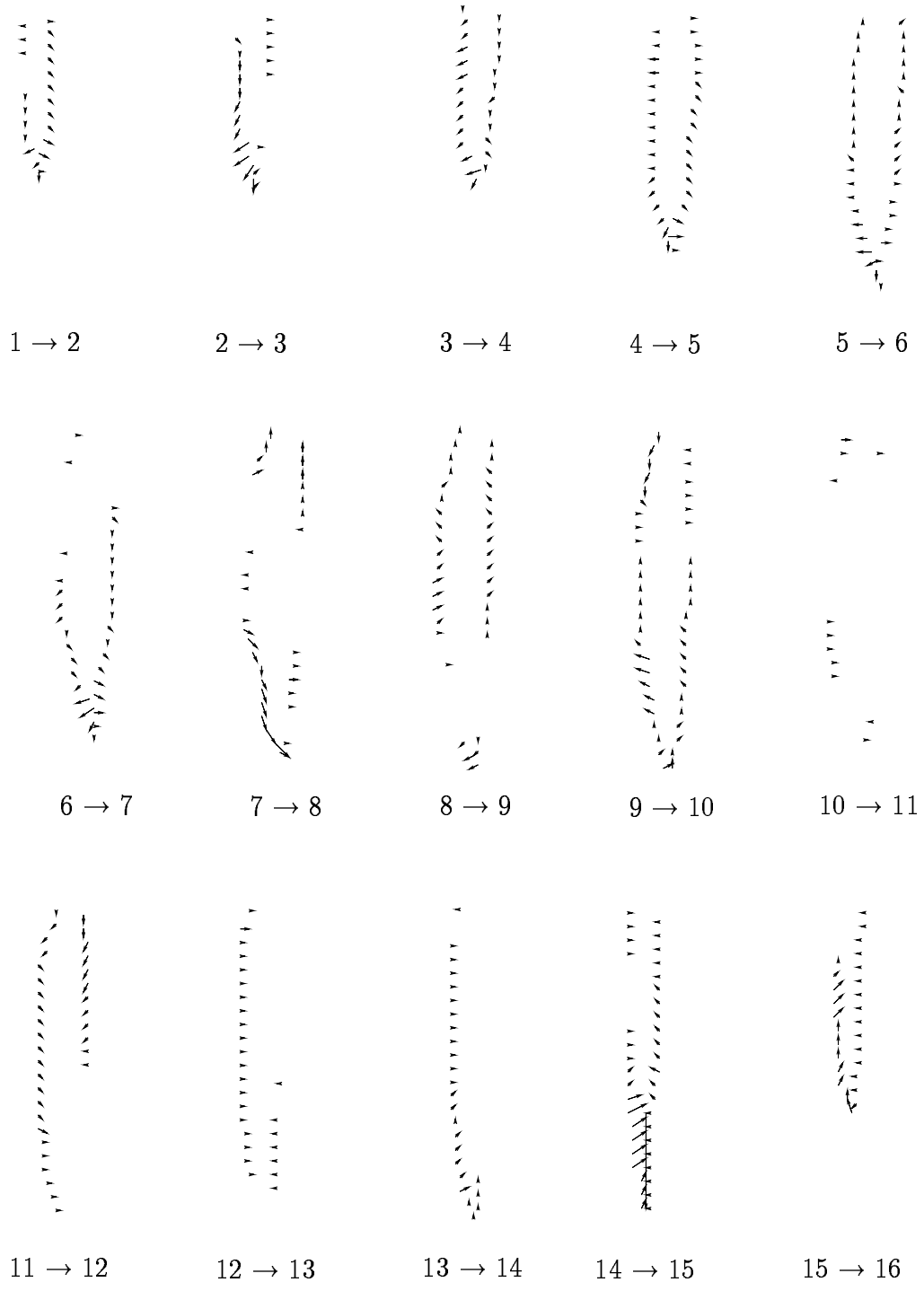


Figure 8: The registration of the successive boundaries of the sequence of videostroboscopic images of a larynx with a polyps using the proposed simulated annealing based approach runs from top left to the bottom right, the displacement vector field are sampled and magnified.

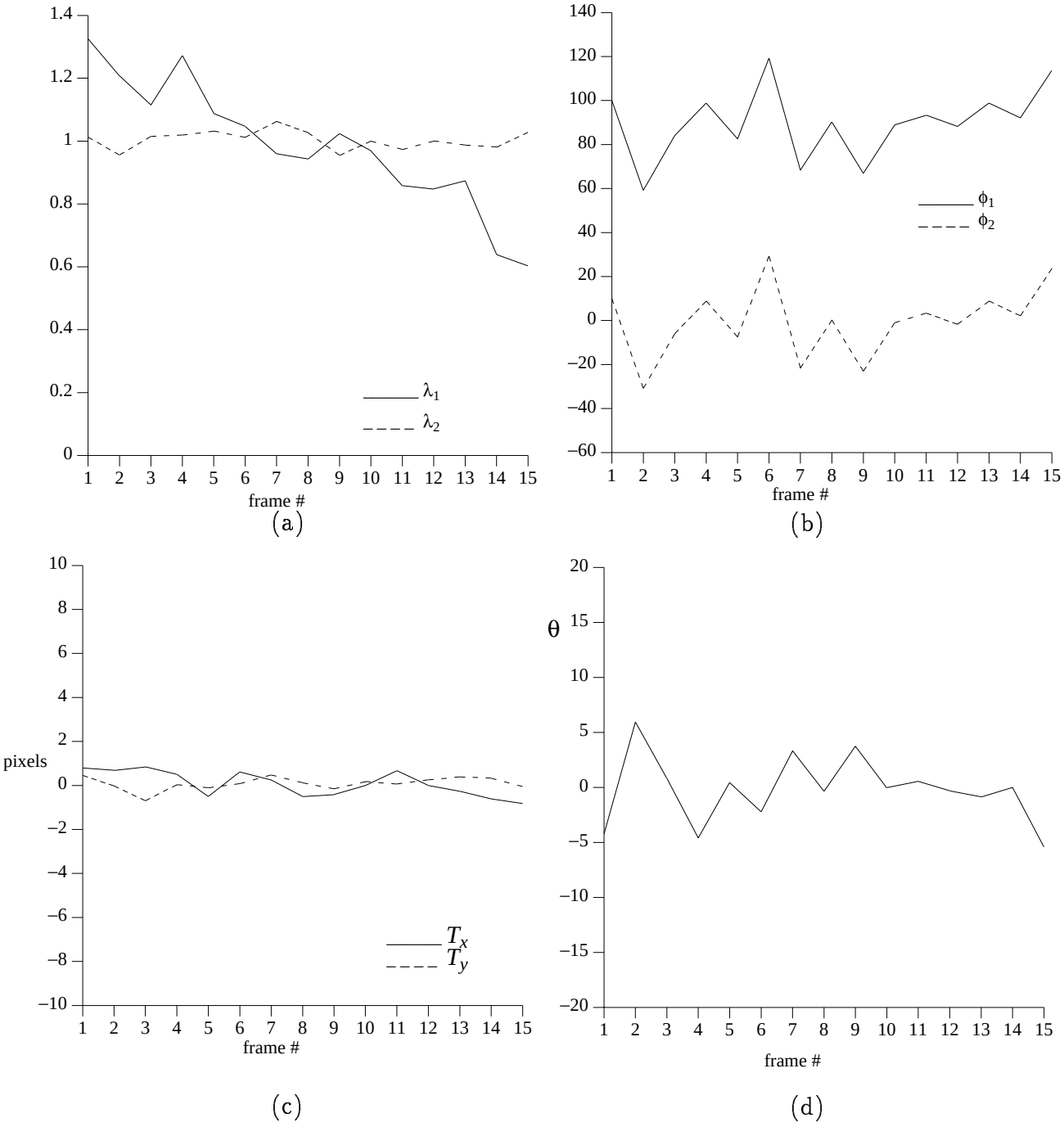


Figure 9: The figures in (a), (b), (c) and (d) represent time evolution of the eigenvalues, principal directions of deformation, the translation along the x and y axes and the rotation angle for the frames of the first abnormal videostroboscopic image sequence respectively. The units of the vertical axes in (b) and (d) are degrees.

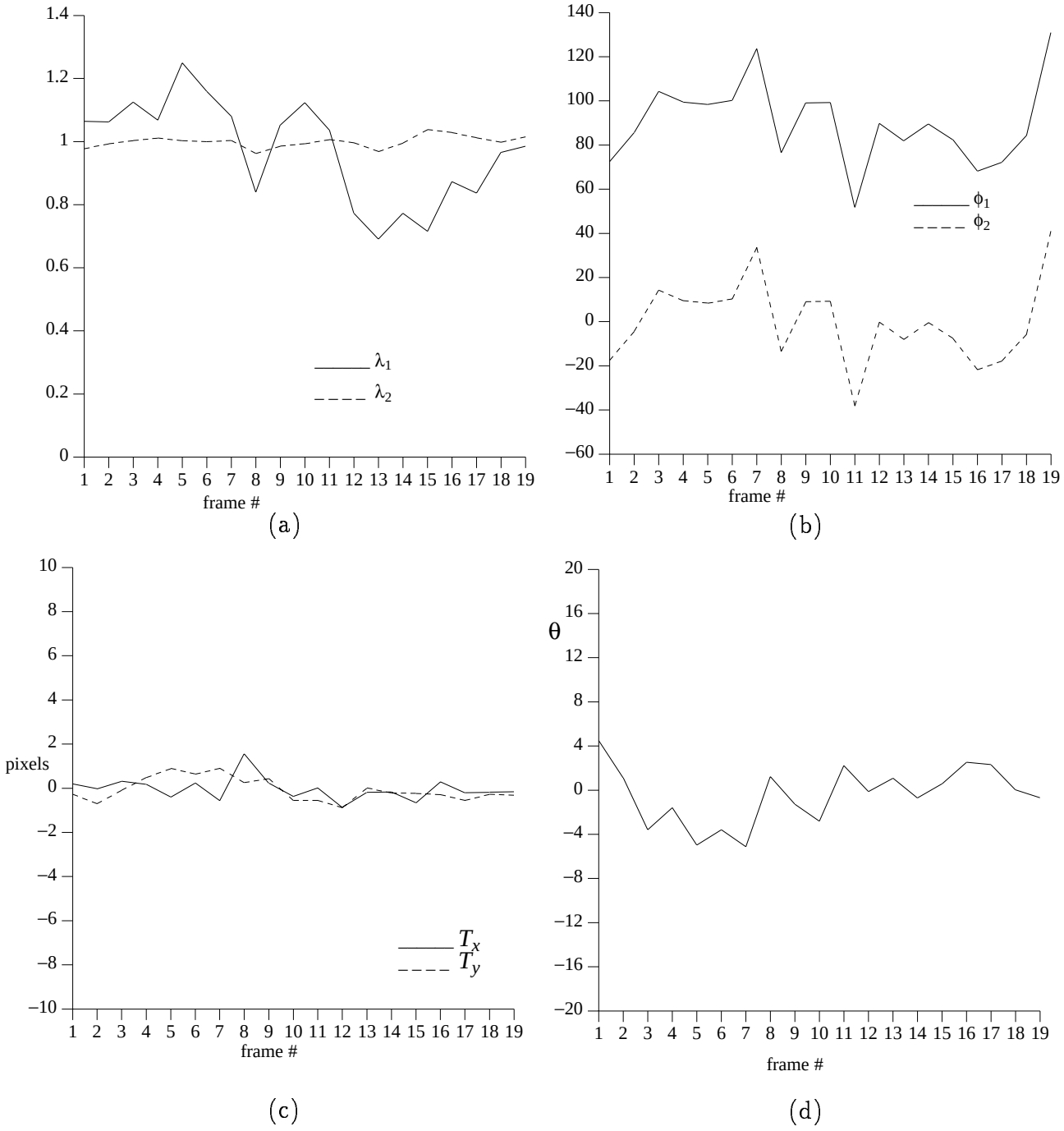


Figure 10: The figures in (a), (b), (c) and (d) represent time evolution of the eigenvalues, principal directions of deformation, the translation along the x and y axes and the rotation angle for the frames of the second abnormal videostroboscopic image sequence respectively. The units of the vertical axes in (b) and (d) are degrees.