

3

# 一种用于声带振动图像自动分割的自适应可变形模型\*

赵雪春 戚飞虎<sup>1</sup> 王磊

(上海交通大学计算机科学与工程系, 上海, 200030)

TP391.42

**摘要** 提出采用基于迭代算法的分级自适应可变形模型,用于声带振动图像的自动分割.该模型由基本可变形模型和高阶可变形模型构成二级模型,基本可变形模型由2条描述声带内壁的样条曲线组成,依据声带振动目标形状的先验知识定义及初始化,高阶模型通过提高每条曲线的阶次,增加控制点数来捕获目标的精细边缘,以得出精确的声带形状分割结果.本文模型充分利用了有关声带形状的先验知识及全局特征,增加了图像分割的正确性及稳定性.

**关键词** 主动轮廓, 可变形模型, 图像分割, 动态规划

## 引言

声带

为了帮助由于声带缺陷造成发声困难或语言障碍的人进行发声训练及语音矫正,采用计算机进行声带振动图像序列分析来辅助治疗,已经显示良好的效果并具有巨大的潜力.在声带振动图像序列分析过程中,声带振动图像的自动分割是其关键的一步.由于自然界中大多数的物体不能由简单的几何形状(如直线、圆、椭圆等)来描述,它们具有非常复杂的轮廓特征,为建立复杂形状的轮廓模型,Terzopoulos, Witkin 及 Kass 等提出了 Snakes 模型,也称可变形模型<sup>[1,2,3]</sup>.近年来,由于可变形模型匹配非刚体目标的强大能力,在基于内容的图像分割、目标匹配、序列图像中目标跟踪等方面有了较为深入的研究,在图像处理及计算机视觉,特别是在医学图像处理领域得到了广泛的应用.

本文针对声带图像分割问题,提出采用基于迭代算法的分级自适应可变形模型,用于声带振动图像的自动分割.该模型由二级模型——基本可变形模型及高阶可变形模型构成.基本模型由2条样条曲线构成,每条曲线上由3个控制点(包括2个角点及1个可由基本能量函数优化的第三控制点),针对声带振动目标形状的先验知识来建立基本可变形模型并初始化该模型.当基本模型给出声带内壁的粗细定位以后,调用高阶模型进一步对图像作精细分割.高阶模型通过提高样条曲线的阶次,增加控制点数来捕获图像中的精细目标,从而可以得出更精确的图像分割结果.由于该自适应可变形模型充分利用了分割目标的先验知识及全局特征,因而增加了图像分割的正确性及可靠性,在进行声带振动图像分割中显示了明显的优越性及自适应能力.同时,采用分级模型克服了传统可变形模型中存在的优化搜索时运算量大、计算速度慢、优化搜索过程不能保证收敛到全局极小点的缺点,从而提高了图像分割速度及稳定性.

\* 国家自然科学基金(编号 69772002)资助项目  
稿件收到日期 199-03-09,修改稿收到日期 1999-06-07

## 1 声带可变形模型

声带由于其本身的物理特性,具有很大的变形能力,因此采用刚体模型建立声带振动模型是非常困难的,而可变形模板模型中,模板可在一定范围内根据约束条件进行特定的变形以实现与待识别目标的最佳匹配,因此采用可变形模型来建立声带振动模型是一种合理有效的方法.声带振动过程中,我们主要观察其内壁的振动变形过程,由2条样条曲线来描述(见图1).声带振动图像自动分割过程概述为:(1)定位声带内壁的2个角点,作为描述声带内壁的2条样条曲线的控制点,即基本模型位置的初始值,若找不到一对合理的角点,则停止;(2)初始化基本模型,分别用3个控制点给出每条样条曲线的参数化描述,并置初值;(3)采用动态规划法对基本模型的能量函数进行优化,搜索出每条曲线的最佳中间控制点;(4)调用自适应高阶模型,增加控制点数,并采用动态规划法对高阶模型的能量公式进行优化搜索,以调整控制点位置,使得能量函数最小化,通过完成对基本模型及高阶模型的能量函数的优化搜索,即可实现声带内壁轮廓的分割,下面我们对以上各个步骤分别进行讨论.

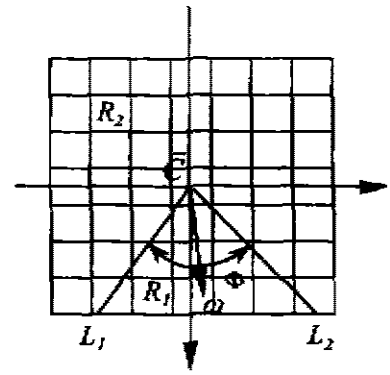
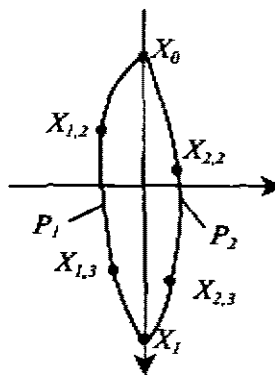
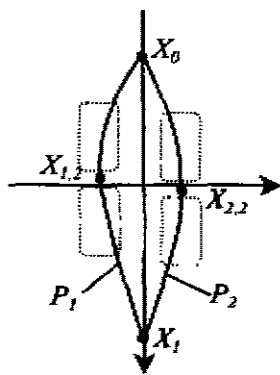


图1 (a)声带内壁的基本可变形模型,(b)高阶可变形模型  
Fig.1 (a) Primary deformable model for vocal cords' inner boundary, (b) high-order deformable model for vocal cords' inner boundary

图2 角点特征示意图  
Fig.2 Characteristic of the corner points

## 2 声带角点检测

在进行声带振动图像分析时,首先需定位声带内壁的两个角点,声带内壁的振动部位位于这2个角点之间,而角点可以看作是由2条直线或曲线交叉点.实验分析表明,满足以下特征的角点可被认为是候选的声带内壁角点.

### 2.1 角点特征

设 $\Phi$ 为介于2条直线 $L_1$ 和 $L_2$ 之间的锐角, $\omega$ 是二分锐角 $\Phi$ 的直线方向角, $\tilde{C}$ 为角点(见图2).区域的相似性 $S$ 用于测量2个区域 $R_1$ 和 $R_2$ 的相似性,即

$$S(\tilde{C}) = \frac{1}{n_1} \sum_{k_1} I(x, y) - \frac{1}{n_2} \sum_{k_2} I(x, y), \quad (1)$$

其中,  $J(x, y)$  是灰度图像,  $n_1$  和  $n_2$  是区域  $R_1$  和  $R_2$  中的像素. 如果声带图像中的角点  $X_0$  和  $X_1$  的角度在设定的范围内(见表 1), 就作为声内腔的可能角点.

表 1 候选角点特征参数表  
Table 1 Parameters of the candidate corners

角点	$\phi$	$\omega$	S
$X_0$	52.5~90	-45~-135	<-10
$X_1$	52.5~90	45~135	<-10

检测角点的第一步是抽取图像中的边缘, 然后检测边缘上的每一点是否为角点. 由于声带振动图像的低对比度, 传统的边缘检测方法常常失效. 本文采用多段二值图像来提取图像边缘. 虽然这样可能造成许多伪角点, 但通过下面的角点选择程序, 通常这些伪角点能够很容易地除去.

## 2.2 选择和检测角点位置

通过角点筛选程序, 可从候选角点中选取最佳角点<sup>[4]</sup>. 首先将候选的角点分为几组. 如果 2 个角点  $X_i$  和  $X_j$ , 其坐标分别为  $(x_i, y_i)$  和  $(x_j, y_j)$ , 若满足  $|y_i - y_j| + |x_i - x_j| < T$  则他们属于同一个组. 其中,  $T$  是阈值,  $T$  的选择由该对角点之间距离的先验知识给定, 选择具有最大  $|S|$  的角点作为该角点组的代表. 给定 2 个角点  $X_0$  和  $X_1$ , 其坐标为  $(x_0, y_0)$  和  $(x_1, y_1)$ , 则设  $y$  轴的距离,  $dy = |y_1 - y_0|$ , 近似等于声带的长度. 假设整个声带图像只有轻微的旋转, 要求声带内壁的一对角点  $(X_0, X_1)$  满足

$$0.15 * \text{Vocal\_length} < dy < 0.9 * \text{Vocal\_length}, \quad (2)$$

其中  $\text{Vocal\_length}$  是声带的长度, 可以通过图像的边缘轮廓检测实验求得. 如果式(3)的代价函数最小, 则选择该对角点为候选角点组的代表. 即

$$\text{cost} = \rho_1 \frac{|dx|}{|dx|_{\max}} + \rho_2 \frac{|S|_{\max}}{|S|}. \quad (3)$$

其中,  $\rho_1$  和  $\rho_2$  是归一化的代价项的权值,  $|dx|_{\max}$  和  $|S|_{\max}$  分别是所有候选角点对  $(X_0, X_1)$  中的  $dx = |x_1 - x_0|$  和  $|S|$  的最大值.

## 3 基本声带可变形模型

声带的基本模型由 2 条声带内壁的轮廓曲线构成, 它们分别为  $P_1$  和  $P_2$  (见图 1(a)). 其中每条曲线由 2 个角点  $X_0$  和  $X_1$ , 及需优化的第三控制点  $X_{i,2}$  ( $i=1,2$ ) 描述.

### 3.1 基本模型能量函数

通过基本能量函数优化来寻找声带内壁的粗细轮廓并定位角点. 基本能量函数由一组能量项来对声带内壁轮廓加以限制, 其中的每一个能量项都是用来捕获声带的一个具体特征, 这些特征为: 2 条声带内壁边缘之间的距离、角点能量、边缘轮廓能量及谷点能量. 由 2 个角点  $X_0$  和  $X_1$ , 及需优化的第三控制点  $X_{i,2}$  (下标  $i$  表示第  $i$  条曲线, 下标 2 表示第三控制点) 给出的曲线  $P_i$  ( $i=1,2$ ) 的基本模型能量函数  $E_i^c$  定义为

$$E_{f_i}^*(X_0, X_1, X_2) = E_i(P_1, P_2) + E_e(P_i, X_0, X_1, X_{i,2}) + E_v(P_i, X_0, X_1, X_2) - E_{dp}(P_1, P_2), \quad (4)$$

其中角点能量  $E_i$  项用于测量边缘上的点是否为合适的角点, 能量定义为

$$E_i(P_1, P_2) = |S(X)|, \quad (5)$$

$X$  是由两条轮廓曲线  $P_1$  和  $P_2$  形成的角点. 边缘能量项用于测量 2 个角点  $X_0$  和  $X_1$  之间沿着曲线  $P_i$  的边缘强度  $\Phi_e$  之和, 即

$$E_e(P_i, X_0, X_1, X_{i,2}) = \frac{1}{L_{P_i}} \int_{r_1}^{r_2} \Phi_e(x) dx. \quad (6)$$

谷点能量项给出了 2 个角点  $X_0$  和  $X_1$  之间的沿着曲线  $P_i$  的谷点能量  $\Phi_v$  之和, 即

$$E_v(P_i, X_0, X_1, X_{i,2}) = \frac{1}{L_{P_i}} \int_{r_1}^{r_2} \Phi_v(x) dx. \quad (7)$$

距离能量  $E_{dp}$  项用于控制声带内壁形状, 即

$$E_{dp}(P_1, P_2) = |\mu_1 L_{vocal} - \mu_2 d(P_1, P_2)| + \rho_1 B_1 + \rho_2 B_2, \quad (8)$$

其中  $L_{vocal} = |X_1 - X_0|$ ,  $\mu_1, \mu_2$  为归一化系数,  $B_1 = |b - b_1|$  和  $B_2 = |b - b_2|$  (如图 3 所示),  $\rho_1$  和  $\rho_2$  是归一化的权重因子. 该能量项用于测量声带形状并控制声带的边缘之间的距离. 它测量了期望的声带内壁两边缘的相对长度  $\mu_1 L_{vocal}$  及相对厚度  $\mu_2 d(P_1, P_2)$ , 同时它也把第三控制点限制在 2 个声带角点的中点附近.

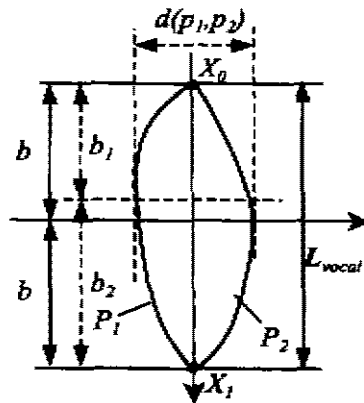


图 3 声带形状控制参数说明  
Fig. 3 Parameters to control vocal cords' shape

在式(6)和(7)中,  $L_{P_i}$  表示曲线  $P_i$  的长度,  $\Phi_e$  和  $\Phi_v$  分别为平滑边缘和谷点因子, 图像边缘和谷点分别由 Canny 边缘检测算子及谷点检测算子求得, 边缘平滑操作是通过边缘图像与平滑因子  $\exp[-\rho(x^2 + y^2)]^{1/2}$  卷积求得, 其中  $\rho$  为平滑因子, 边缘平滑操作的作用是使可变形模板能捕获较大范围的声带变形<sup>[2,6]</sup>.

### 3.2 基本能量函数优化

基本能量函数优化过程按每条曲线顺序进行. 曲线  $P_1$  到  $P_2$  分别对应其能量函数  $E_{f_1}^*$  及  $E_{f_2}^*$ . 在每一优化过程中, 找出各能量项可能值中的最小值, 并将它们归一化为相应的项. 搜索过程采用主动轮廓模型 (Active Contour Model) 优化技术的动态规划算法<sup>[1]</sup>. 每条声带

的轮廓曲线的总能量可沿着曲线的一个领域来进行优化. 每一个控制点都可以移动到它的 8 邻域窗中的任何一点, 如果新点可以给出优化的能量函数值, 则零点替换原来点. 当动态规划优化搜索完毕, 基本可变形模型就给出粗细的定位结果, 则调用自适应高价可变形模型

进入精调阶段.

#### 4 自适应高阶可变形模型

由基本优化过程得到的优化模型可以通过增加声带轮廓曲线的阶次,我们采用了三次 B 样条曲线作为高阶可变形模型<sup>[3]</sup>,通过增加更多的控制点进行精细调整.

由  $m$  个控制点  $(X_0, X_1, \dots, X_{m-1})$  (包括声带的两个角点) 的声带内壁曲线  $P$  的高阶模型的能量函数  $E_p^h$ , 定义为

$$E_p^h(X_0, X_1, X_{1,2}, \dots, X_{1,m-1}) = E_v(P, X_0, X_1, X_{1,2}, \dots, X_{1,m-1}) \\ + E_e(P, X_0, X_1, X_{1,2}, \dots, X_{1,m-1}) + E_{dh}(P_1, P_2). \quad (9)$$

其中,谷点能量项  $E_v$  与式(6)相似,边缘能量项  $E_e$  与式(7)相似,距离能量项  $E_{dh}$  用于在高阶边界曲线中测量声带形状,与式(8)相似,即

$$E_{dh}(P_1, P_2) = |\mu_1 L_{total} - \mu_2 d(P_1, P_2)|. \quad (10)$$

通过高阶模型捕获精细的声带轮廓,事实上高阶模型能量函数是基本能量函数的部分分项.在该能量函数中,角点能量项被取消了,这是基于对基本模型优化中已经正确检测到角点的合理假设基础上的,以达到降低计算代价之目的.同样,我们采用了动态规划法来进行优化.下面以实例来解释增加控制点的过程(如图 1).在新的模型中,右侧声带轮廓曲线  $P_1$  中角点  $X_0$  和  $X_1$  保留,但中间的控制点  $X_{1,2}$  由 2 个控制点  $X_{1,2}$  和  $X_{1,3}$  替代,并且  $X_{1,2}$  和  $X_{1,3}$  被初始化为位于  $X_{1,2}$  处.图 1(a)中虚线矩形区域是新模型中的控制点的搜索区域.同样的方法可以方便地用于增加更多的控制点.

#### 5 实验与讨论

我们用一组声带振动图像系列来测试本文模型,算法用 Matlab 语言编写,并在 DELL P I-400 上运行.图 4 和 5 给出了一些声带振动图像分割的例子,图 4 和 5 中,(a)、(d)为基本模型在图像中的初始位置,(b)、(e)为基本模型的优化结果,声带内壁轮廓被大致定位,(c)、(f)为高阶模型优化结果,正确分割出声带内壁.采用参数化样条曲线对变形模板的初始化,每条曲线采用 3 个控制点,并且这 3 个点通过式(4)进行优化.模型自适应地调整,当基本模型在一定数量的迭代以后,其能量函数已经不再发生变化,则调用自适应高阶模型进行精调.优化方法采用了动态规则法,其计算复杂度为  $O(m^2)$ ,  $m$  为一样条曲线上的控制点数(不包括 2 个角点),实验中一般  $m \leq 10$  即可达到良好的分割效果.

#### 6 结论

我们提出了用于声带图像分割的分级自适应可变形模型.这种新模型采用了声带形状的先验知识来定义能量函数,用于捕获声带形状特征及控制其变形.每一条轮廓曲线由一组控制点给出模型,控制点定位由动态规则优化算法来实现.高阶的模型可以捕获精细的轮廓,通过增加基本模型的控制点来精确调整控制点.实验结果表明该自适应模型能够对不同的人的声带振动图像作自动分割.同时由该模型采用分级模型方法,采用基本模型可以快速



图4 声带振动图像分割结果示例(输入图像  $320 \times 240$  像素)

(a)、(d) 基本模型初始化, (b)、(e) 基本模型优化结果, (c)、(f) 高阶模型优化结果

Fig. 4 Some segmentation results of vocal cords quivering images (input image size is  $320 \times 240$  pixels), (a), (d) initialization of primary model, (b), (e) optimization results of primary model, (c), (f) optimization results of high-order model

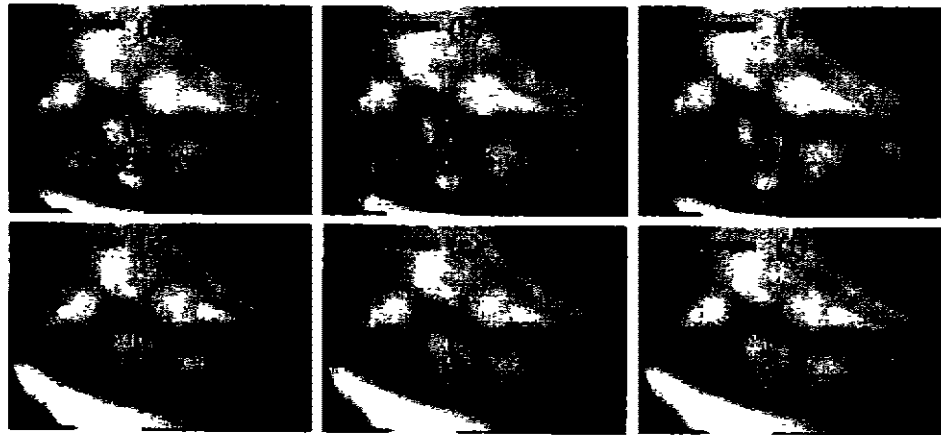


图5 声带振动图像分割结果示例(输入图像  $266 \times 220$  像素)

(a)、(d) 基本模型初始化, (b)、(e) 基本模型优化结果, (c)、(f) 高阶模型优化结果

Fig. 5 Some segmentation results of vocal cords quivering images (input image size is  $266 \times 220$  pixels), (a), (d) initialization of primary model, (b), (e) optimization results of primary model, (c), (f) optimization results of high-order model

定位声带轮廓,而高阶模型又实现精细轮廓分割,从而降低新模型总的计算复杂度,本文方法同样适用于对分割对象具有先验知识的可变形目标分割,如人脸识别中的眼、嘴轮廓的特征提取等等。

## REFERENCES

- 1 Kass M, Witkin A, Terzopoulos D. Snakes: Active contour models, *Proc. of Int. Conf. on Computer Vision*, 1995, 259—269
- 2 Terzopoulos D, Witkin A, Kass M. Symmetry-seeking models for 3D object reconstruction, *Proc. of Int. Conf. on Computer Vision*, 1987, 269—278
- 3 Blake A, Yuille A L. *Deformable Templates in Active Vision*, Cambridge: MIT Press, 1992
- 4 Xie X, Sudhakar R, Zhuang H. Corner detection by a cost minimization approach, *Pattern Recognition*, 1993, 26(8): 1235—1243
- 5 Yuille A L, Hallinan P W, Cohen D S. Feature extraction from faces using deformable template, *Int. J. Comput. Vision*, 1992, 8(2): 99—111
- 6 Brunelli R, Poggio T. Face recognition: features versus templates, *IEEE Trans. PAMI*, 1993, 15(10): 1042—1052
- 7 Amir A A, Terry E W, Ramesh C J. Using dynamic programming for solving variational problems in vision, *IEEE Trans. PAMI*, 1990, 12(9): 855—867
- 8 Boor C D. *A Practical Guide to Spline*, New York: Springer-Verlag, 1978

## ADAPTIVE DEFORMABLE MODEL FOR AUTOMATIC VOCAL CORDS QUIVERING IMAGE SEGMENTATION\*

ZHAO Xue-Chun QI Fei-Hu WANG Lei

(Department of Computer Science &amp; Engineering, Shanghai Jiaotong University, Shanghai 200030, China)

**Abstract** A new hierarchical adaptive deformable model based on iteration algorithm was proposed to automatically segment vocal cords quivering images. The new model consists of two submodels, i. e. primary model and high-order model. The primary model formed by two spline curves that describe two vocal cords' inner boundary is defined and initialized according to the prior knowledge about the shape of vocal cords. The primary model is further tuned by higher-order model that can capture subtle vocal cords' edges by increasing the order of the curves and the number of control points. The experimental results show good performances and robust of the model since it takes full advantage of the prior knowledge and global information of vocal cords.

**Key words** active contours; deformable model; image segmentation; dynamic programming.

---

\* The project supported by the National Natural Science Foundation of China  
Received 1999-03-09, revised 1999-06-27