

文章编号: 1001-8360(2006)05-0046-06

单目模糊点脊的窗口手势跟踪

周 航, 阮秋琦

(北京交通大学 信息科学研究所, 北京 100044)

摘 要: 相对于需要第三方辅助设备的系统, 基于视觉的手势跟踪更方便自然。单目的手势识别跟踪可以分为手部的定位和手指的跟踪。它涉及到对颜色空间的选取, 以及由点到面再回到点的分析方法。首先, 从含有手的场景中检测出手, 并且得到手所在的大致区域, 提取出手掌的中心位置作为追踪的前提。然后, 应用模糊点脊模型来描述手, 对于基本的伸展状态进行了分类, 为语义识别打下基础。为了控制不确定环境条件的影响, 提出了一种直方图匹配算法以及使用搜索窗搜索定位方法。整个试验分成手指定位和手部区域跟踪, 得到的最终数据是手指的伸展方向和角度。在有限的限定条件下, 系统的初始化解算法能得到更多的手部特征, 相对于基于神经的方法减少了训练强度, 但又保持了其准确智能的特点, 能够满足实时的要求。实验的目的在于尽量减少人工干预, 结果证明: 这种方法的可靠性和效率都有一定提高, 具有研究价值。

关键词: 手势识别; 模糊点脊; 窗口; 搜索

中图分类号: TP18 **文献标识码:** A

Window Gesture Tracking Based on BR with Single Camera

ZHOU Hang, RUAN Qiu-qi

(Institute of Information Science, Beijing Jiaotong University, Beijing 100044, China)

Abstract: Compared with other systems in need of third-part equipment, the vision-based gesture tracking system is more convenient and natural. Single camera based gesture tracking includes locating of the hand and tracking of the fingertip. The algorithm is concerned with selection of color spaces and analysis from point to plane and back from plane to point. Firstly, the hand and the rough area of its location are checked out from the scene, and the center of the palm is extracted to facilitate further tracking. Then, a BR model is built to express the postures of the hand and the usual extending states are classified to lay foundation for semantic recognition. A new histogram matching and window searching & locating tracking algorithm is proposed in real-time application to avoid the influence of negative conditions from the scene. The experiment consists of fingertip locating and hand area tracking. The final data results demonstrate the directions and angles of fingertip at different time. Compared with the neural method, the experiment gets more hand features, maintains intelligent accuracy and reduces the training strength under finite limited conditions. For a goal of less manual intervening, it achieves real-time processing. The results prove the algorithm has high reliability and efficiency and is worth researching.

Key words: gesture recognition; BR model; window; searching

随着多通道人机界面技术以及多功能感知机技术的研究和应用, 手的识别定位及理解在许多复杂的现场环境下具有广阔的应用前景。许多场合具有高噪声、环境复杂、粉尘多并且自动化程度要求高的特点。

传统的交互方式(例如鼠标键盘)对于操作现场具有较高的环境要求(稳定性、清洁度和操作的复杂性等)。在这些场合, 如果能够结合传统的交互方式加入手势的识别理解, 能够有效的提高交互界面的自然化和自动化程度。用于控制和理解的识别算法可以分为主动识别和被动识别。主动识别主要依赖于数据手套采集取样信息, 考虑到应用现场的特点, 本文的研究定位在

收稿日期: 2005-06-10; 修回日期: 2005-09-15
基金项目: 国家自然科学基金项目(60472033)
作者简介: 周 航(1974-), 男, 四川成都人, 博士研究生。
E-mail: chouhangcn@yahoo.com.cn

被动识别。这种识别主要依赖于摄像头采集、使用计算机视觉的算法来分析得到的图像序列,最终得到准确的手形并且加以解释。在许多关于跟踪和识别的研究工作中,人们开始聚焦于非自然形态(规则)物体的刚体模型,这些模型具备各种边和角度。基于视觉的识别工作可以利用这些特征并且把手简化成规则的刚体几何集合。

J. Kender 和 R. Kjeldsen^[1]提出了一个实时手势识别系统。这个系统替代了鼠标来移动和调整界面的窗口。在这个系统中,利用肤色特征,手被计算机从背景中分割出来,得到的静态手形应用神经算法来进行分类。C. von der Malsburg 和 Triesch^[2]实现了运动和肤色结合来进行识别,对象的彩色信息被应用到手的定位和跟踪上。静态手形的识别依赖于一种能应用在实时条件下较复杂背景的弹性图表匹配算法。但是,由于它的肤色检测还是基于一种原始的经典的 HS 平面的肤色点集,这种点集对于外界变化波动性很大,所以这样的方法最终对于光照和噪音的变化非常敏感。Davis 和 Shah^[3]在他们的实验中使用了指尖标记的方法并且使用了指尖轨迹追踪来识别 7 个手势。对于不对称特征的遴选使识别更有效率。在本文结合相对于肤色高鲁棒性的色度空间,提出了一种新的手的搜索方法和指尖的预测定位来解决精确定位,减轻环境的干扰。本文的贡献在于:首先提出了一种在 HSV 空间下直方图匹配算法用于肤色的分割,对于开始若干帧的处理取得很好的效果并为下面的跟踪提供了条件。其次在实时跟踪过程中,提出把搜索区域缩小到一个小搜索窗口的方法。把整个手的跟踪具体到了点和线的搜索,减小了运算量并且能够得到所有识别所需要的参数。

1 手部位的分割算法

作为劳动和生活的主要部位,手是一个能高度形变、骨骼结构复杂的部分。通过不同的手形和动作,它能表达出丰富的信息。需要分层次在不同的范围来处理,结合手的肤色信息,本文通过 3 个步骤来进行定位追踪。

1.1 颜色空间选择

通过对以往文献的研究和作者在此之前的实验,相对于 RGB 及其衍生空间,最终选择 HSV 颜色空间作为实验的基础^[4]。在每一帧使用肤色在全肤色光带中的覆盖范围和形状模板来定位手掌和手指区域。利用肤色的背景消减法,能够从 HSV 色度空间中提取出手部区域。所得到的区域点坐标为

$$x = S(V/100)\cos H$$

$$y = S(V/100)\sin H$$

$$z = wV$$

其中, $0 \leq w \leq 1$, $0 \leq H \leq 2\pi$, $0 \leq S \leq 100$, $0 \leq V \leq 100$, 对于 V 的变化, w 是归一化权值。

1.2 初始分割过程

首先对于得到的每一帧,使用背景表观模型来进行预处理^[5]。定义前景为手,包括手指,背景包括除了手以外的任何物体集合。背景模型精确到像素,在手还没进入采集区域的时候,对采集图像序列进行统计,得到初始化的模型。这样随后得到的图像序列和之前的初始模型相比,拥有基于直方图的特征区别明显的肤色信息。这可以作为分割的基本出发点。作者的映射关系结合了贝叶斯法则给出的前景区域的后验概率映射,这些映射都必然相关于需要提取的手部区域。

在实验中,手部的中心定义为是一个距离所有手部区域边界点欧氏距离的期望值最高的点。在算法上,利用形态学腐蚀算子来处理提取手部区域。具体实现过程是首先需要切除手腕部以下的手臂区域,从而得到一个完整有规则的手。这个粗略的结果以手掌为核心,并且一定包含了手部中心点。然后利用形态学腐蚀算子,在反复运算后获得达到规定阈值的期望手形。在腐蚀中规定了 HSV 色度空间的范围,在遇到落入这些范围的值以后,腐蚀停止。最后通过上面给出的规则计算出手的区域中心点,这个中心点就是实验中的手部中央。通过它来确定后面手指的弯曲角度和指向。

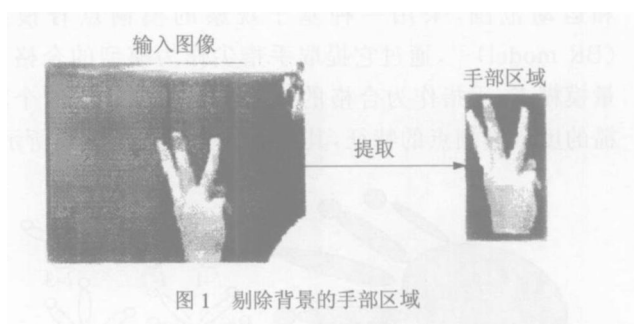


图 1 剔除背景的手部区域

结果如图 1 所示。图中左侧为输入图像。所有处理的前提是认为此时手部一定出现在拍摄的场景中。由于采取了初始化的场景模型,所以可以避免手没有出现就进行分割的情况发生。由于光照的影响,具体的肤色是未知的,但是肤色区域的可预见性减少了后期查找的复杂度。而且由于单目拍摄的特点,出现在相近帧里连通区域的手部皮肤的颜色应该是连续的,这在相关文献中可以得到证明。另外,由于把分割和追踪分开处理,所以在分割阶段,对于至少 15 帧/s 的采样,每帧手势的偏移可以忽略不计。所以在初始阶

段的开始几帧,可以提取出每帧的各项参数,最后加权得到平均的中心点和手部边界位置。如图 2 所示。

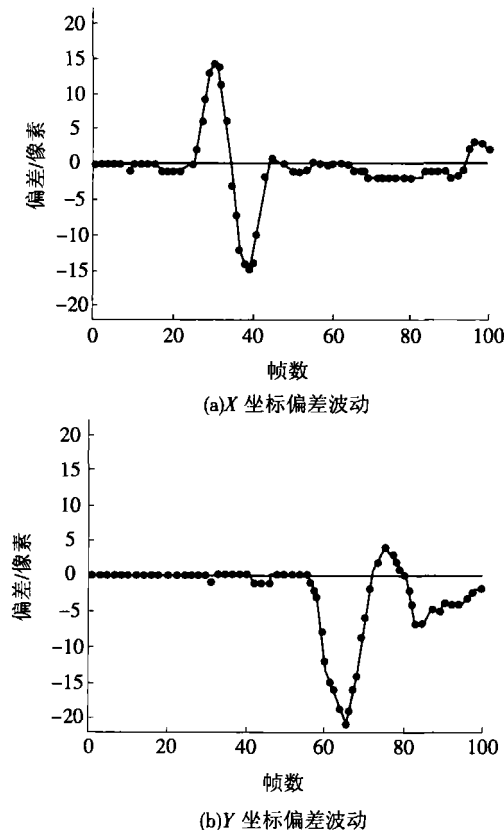


图 2 初始 100 帧中中心点 X,Y 轴偏差波动

1.3 模糊点脊手模型

在找到手部基本位置之后,采取一种更为精细的定位方法来完成后续工作。考虑到相对的方向、位置和运动范围,采用一种基于观察的模糊点脊模型(BR model)^[6],通过它提取指尖作为模型的合格度量模糊点、手指作为合格的度量脊和手掌作为一个粗糙的度量模糊点的特征,其三者的关系图如图 3 所示。

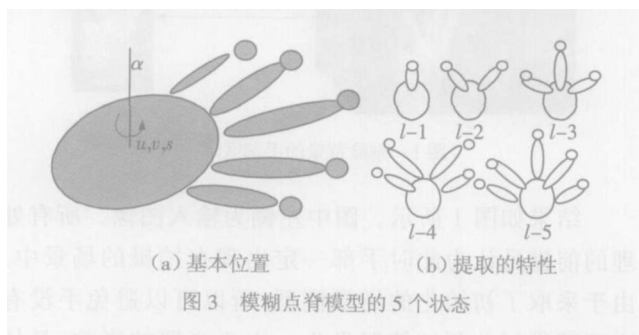


图 3 模糊点脊模型的 5 个状态

定义参数向量 $V = (u, v, s, a, l)$ 表示手部的平移、旋转和变形。 (u, v) 是中心点的坐标; s 表示大小; a 表示手的方向;离散状态 $l = 1, 2, \dots, 5$ 。按照概率优先原则,定义肤色信息为:

(1) 在大约 25 帧中,对于前景 R_s 和背景 R_b ,手部区域从背景中分割出来,并且从包含彩色信息的点

(u, v) 得到相应的直方图;

(2) 直方图被叠加并且归一化;

(3) 任何像素点 (u, v) 归类为皮肤的概率可以表示为

$$w_{skin}(u, v) = \frac{\max(0, aR_s(u, v) - R_b(u, v))}{\sum_{u, v} \max(0, aR_s(u, v) - R_b(u, v))}$$

其中, $a = 0, 1$, 是一个恒判别因子,用来标记皮肤和背景。

1.4 色度直方图匹配

由于采用了 HSV 空间,所以直方图只依赖于色度和饱和度两项指标。根据 Foley^[7] 的方法,在转换 HSV 空间的时候,色度转换成 $0 \sim 360^\circ$ 之间,值为 $0 \sim 1$ 之间的实数。饱和度也被定义在 $0 \sim 1$ 之间,随光照的变化而改变数值。在实验中,这些值会在同一个位图、不同位图之间和不同背景之间发生线性变化。可以把直方图的每一个坐标组成一个二进制的 $M \times N$ 规模的数组。直方图中每一个非连续的二进制数字都被初始化为零。对于每一个给定的肤色面片(模板),都往对应的二进制数组里面加“1”。从而正确表示每个点色度和饱和度。为了进行检测,算法在每个时间搜索一块区域,并且把得到的控制直方图(C)和从图像(I)得到的直方图进行比较。这里参照了 Swain^[8] 的方法来进行比较。两者相似的数量表示为

$$U_{C,I} = \frac{\sum_{i,j} \min(H^C(i, j), H^I(i, j))}{\sum_{i,j} H^C(i, j)}$$

具体的比较算法如下:

(1) 水平和垂直轴选择

图像 I 的 $M \times N$ 直方图大小,让肤色控制直方图 C 扫描目标直方图;

(2) 用模板 C 比较目标

边界值相似检查

用伪代码表示为:

```

counter = 0;
for i = 1; mTRows,
    for j = 1; nTCols,
        if ( T(i, j) eq. skin range ) & ( Canny(i, j) eq.
            skin range ).
            counter = counter + 1
    end
end ( skin range = 0.78)

```

$M_E = \text{counter} / (N * M)$;

边界方向相似度检查

当两个对比部分拥有同一个方向的时候,它们的内积等于 1;当其他情况的时候,则为很小的数字。这

样就找到了方向相似的判断标准:

```

for i = 1; mTRows,
  for j = 1; nTCols,
    if ( T(i,j) eq. skin range) & ( Canny(i,j) eq. skin
      range),
      if abs(template_angle(i,j)-Angle(i,j)) <= a_
        small_number,
        counter = counter + 1;
    end
  endend

```

$U_A = \text{counter} / (N * M)$

这里角度矩阵通过 Canny 滤波器来确定

$$U_A(t, I_k) = \frac{\sum_{i,j} \nabla I(i,j) \bar{t}_k}{N_k}$$

式中, I_k 代表后面的图像区域。

总的相似度

$$U_{C,1} = U_E + U_A$$

(3) 重复步骤(2)直到最后。依照 1 进行扫描。

(4) 利用求得的最大相似度选择区域。

当得到的值大于设定的一个固定值时(在实验中值为 0.78),就认为两个肤色点匹配。而这个肤色点又正好处在连通区域的边缘,也就是说这个点处在一系列同质点(比该点多一个连通区域)的临边,则可以认为这个点是手指尖的一部分。把相同的点归为一个模糊点,之前另外的一系列点作为脊,连接在初始化过程中得到的中心点,就可以从模糊点脊模型中得到需要的方向和角度的参数。



图 4 运用直方图匹配的最终分割结果

在实验中,操作者需要定义一个控制点作为比较的起始种子点,这对于快速识别很有帮助。这样定义并不影响实时的效果,因为它是在正式识别开始前就可以进行的。当然也可以借助于自动定义,但是效率会下降很多,下一个识别周期也会延长。在第一次计算后,就可以自动进行迭代计算,直到遇到规定的阈值,这期间不需要再加入人工干预。前面的训练决定

阈值的选择。有时候,来自更复杂的孤立面片和其他的细条会干扰结果。它们的特点就是每个的周围 8 个面片都没有被匹配算法认定为是肤色区域。如果在每次迭代之前及时去掉这些孤立的部分,在选择控制点的时候就不会被它们影响。所以在每次迭代之前都要消除孤立的部分和其他噪音的影响,这些处理可以通过简单的腐蚀运算来解决。算法的实验结果如图 4 所示,是令人满意的。

2 序列中指尖的搜索和定位

前面一系列工作最终目的是为了在经过若干帧之后能够快速搜索找到指尖的位置并保持对它的定位。对于每一帧已经能够找到它的五维特征向量,那么对于运动序列,如何实时跟踪和定位是手势识别的关键。相对于手臂的追踪,采取了搜索窗口的方法来追踪。基于方向特征的搜索窗是沿着提取出的腕部区域主轴的最高 2 次阶矢量来建立的。因为手指的水平投影形状近似于一个长方体加上一个半圆形,所以映射到窗口内的形状也如此,且搜索窗可以利用几何特征来定位。

吸取了前人的状态空间估计算法^[9]的特点,作者提出了一个新的算法,它能把搜索区域缩小到一个小搜索窗口。这个窗口能预测跟踪指尖位置。首先把窗口的位置定义在上节中计算出的模糊点上,以此为基础进行下面的计算。需要测量每个指尖的位置和速度(最初的速度默认为 0),定义状态向量 $X_t = (x, y, v_x, v_y)^T$, 这里 x, y 代表指尖的位置; v_x, v_y 分别代表在 t 帧时候指尖在两个坐标轴方向上的速度。之后定义观察向量 Y_t , 它表示在 t 帧时候的指尖位置,两个向量的关系可以表示为

$$X_{t+1} = FX_t + Gw_t \quad (1)$$

$$Y_t = HX_t + v_t \quad (2)$$

式中, F 为状态转移矩阵; G 为驱动矩阵; H 为观察矩阵; w_t 是 X_t 的系统噪声,而 v_t 是观察噪声。

因为手势挥动的过程中,时间非常短,通常是 ms 级的,所以在后继一帧中的手指运动可以近似地忽略不计。这样就能得到下面的矩阵。

$$F = \begin{bmatrix} 1 & 0 & T & 0 \\ 0 & 1 & 0 & T \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$G = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}^T$$

$$H = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

筛选过程假定这些模型参数 $\{F, G, H, R, Q\}$ 是准确的^[10]。一旦这个假定不成立,筛选器将会放慢处理速度,而且其中的模型参数会随之改变,从而能够一定程度上消除带来的影响。我们把给出的参数 $\{F, G\}$ 当作标称值并且认为其实际值会围绕着它们做小范围的波动。用状态空间模型替换式(1)可以得到

$$X_{i+1} = (F + F_i) X_i + (G + G_i) w_i$$

式中状态变量 $\{F, G\}$ 可以表示为

$$\begin{bmatrix} F_i & G_i \end{bmatrix} = M_i \begin{bmatrix} E_f & E_g \end{bmatrix} \quad (3)$$

对于矩阵 $\{M, E_f, E_g\}$ 以及任意的收缩因子 α_i , α_i 在实验里值取作 0.68。当在某个特定时间模型变化非常大的时候,可以让 $\{M, E_f, E_g\}$ 随着时间进行调节。公式(3)允许在某个波动范围能限定不确定因素的来源。这些不确定因素包括了光线条件的变化,背景和物体的自由移动以及手部动作的突然变向和变速等不正常情况。

3 实验结果和分析

在本实验里,一开始利用背景表观模型进行了初始化分割。利用相应的模型,每个像素属于前景或者背景的相似度可以通过计算得出。然后计算出手的中心位置,为下一步的分割识别打下基础。

在把 RGB 空间转换成 HSV 空间之后,通过直方图匹配检测到了肤色区域。被称作控制种子的初始化的肤色块用作迭代法的初始化。手的模糊点脊模型非常好的描述了手部跟踪。在跟踪手指过程中,置信度被指定为一个在 15 至 180 之间的角度,它给出了搜索窗口的方向信息。在测试过程中,根据以前的数据并且考虑实际情况,赋予 $\{M, E_f, E_g\}$ 以下的值

$$M = \begin{bmatrix} 1 & 0.5 & 0.5 & 0.25 & 0.25 & 0.125 \end{bmatrix}^T$$

$$E_g = \begin{bmatrix} 0 & 0 & 0.316 & 0.316 & 10 & 10 \end{bmatrix}$$

$$E_f = \begin{bmatrix} 0 & 0 & 2 & 2 & 4 & 4 \end{bmatrix}$$

搜索窗口被安置在手的随机运动方向的 \pm 角度之内。

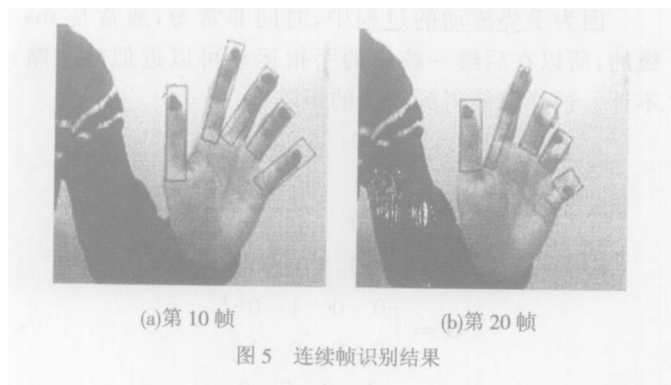


图 5 是手指的跟踪和定位结果。指尖位置和其方向被准确检测。通过和同类方法比较,作者的方法降低了复杂环境下错误的识别率,取得了预期的效果,本文方法与部分算法的比较结果如表 1 所示。

表 1 本文方法与部分算法的比较

算 法	误否定率/ %	误肯定率/ %	总错误率/ %
贝页斯法(真实值)	7.5	4.0	11.6
贝页斯法(上界)	3.1	3.6	6.7
简化算法(上界)	6.4	11.9	17.8
本文方法	4.2	3.0	5.7

4 结论

本文提供了一个稳定的在 HSV 空间单目条件下对手进行分割跟踪的处理方法。模糊点脊模型算法的使用使得最终的分割取得了很好的效果。在实时状态下,为了对手进行定位,提出了新的搜索窗法来减少实时系统的处理区域。实验得到了精确的手指伸展位置和方向,这些数据为手势的识别和解读提供了重要的判定依据。根据得到的方向和角度,可以设计语义库解读手势含义。相对于适应平面运动识别的 Alberto^[11]的直方图算法,作者的匹配方法不需要复杂的训练步骤,并且高效的完成了肤色区域的提取。在实验中,初始化过程变得仅仅是为了提高效率而采取的优化措施,并且在整个过程中对外部变化的依赖也减轻了,这就最大程度上减少了致命错误的发生,在特殊应用场合更具有独特的优势。此外通过实验也可看到手势识别完全可以应用到信号控制、人机交互等方面,从而提高人机界面的水平。

参考文献:

- [1] Kjeldsen R, Kender J. Finding skin in color images [A]. In: Proceedings of International Conference on Automatic Face and Gesture Recognition [C]. (Killington, Vt.), 1996. 312—317.
- [2] Jochen Triesch, Christoph von der Malsburg. Classification of hand postures against complex backgrounds using elastic graph matching [J]. Image and Vision Computing, 2002, 20: 937—943.
- [3] Davis J, Shah M. Visual gesture recognition [A]. In: IEE Proc. - Vision, Image, Signal Processing [C]. 1994. 141, 101—105.
- [4] Foley J D, A Van Dam, Feiner S K, Hughes J F. Computer Graphics: Principles and Practice [M]. 2nd ed. Mass: Addison-Wesley, 1993. 590.
- [5] McAllister G, McKenna S J, Reketts I W. Towards a non

- contact driver-vehicle interface [A]. In: IEEE Conference on Intelligent Transportation Systems [C]. Dearborn, Michigan: October, 2000.
- [6] Bretzner L, Lindeberg T. Qualitative multi-scale feature hierarchies for object tracking [A]. In: Journal of Visual Communication and Image Representation [C]. Nov. 2000, 115—129.
- [7] Foley J, Van Dam A, Feiner S K, Hughes J. Computer Graphics [M]. Addison-Wesley: Reading, 1992. 590—593.
- [8] Swain M J, Ballard D H. Color Indexing [J]. International Journal of Computer Vision, 1991:11—32.
- [9] Sayed A H. A framework for state-space estimation with uncertain models [A]. In: IEEE Transactions on Automatic Control [C]. July 2001, 46(7):998—1013.
- [10] Mehra R K. On the identification of variances and adaptive Kalman filtering [J]. IEEE Transaction on Automatic Control, 1970, AC-15:175—183.
- [11] Tomita A, Jr., Ishii R. Hand shape extraction from a sequence of digitized gray-scale images Industrial Electronics [A] In: Control and Instrumentation, IECON '94, 20th International Conference [C]. Sept, 1994, 3(5-9):1925—1930.

(责任编辑 江 峰)

詹天佑科学技术发展基金会简介

詹天佑科学技术发展基金会 詹天佑科学技术发展基金会经铁道部同意,报民政部审定,国务院领导批准,于 2005 年 12 月 17 日在北京正式成立。

詹天佑科学技术发展基金会是在中国科学技术发展基金会原詹天佑铁道科技发展基金的基础上,按国家新颁布的《基金会管理条例》的要求设立的。基金会是从事社会公益性活动的非营利民间组织,独立社团法人。基金会实行双重管理体制,业务主管单位为铁道部,登记管理机关为民政部。

宗旨与业务 宗旨是缅怀詹天佑先生对发展中国铁路事业做出的突出功绩,继承和弘扬他爱国、创新、自力更生、艰苦奋斗的精神,支持和赞助科技活动,表彰奖励有突出贡献的科技人员,促进中国铁道科技事业的繁荣和优秀人才的成长,为铁路跨越式发展做出贡献。任务是充分利用国内外各种渠道募集资金,表彰奖励在铁路科技发展中做出突出贡献的优秀科技人员,资助重大科技活动,组织管理“詹天佑铁道科学技术奖”的评选工作,与国内外科技团体建立友好与合作关系,为企业事业单位和团体设立专项奖励基金。

组织机构 詹天佑科学技术发展基金会理事会为决策机构,并设监事会。理事会下设秘书处和奖励评审委员会,秘书处为基金会的日常办事机构,由秘书长或副秘书长主持工作。

基金来源与管理 詹天佑科学技术发展基金会的基金来源有:国内外友好团体、企事业单位和个人的捐赠,政府资助,以及其它符合国家法律规定的各类捐赠。到 2005 年末,基金总额约 2500 万元,其中专项基金 577 万元。基金的使用由理事会管理,实行独立核算,受铁道部的审计监督。

詹天佑铁道科学技术奖 自 1993 年设立詹天佑铁道科学技术奖以来,每两年举办一届,已成功地举办了七届奖励活动,共奖励 631 名科技人员,其中获专项基金奖的人员 300 名。詹天佑铁道科学技术奖是首批经国家科学技术奖励工作办公室审查批准予以登记的奖项,已成为国家科学技术奖励重要的、有益的补充,也是铁路科技领域有重大影响的奖项,受到了广大科技工作者的欢迎,对促进铁路科技进步和优秀人才成长发挥了积极作用。该奖项设大奖、成就奖、青年奖、专项基金奖,奖金额大奖 6 万元/人、成就奖 2 万元/人、青年奖 0.8 万元/人。遵照刘志军名誉理事长的指示精神,从第八届奖励活动开始,增设贡献奖,奖励在铁路运输组织、经营管理、挖潜提效、安全生产方面做出突出贡献的科技人员。

(下转第 64 页)