

变形网格及其在图像识别中的应用¹

燕继坤 张震* 郑辉

(西南电子电信技术研究所现代信号处理国家重点实验室 成都 610041)

*(信息工程大学通信工程系 郑州 450002)

摘要: 网格特征是图像识别中一类重要特征, 而变形模板对于复杂的图像识别问题如字符识别、数字识别、图标识别等表现出很好的性能, 但变形模板很费时. 该文针对网格特征提出了变形网格, 并分析了变形模板与变形网格之间的近似等价性. 所提方法对网格进行变形而不是对图像变形, 因而速度比变形模板快得多, 而且性能相差不大. 把这种方法分别应用于图标识别和脱机手写汉字识别. 图标识别实验中变形模板使识别率提高了 7.5%, 而变形网格使识别率提高了 7.3%. 手写汉字识别实验中变形模板使识别率提高了 6.1%, 而变形网格使识别率提高了 5.8%. 考虑到变形网格比变形模板快得多, 所以这种方法是具有优势的.

关键词: 图像识别, 变形模板, 变形网格, 图标识别, 汉字识别

中图分类号: TP391.41 **文献标识码:** A **文章编号:** 1009-5896(2004)08-1183-07

Application of Deformable Grid in Image Recognition

Yan Ji-kun Zhang Zhen* Zheng Hui

(LISP, The South-west Inst. of Electron. & Telecom. Tech.,

Chengdu 610041, China)

*(The Information Engineering University of Zhengzhou, Zhengzhou 450002, China)

Abstract Grid feature is a popular feature extraction scheme in image recognition, and usually higher performance could be obtained with the help of deformable template, especially in such difficult image recognition tasks as character recognition, digit recognition, logo recognition etc. The critical shortcoming of deformable template is time-consuming. In this paper the method of deformable grid is proposed to compensate for the shortcoming of deformable template. Firstly certain grid must be superimposed on image, then unlike deformable template where deformation is applied to image, various deformations are applied to grid. Because the number of grid is much less than that of pixels in image, the method is much more timesaving comparing to deformable template. The approximate equality of deformable template and deformable grid is also analyzed. The method is evaluated by two image recognition experiments, namely, logo recognition and off-line Chinese character recognition. The improvements in recognition rate by 7.3% in first experiment, and 5.8% in second one are obtained by the use of deformable grid.

Key words Image recognition, Deformable template, Deformable grid, Logo recognition, Chinese character recognition

1 引言

特征抽取是图像识别的基本问题. 研究者提出了多种特征抽取方法, 其中很重要的一类特征我们称为网格特征. 这类特征首先在图像上划分 $m \times n$ 的网格, 然后抽取每个网格内的特征,

¹ 2003-03-17 收到, 2003-08-29 改回
国防预研基金资助项目 (编号: 514950307)

得到 $m \times n \times k$ 维向量 (假设每个网格中抽取到的特征是 k 维)。所抽取的特征有灰度平均值、黑像素数目、边缘方向线段直方图等。有的研究者注意到靠近网格边界处图像的微小变化会带来相邻网格内特征比较大的变化, 因而进一步采取了“模糊边界”^[1-3]。

变形模板是提高图像识别准确率的有效方法, 尤其是对字符识别、数字识别、图标识别等图像合法变形极多的复杂图像识别问题^[4-6]。这种方法认为每个模式都有若干标准模板 (图像), 属于该模式的其它图像是由标准模板变形而来, 因此如果知道变形的方法就可以通过对待识别图像进行逆向变形, 使之接近或相等于某一标准模板。由于不知道待识图像的变形方法, 所以一般的做法是用“穷举搜索”对待识图像进行所有可能的变形, 抽取所有这些变形的特征, 用加权表决法对所有特征的匹配结果进行综合。投票所用的权值一般与变形的大小有关, 变形越大则权值越小, 变形越小则权值越大。图 1 表示了在使用网格特征时基于变形模板的识别方法, 其中“变形图像_{*i*}” ($i = 1, 2, \dots, n$) 表示由输入的待识图像得到的变形图像, “网格_{*i*}” ($i = 1, 2, \dots, n$) 表示对“变形图像_{*i*}”生成的网格, “特征抽取和匹配”表示对“变形图像_{*i*}”根据“网格_{*i*}”抽取某种网格特征并匹配。

变形模板在许多测试中表现出很好的性能, 但是变形模板很重要的缺点就是速度很慢, 因为“穷举搜索”需要生成很多变形图像, 而图像数据量比较大, 需要耗费大量时间。例如文献 [5] 中对手写数字识别进行了 25 种图像变形, 这样耗时也基本上是不使用变形模板而直接识别的 25 倍。本文针对网格特征提出了对变形模板的改进方法, 我们的方法不是对图像进行变形, 而是对网格进行变形。与图像相比网格数目很少, 所以我们方法的速度比变形模板快得多, 而且有和变形模板相近的效果。与变形模板相对应, 我们的方法称为变形网格。

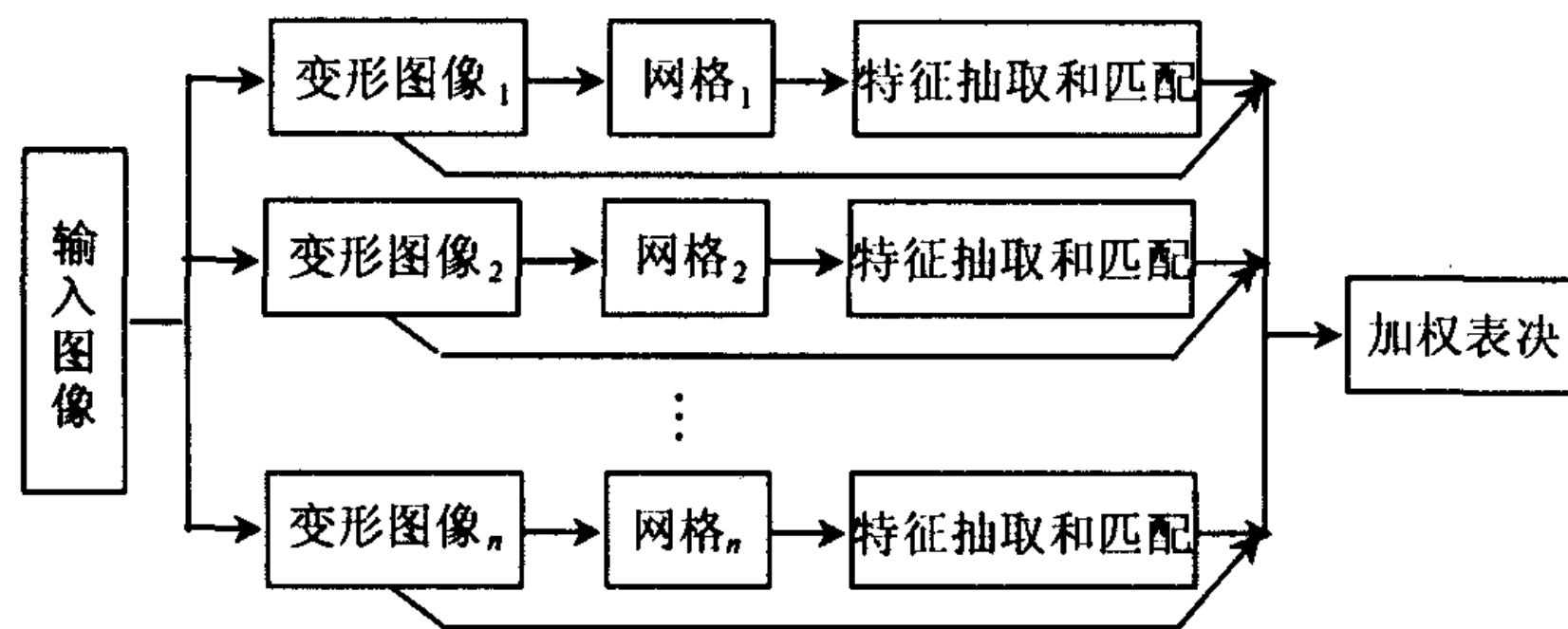


图 1 基于变形模板的识别方法

2 变形网格

我们首先通过汉字识别的例子来说明变形网格的思路, 再通过两个应用实例证明这种方法的有效性。因为书写习惯的不同手写汉字有很多变形, 例如笔划挤到上边或下边, 图 2 中左边是一个挤到上边的汉字“平”。如果使用变形模板, 需要把图像向下推拉得到与标准写法相近的图像, 然后对该图像划分网格并抽取特征。但是也可以不对图像进行变形, 而是首先在待识别的汉字上划分网格, 然后把网格向上推拉, 再抽取特征。可以看出, 这两个方法得到的特征相近。对于其它的变形如旋转、扭斜等可得到同样的结论。

我们进一步说明变形网格的实质, 设 P 是待识别的图像。

$$P = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1N} \\ p_{21} & p_{22} & \cdots & p_{2N} \\ \vdots & \vdots & \vdots & \vdots \\ p_{M1} & p_{M2} & \cdots & p_{MN} \end{bmatrix} \quad (1)$$

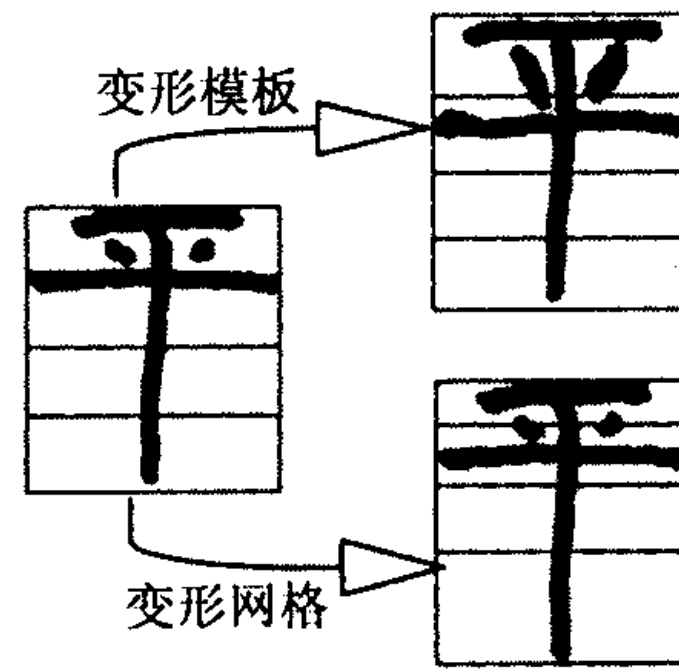


图 2 变形网格与变形模板的对照

设 C 是划分网格的函数， C 把 P 划分成网格。

$$C(P) = \begin{bmatrix} P_{11} & P_{12} & \cdots & P_{1n} \\ P_{21} & P_{22} & \cdots & P_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ P_{m1} & P_{m2} & \cdots & P_{mn} \end{bmatrix} \quad (2)$$

其中 $P_{ij}(1 \leq i \leq m, 1 \leq j \leq n)$ 是 P 中位于网格 (i, j) 的格子内的元素组成的子阵。设 E 是特征抽取函数， E 以 $C(P)$ 作为参数，抽取每个 P_{ij} 的特征，排在一起得到特征 F 。即

$$F = E[C(P)] = E(P_{11}, \cdots, P_{1n}, P_{21}, \cdots, P_{2n}, \cdots, P_{ij}, \cdots, P_{m1}, \cdots, P_{mn})$$

设 D 为变形函数， D 以切分出的子阵为参数，得到变形后的图像

$$P'_{ij} = D(P_{ij}) \quad P' = D(P) = \begin{bmatrix} P'_{11} & P'_{12} & \cdots & P'_{1n} \\ P'_{21} & P'_{22} & \cdots & P'_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ P'_{m1} & P'_{m2} & \cdots & P'_{mn} \end{bmatrix} \quad (4)$$

其中 P'_{ij} 为变形后的图像块， P' 为变形后的图像。

对于变形图像，特征抽取可表示为（运算顺序由右到左）：

$$E \cdot C(P') = E \cdot C \cdot D(P) \quad (5)$$

其中“ $E \cdot C$ ”表示 E 操作的输入为 C 操作的输出，其它操作的含义相同。

当变形较小时，就像图 2 中观察到的那样，式 (5) 中的两个函数 C, D 近似满足交换律，即

$$E \cdot C \cdot D(P) = E \cdot D \cdot C(P) \quad (6)$$

式 (6) 右端表示先划分网格，再对网格进行变形，最后抽取特征，这就是变形网格。而左边对应的是变形模板。所以变形网格实质上是利用两个函数 D, C 在变形较小时近似满足交换律这一事实大大减少运算量。图 3 表示的是基于变形网格的识别方法，对输入图像首先划分网格，对网格进行变形生成若干个变形网格，以每个变形网格为基础抽取某种网格特征并匹配，最后根据所有匹配结果进行加权投票以确定最终结果。

为了控制变形的大小，我们给 D 函数加上一个控制参数 a 。即

$$P' = D(P, a) \quad (7)$$

设 a 的取值范围为 A 。使用中, 在 a 的取值范围 A 之内取有限个离散值 a_1, a_2, \dots, a_n 可达到近似“穷举搜索”的目的。

变形网格存在着因变形过大而使抽取的特征发生畸变的危险, 因而我们给每个 a_i 定义一个可信因子 w_i 。变形越大 w_i 越小, 变形越小 w_i 越大。 w_i 作为权值参加加权投票。

我们将通过图标识别和脱机手写汉字识别两个实验说明变形网格方法的有效性。

3 图标识别

文档图像指图像格式的电子文档, 例如传真、扫描得到的文档等。随着大量电子文档的出现, 完全依靠手工标注是一件困难的事情。图标指文档图像中的小型图形标志, 如公司的标志、产品标志等。图标很好地反映了文档的来源, 表达了文档的语义信息, 所以图标识别是文档图像标注的有效途径。完整的图标识别主要包括图标定位和图标识别两大部分, 本文仅介绍我们在图标识别部分所做的工作。我们把网格变形法应用于图标特征的提取。

在实验中我们使用的网格特征以边缘方向特征为基础, 反映了图标形状的局部统计特性^[3]。我们认为不像文字, 图标的边缘并不能代表全部信息, 因此我们还加上了内部黑像素(相对于边缘黑像素)的特征。

我们使用的网格划分方法是重心网格, 这种网格在黑像素密集的地方划分较细, 在黑像素稀疏的地方划分较粗, 因而具有“自动变焦”的能力。以重心网格作为初始网格, 对它进行多种变形, 根据每种变形抽取相应的特征, 并分别进行识别, 最后用加权表决的方法得到最终识别结果。我们使用了 16×16 的网格, 每个网格内抽取到的特征为 5 维, 这样图标的特征为 1280 维, 我们用欧氏距离度量两个特征之间的距离。

我们使用的实验数据主要来自于传真图像。图 4 是一些图标的例子, 箭头左边是“标准”图标, 右边是实际中出现的变形图标, 箭头表示变形的方式。图标比较常见的几类变形有:

(1) 推拉变形: 像素挤到某一部分。在水平方向有挤到左边、中间、右边几种; 在垂直方向有挤到上边、中间、下边几种。图 4 中图标 1 的像素挤到下边;

(2) 扭斜变形: 左边升高, 右边降低, 或者相反地左边降低, 右边升高。图 4 中图标 2 左边升高, 右边降低;

(3) 图像旋转: 图 4 中图标 3 发生了旋转。

(4) 腐蚀或膨胀: 图 4 中图标 4 发生了腐蚀。

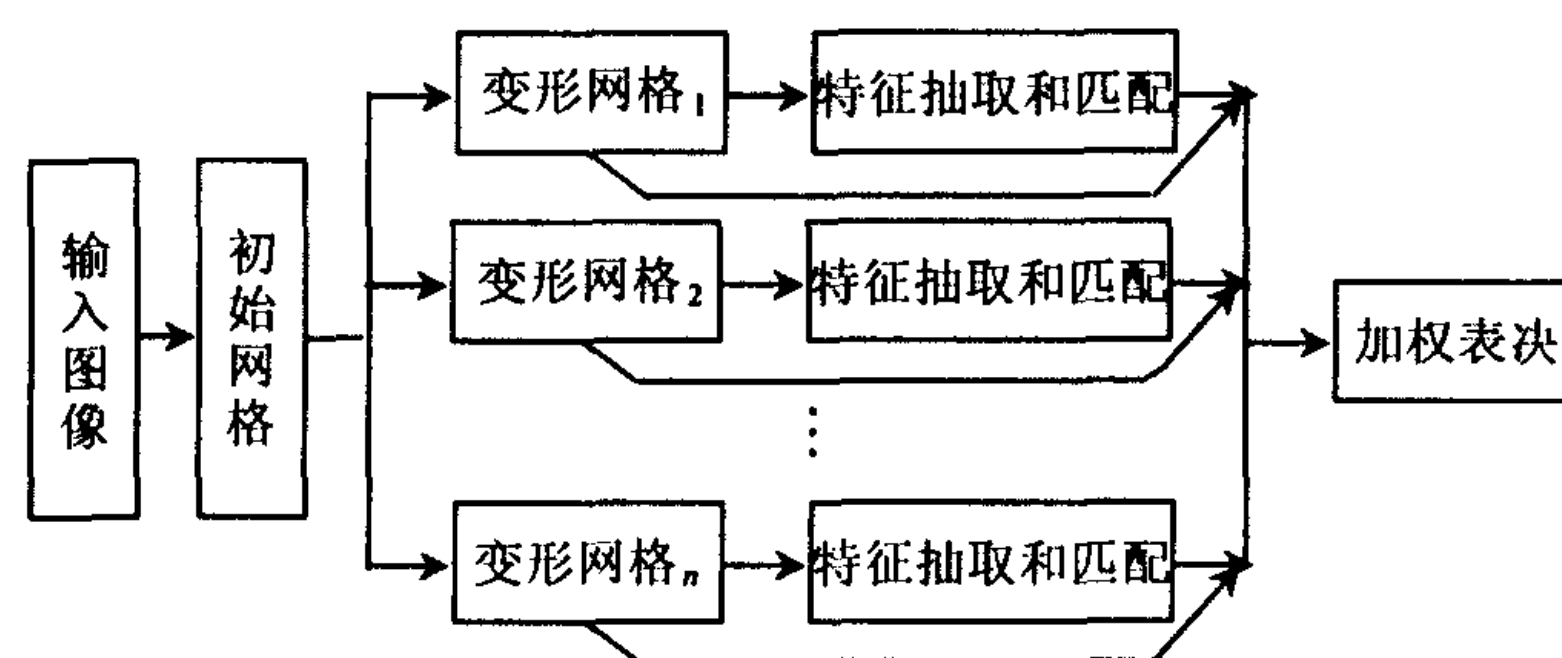


图 3 基于变形网格的识别方法

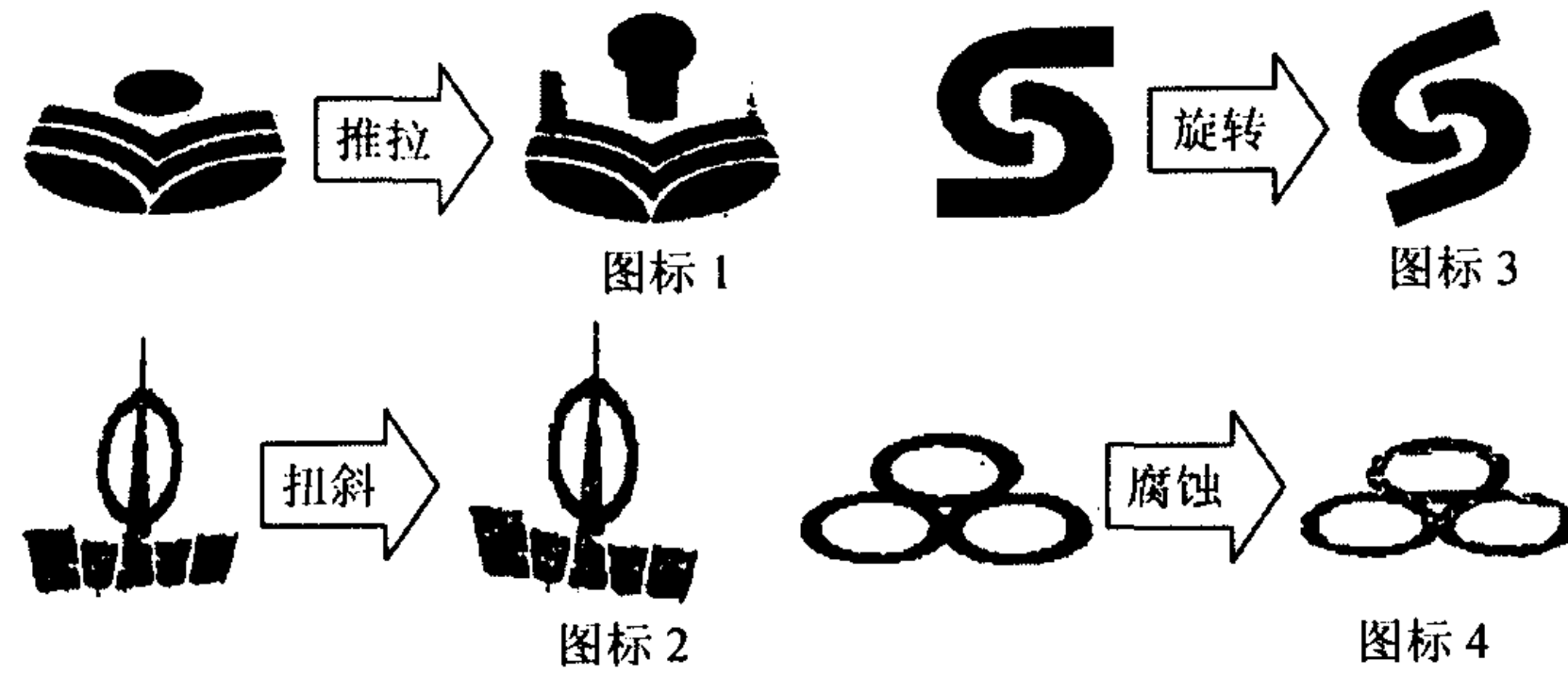


图 4 图标变形的示例

针对前 3 种变形，我们对网格采用了水平和垂直方向的推拉变形，垂直方向的扭斜变形，以及旋转变形，如图 5 中的“推拉”，“扭斜”，“旋转”所示。对于腐蚀或膨胀变形，无法通过网格变形来近似，所以没有考虑这两种变形。下面介绍我们为这 3 类变形设计的变形函数，变形函数的参数和函数值均指图像或网格的坐标。

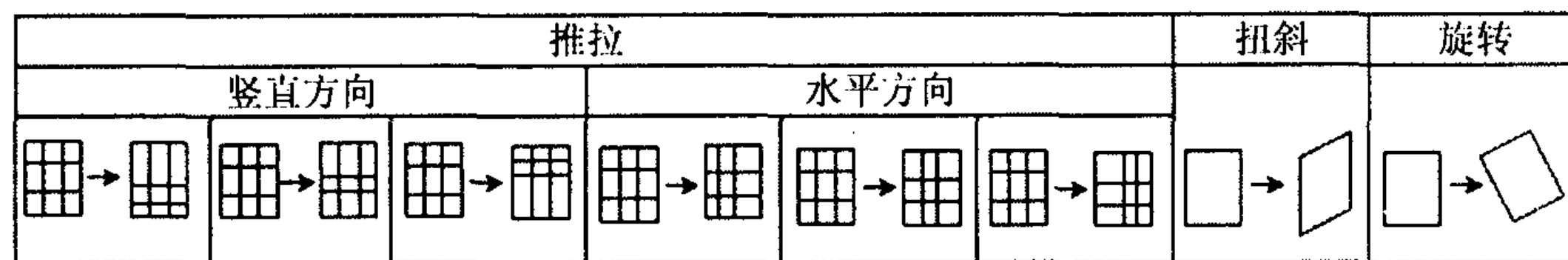


图 5 各类变形网格

3.1 推拉变形

水平方向的推拉变形和垂直方向的推拉变形类似，仅以垂直方向为例来说明。取三次函数 $f(x) = ax^3 + bx^2 + cx + d (0 \leq x \leq 1)$ 为变形函数。下面的限制是必需的：

$$f(0) = 0, \quad f(1) = 1 \tag{8}$$

$f'(x) = 3ax^2 + 2bx + c, f''(x) = 6ax + 2b, f(x)$ 的拐点 $x_c = -b/(3a)$ 。

令 $x_c = 0$ ，结合式 (8) 得 $f(x) = a(x^3 - x) + x$ ，如图 6 所示。对于给定的 x ， $f(x)$ 有向下推拉 x 的作用，而且 x 越接近 0，推拉程度越大。

令 $x_c = 0.5$ ，得 $f(x) = a(x^3 - 1.5x^2 + 0.5x) + x$ ，如图 7 所示。对于给定的 x ， $f(x)$ 有向中间推拉 x 的作用，而且 x 越接近 0.5，推拉程度越大。

令 $x_c = 1$ ，得 $f(x) = a(x^3 - 3x^2 + 2x) + x$ ，如图 8 所示。对于给定的 x ， $f(x)$ 有向上推拉 x 的作用，而且 x 越接近 1，推拉程度越大。

3.2 扭斜变形

取函数 $f(x) = g(x) + x$ 为变形函数，其中 $g(x) = ax$ ，它的图像如图 9。若 $a > 0$ 则 $f(x)$ 对于网格的右半部分有抬升作用， $a < 0$ 时则相反。

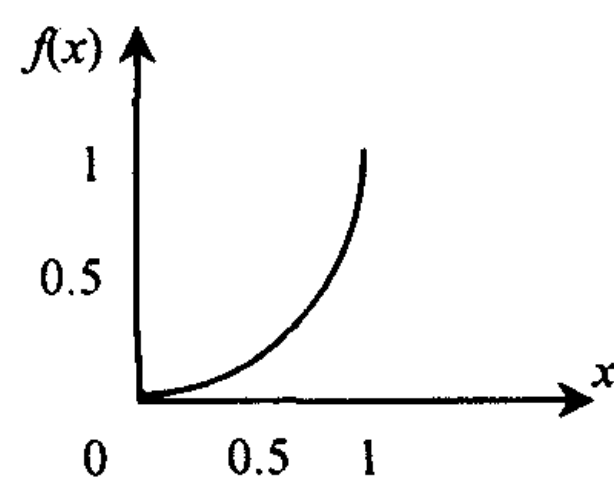


图 6 向下推拉的变形函数

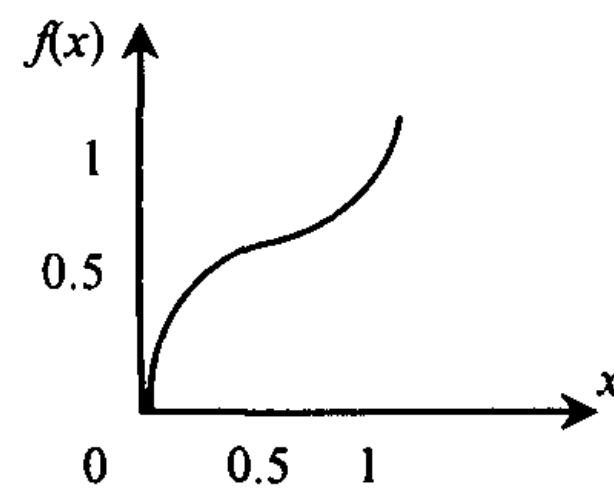


图 7 向中间推拉的变形函数

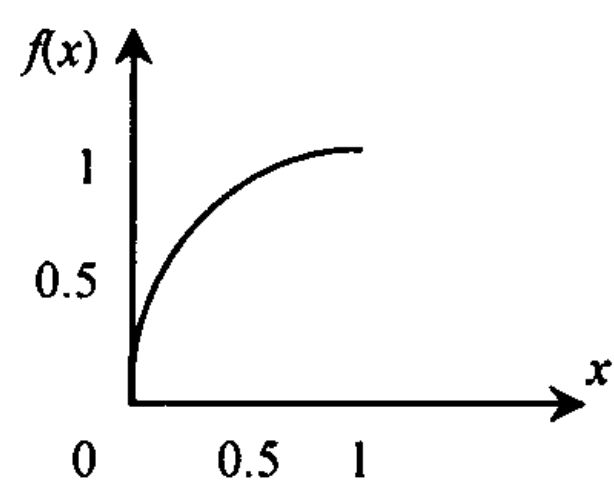


图 8 向上推拉的变形函数

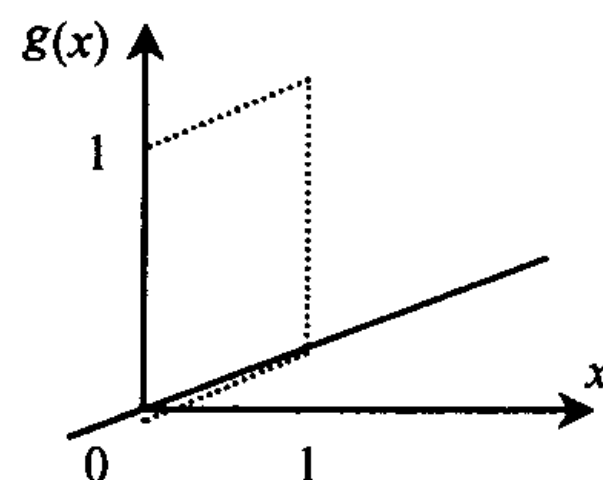


图 9 扭斜变形的函数 (虚线所示为变形效果)

3.3 旋转变形

取函数 $x' = x \cos a - y \sin a$, $y' = x \sin a + y \cos a$ 为变形函数, $a > 0$ 时有逆时针旋转的作用, $a < 0$ 时有顺时针旋转的作用。

对图 5 中列出的 8 类变形, 每类变形的控制参数 a 取 4 个值, 以重心网格为初始网格进行变形, 共得 33 个网格 (包括 1 个初始网格), 根据每个网格抽取相应的特征得到 33 个特征。

3.4 图标识别的实验结果

The University of Maryland 收集了用于研究的图标库^[7], 共含 106 个图标。这个数目比较少, 我们在这个图标库的基础上从传真图像中又收集了部分图标得到含 4060 个图标的图标库, 把这个图标库作为训练集。只有很少部分图标在多幅传真图像中出现, 对于出现次数很少的图标, 人工随机进行涂污、加噪、推拉、扭斜、旋转等变形。最后每个图标均得到 4 个变形图像用于构成测试集。虽然这些变形不能反映实际中遇到的所有变形, 但包含了大部分的变形。

我们对变形模板和变形网格进行了实验, 两者使用相同的特征, 使用的网格均是 16×16 的重心网格, 变形函数、控制参数也相同。表 1 给出了实验结果, “不使用变形方法”指对原始图像进行特征抽取和匹配。使用变形模板使识别率提高了 7.5%, 使用变形网格使识别率提高了 7.3%。

表 1 图标识别测试结果

方法	不使用变形方法的识别率	使用变形方法的识别率	识别率的提高
变形模板	85.0%	92.5%	7.5%
变形网格	85.0%	92.3%	7.3%

4 脱机手写汉字识别

我们用于验证变形网格的第 2 个实验是脱机手写汉字识别。手写汉字识别的一个最大困难在于字的合法变形极多, 不仅不同书写者写出的字风格不同, 就是同一书写者在不同条件下写出的字也有差异, 因此是很困难的图像识别问题。

我们使用的特征是自己提出的“模糊线元特征”, 网格是 8×8 的重心网格^[8], 并使用与图标识别实验中相同的变形函数、控制参数。同样地, 也对变形模板进行了实验。

中科院自动化所建立了手写汉字的样本库, 该库含 3755 个一级汉字, 每字有 100 个样本, 我们用每字的前 50 个样本建立特征库, 后 50 个字用于测试, 需要指出的是这个实验中没有“人造”样本。表 2 给出了实验结果, 使用变形模板使识别率提高了 6.1%, 使用变形网格使识别率提高了 5.8%。

表 2 脱机手写汉字识别的测试结果

方法	不使用变形方法的识别率	使用变形方法的识别率	识别率的提高
变形模板	84.3%	90.4%	6.1%
变形网格	84.3%	90.1%	5.8%

5 结论

本文针对变形模板速度慢的缺点, 对网格特征提出了变形网格法。变形网格与变形模板在变形较小时是近似相等的。我们还通过图标识别和脱机手写汉字识别两个图像识别实验说明变形网格的有效性。在实验中, 变形网格的性能略低于变形模板, 这是因为变形网格只是近似与变形模板等价, 尤其当变形较大时两者之间差距较大。另外, 无论是变形模板还是变形网格在图标识别中的性能都比脱机手写汉字识别中好, 这是因为图标的变形比手写汉字少, 同样的变形函数和变形参数能“捕捉”到更多的变形。这说明为了使变形网格充分发挥作用, 必须仔细分析图像的变形方式, 设计更好的变形函数和变形参数, 以“捕捉”更多的变形。考虑到变形网格需要的运算量远远小于变形模板, 所以变形网格是有优势的。需要指出的是变形网格只对网格特征有效, 对诸如不变矩、Fourier 描述子、正交变换系数等特征无效; 另外变形网格只能捕捉“仿射”变形, 无法捕捉腐蚀膨胀等变形。这些是变形网格的局限所在。

人脸识别、数字识别、图标识别、脱机手写字符识别等图像识别问题的变形以仿射变形为主, 当变形不大时, 变形网格有比较好的效果, 而且速度很快。当图像变形较大时, 我们现在的考虑是使用更加细密的网格, 这样会减弱特征对噪声的抵抗力, 因此需要进一步增大相邻网格间特征的“模糊化”程度来抵抗噪声, 目前我们正在进行这个方法的实验。

文献 [9] 中为了改善文本分类的性能, 由已知样本根据语义不变性构造人工样本, 并把这些样本加入训练样本, 使识别率提高约 3%。人工样本由已知样本通过同义词替换生成。按照变形网格的思路, 可以不经过人工样本, 直接把语义不变性应用于由文本生成的特征向量, 即对特征向量中的各维进行同义词替换和权值调整, 这样有与对人工样本进行特征抽取同样的效果。目前我们也正在进行这方面的实验。可以看出, 变形网格的思路有更广泛的应用。

参 考 文 献

- [1] Trier Q D, Jain A K, Taxt T. Feature extraction methods for character recognition—A survey. *Pattern Recognition*, 1996, 29(4): 641–662.
- [2] Zhang Dengsheng, Lu Guojun. A comparative study of three region shape descriptors, *Digital Image Computing Techniques and Applications(DICTA2002)*. Melbourne, Australia, 2002: 21–22.
- [3] Yan Ji-Kun, Zheng Hui, Xi Jian-Min. Logo recognition in low quality document images. *Proceedings of International Conference on Intelligent Information Technology (ICIIT2002)*. Beijing, China, 2002: 185–189.
- [4] Jain A K, Zongker D. Representation and recognition of handwritten digits using deformable templates. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 1997, 19(12): 1365–1388.
- [5] Ha T M, Bunke H. Off-line handwritten numeral recognition by perturbation method. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 1997, 19(5): 535–539.
- [6] Jain A K, Vailaya A. Shape-based retrieval: A case study with trademark image databases. *Pattern Recognition*, 1998, 31(9): 1369–1390.
- [7] <ftp://ftp.cfar.umd.edu/pub/documents/contrib/databases/UMDlogo-database.tar>.
- [8] 燕继坤, 郑辉, 张震. 基于网格变形的脱机手写汉字识别. *计算机研究与发展*, 2001, 38(增刊): 127–131.
- [9] 李辉, 史忠植, 许卓群. 运用文本领域的常识改善基于支撑向量机的文本分类器的性能. *中文信息学报*, 2002, 16(2): 7–13.

燕继坤, 男, 1973 年生, 博士生, 主要研究方向为模式识别与机器学习。

张震, 男, 1974 年生, 博士生, 主要研究方向为模式识别。

郑辉, 男, 1957 年生, 研究员, 博士生导师, 主要研究方向为信号与信息的智能处理。