一种基于变换特征和分层模型的静态手势检测方法

赵颜果 宋 展

（中国科学院深圳先进技术研究院 深圳 518055）

摘 要 本文提出一种基于变换特征和分层模型的静态手势检测方法，所采用的分层模型由一系列手势表观模型和一个总的判别模型构成，其中每个手势表观模型各包含一个通用模板和一系列子类模板。将这些模板作为转移函数，可以从原始的梯度方向直方图特征中得到一组新的特征表示，即变换特征。将此变换特征用于构造分层模型中的判别模型，可以实现背景与手势以及不同手势间的精确分类。为了提高检测速度，算法在初始阶段引入了肤色滤波器方法，用于排除大部分的非肤色区域。实验表明，所述算法能够有效处理视角变换、手势倾斜、自然形变等因素带来的手势表观波动，处理速度可达20帧/秒以上，在鲁棒性和计算效率方面均体现了明显的优势。

关键词 变换特征；分层模型；表观模型；判别模型

A Novel Method for Hand Posture Detection Based on Feature Transform and Hierarchical Model

ZHAO Yan-guo SONG Zhan

(*Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen* 518055, *China*)

Abstract This paper presents a hand posture detection method based on transform feature representation and hierarchical model. The hierarchical model comprises a series of appearance models and an overall discriminate model. Appearance model for each posture is composed of a general template as well as several sub-category templates. With all the sub-category templates as transition functions, the original gradient histogram features can be converted into a more discriminative representation form. This transform representation is used to construct the discriminative model in the hierarchy model to achieve further posture-background and posture-posture classification. Moreover, to boost the efficiency, a skin-filter is introduced to exclude a wide range of non-skin area. Experimental results show that the proposed algorithm can successfully cope with appearance variability caused by viewpoint changes, posture tilts and natural posture deformation with a detection speed up to 20 frames per second.

Keywords transform feature; hierarchical model; appearance model; discriminative model

1 引 言

手势识别是一种重要的人机交互手段，该技术已经被广泛应用在增强现实、智能家电等设备上[1]。静态手势检测是手势识别系统的关键环节，其核心是对静态手型的识别。静态手型识别法大体上可分为两类：即基于三维模型的方法和基于二维表观模型的方法[2]。三维模型的方法首先依赖于空间手势三维数据的获取，在处理速度上缺乏实时性，且三维手势模型描述参数较多，模型求解和匹配算法的复杂度较高，容易导致由于视角模糊产生的模型求解奇异性[3]。基于以上因素考虑，目前静态手势检测算法研究主要集中在二维表观模型算法领域，基于二维图像数据做分析

基金项目：国家自然科学基金（项目号 61002040），广东省创新团队-机器人与智能信息系统团队项目以及深圳市计算机视觉与模式识别重点实验室项目（项目号 CXB201104220032A）。

作者简介：赵颜果，博士研究生，研究方向为模式识别与机器学习，E-mail: yg.zhao@siat.ac.cn；宋展，博士，副教授，研究方向为计算机视觉与人工智能、三维重建、人机交互、计算机图形学。

和处理，算法复杂度一般较低。

基于表观模型的手型识别算法的主要困难来自手型形变、摄像头视角变化，以及环境光照变化等方面，近期国内外在该领域的研究也主要围绕这些问题展开，尤其是关于鲁棒性图像特征和描述是目前的研究重点，如Haar-like描述子算法[2]、肤色特征模型算法[3,4]、手形轮廓算法[5,6]，梯度方向直方图算法[7]，点对特征算法[8]等。对于手型描述算法，应用较多的包括可形变模型[6]，部件肢解模型[9]，曲线匹配模型[10]等。这些方法有的准确性好但速度慢，有的速度快但限制条件多，不具推广性。因此，寻求一种高效、准确和鲁棒的手势识别算法，依然是目前该领域的一项重要研究课题。

针对以上问题，本文提出一种分层模型用于静态手型检测，在该方法中，我们首先基于梯度方向直方图特征为每类静态手势构造一个独立的表观模型，该表观模型包含一个通用模板和几个子类模板。然后，以这些模板为转换函数，将梯度方向直方图特征转换为变换特征。最后，基于这些变换特征构造一个多类别判别模型，该模型不仅用于区分手势和背景，还可用于对不同的手势做区分。通过在不同实际应用场景下的实验结果分析和对比，证实了该方法具有显著高效性和鲁棒性的优势。

2　基于表观模型的分层建模算法

本文所述静态手势检测算法涉及线下模型学习和在线的手势检测，算法流程图如图1所示。在训练阶段，对标记好的手势样本和背景图像做学习，得到肤色滤波器和分层模型。在检测阶段，肤色滤波器首先被用作手势-背景分类，可排除大部分非肤色的滑动窗口，并将过滤后的窗口图像传递到分层模型做进一步分类。分层模型的建立分两步实现，首先是为每个预定义手势建立表观模型，其次是在背景与手势以及手势之间建立判别模型。在本方法中，表观模型以梯度方向直方图作为特征表示，判别模型则使用由梯度

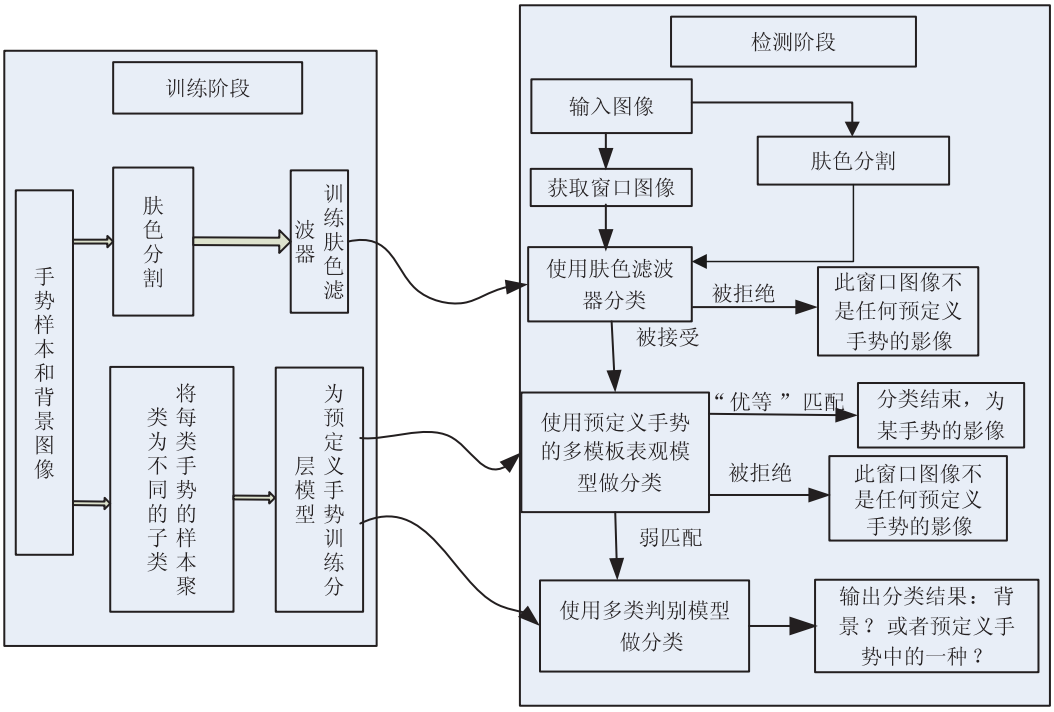

图 1　手势检测系统流程图，包括训练和检测两个阶段

直方图特征变换而来的变换特征表示。

2.1　多模板手势表观模型

如前所述，由于视角变换、手势形变、操作方式等因素的影响，同一手型在二维图像中的表观影像可能存在巨大的差异，如图2所示。这种在同类手型内部所存在的表观差异，使得通过单一模板来描述手势表观成为一件困难的事情。为了处理手势表观的差异

图 2　“拳头”手势的表观差异性

性，我们为每个手型都建立了多个模板表示，其中有

一个通用模板和若干个子类手型模板，我们期望通过多个模板能够捕捉同类手势内部的表观波动，对手势表观做出更紧致的表示。

根据多模板表观表示方法，第i个手势的表观模型可以表示为：

$$M_i=\{d_i,\{p_{(i,j)},f_{(i,j)},\tau_{(i,j)},\mu_{(i,j)}\}_{j=1}^{N_i}\}$$
$$d_i(x)=(\alpha_i,x)+\beta_i,\ p_{(i,j)}(x)=(\alpha_{(i,j)},x)+\beta_{(i,j)} \quad (1)$$
$$f_{(i,j)}(x)=1/(1+\exp(\gamma_{(i,j)}\cdot p_{(i,j)}(x)+\lambda_{(i,j)}))$$

其中(,) 代表两个向量之间的内积，d_i是通用模板，$p_{(i,j)}$是第j个子类模板，每个模板本质上都是一个有着实数输出的线性SVM分类器；$f_{(i,j)}$将$p_{(i,j)}$的实数输出转化到[0,1]区间内作为概率输出，其间以*sigmoid*函数作为转移函数；基于决策规则（通过调整分类面的方向，我们总能得到这样一个决策规则），当$p_{(i,j)}(x)>0$时，模式x被判定为目标类；$f_{(i,j)}$随着$p_{(i,j)}$的增长而增长，参数$\mu_{(i,j)}$的选择要使得当$p_{(i,j)}(x)>\mu_{(i,j)}$成立的时候，必定有$f_{(i,j)}(x)>\tau_{(i,j)}$成立，其中$\tau(i,j)$是取值于[0,1]内的一个较大的正值参数；若有$d_i(x)>0$，且存在某个$j$使得$f_{(i,j)}(x)>\tau_{(i,j)}$，则认为模式$x$与模型$M_i$之间实现“优等”匹配；若$x$不是$M_i$的“优等”匹配，但是$d_i(x)>0$且存在某个$j$使得$f_{(i,j)}(x)>0$，则认为模式$x$与$M_i$之间实现弱匹配。

在为某个手型的表观模型训练子类模板时，需先按照表观的相似性，将该类手势的样本集划分为多个子类的集合。为了降低人工分组的工作量，我们使用一个半监督算法来对样本做聚类，以梯度方向直方图作为特征表示，以概率支持向量机的概率输出作为相似性度量。该半监督聚类算法描述如下：

（1）以人工方式对每个假定的子类$S_i(1\le i\le N)$做初始化，每个子类的初始化样本不少于100枚。

（2）以S_i作为正样本集合，以$(\sum S_j)$-S_i作为负样本集合，提取这些样本的梯度方向直方图表示，基于这些表示训练具有概率输出的线性SVM模型f_i。然后通过下述方式对每个子类进行更新：

$$S_k=\{x\in\Omega;\ k=\arg_i(\max f_i(x)),\ f_k(x)\ge T\}$$

其中$\Omega=\sum S_j$为第i类手势所有正样本所形成的集合，并从经验出发，将阈值T设置为0.5。

（3）重复步骤（2）中所述的训练-更新算法模块，直至样本子集进入稳定状态。在该算法中设定，如果在一次更新中，只有极少数样本发生类别标签改变，则认为迭代进入稳定状态。

在上述聚类算法中，我们使用的是LIBSVM[11]这一公开库所提供的支持向量机算法实现。该算法虽然使我们不必再对样本做繁琐的人工分组，但是，该聚类结果并不一定完全可靠，因此，聚类完成之后，需要人工检验聚类结果，若子集内部存在较大的差异，则需剔除严重不一致样本，并将结果子集作为初始化，重复上面半监督聚类的第2-3步。尽管如此，与人工分组工作量相比，矫正分组结果的工作量是微不足道的，一般情况下，检验结果都是令人满意的，因此无需要重复聚类过程。

获得子类样本集合之后，以第j个子集作为正样本集合，以背景样本作为负样本集合，从中训练出该手型类的第j个子模板$p_{(i,j)}$。以该类所有手势样本作为正样本集合，以背景样本作为负样本集合，从中训练出通用模板d_i，通用模板是对该类手势特性的概括描述，对各个子类手势都有一定的解释能力，但是不能恰好紧致地对各个子类手势做出表达。通用模板相应的阈值要调整到足够低，使其在该类手势正样本训练集合上能有较高的识别率。

为了判定某个模式所对应的梯度方向直方图特征x是否与模型M_i相匹配，首先将x与其中的通用模板做比较，如果通用模板所拒，则代表其必定为M_i所拒。当且仅当x被通用模板和至少一个子类模板所同时接受的时候，才表明x能与M_i匹配上。进一步，如果$d_i(x)>0$，并且存在某个$0\le j\le N_i$使得$p_{(i,j)}(x)>\mu_{(i,j)}$，则认为x与模型M_i之间可实现“优等”匹配。一个优等匹配表明匹配的置信度高，因此被认为是反映了真实情况的正确分类。

上述表观模型通过多个全局模板来捕捉表观波动，与基于部件的可形变模型[12]相比，其优势体现在，在可形变模型算法中，当训练集合发生变化时，整个模型的所有参数都要被重新训练，而且训练计算量繁重过程漫长。而对于手势检测问题来说，有时需要加入一些新的子类来处理更多的形变，对于我们的模型来说，此种情况下，只需根据新增的子类样本，训练该子类的模板以及重新训练通用模板即可，而无需对其他的子类模板再做调整。

2.2 变换特征和判别模型

通过多模板的方法，我们可以使用多个手势模型独立地做多尺度滑动窗口检测，从而判定各类手势的存在与否及具体位置。但可能导致检测结果中存在大量的误判，这主要由于：（1）背景模式的复杂性，即用于训练子模板的负样本集合非常复杂，不仅包括非定义手势的样本，还包括丰富的背景图像模式，这种复杂性降低了手势类和非手势类的可区分度；（2）不同手势的相似性，即，不同手势之间可能有

非常大的相似性，这增加手势间误判的机率；（3）同种手势表观的变异性，即，由于不同用户的操作习惯，在同种手势的内部也存在着非常丰富的变差，同种手势内部的变差的存在模糊了两个相似手势之间的差异。因此，为了降低错误检测率，我们进一步在预定义手型模式和顽固背景模式、以及不同手型模式之间做出区别。

以表观模型中一系列模板作为转移函数，我们可以将原始的梯度方向直方图特征映射为一种新的特征表示：

$$x_p = x_v / \| x_v \|_2 , x_v = (v_1(x), v_2(x), \cdots, v_L(x)),$$
$$v_i(x) = (f_{(i,1)}(x), f_{(i,2)}(x), \cdots, f_{(i,N_i)}(x)) \quad (2)$$

这是一个非线性变换。一般情况下，当x来自于第i个手势的第j类的时候，$f_{(i,j)}(x)$具有较大的值，否则，$f_{(i,j)}(x)$具有较小的值，从图3给出的均值可视化中，我们可以从直观上感受到变换特征的这一特性。经此特征变换，原本低层次的梯度方向直方图表示，被转换为高层次的目标类属信息；新表示的每个维度，都代表了在给定了观察值x的条件下某个手势子类存在的置信度。这样，那些原本混乱的模式分布被转化为一个相对有组织的模式分布，因此，从直观上看，变换表示在不同手势之间以及手势与背景之间将会有更高的区分能力。

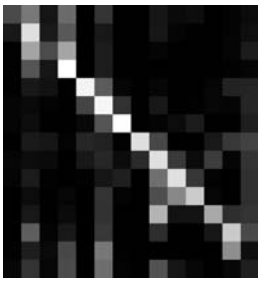

图3　变换特征表示的均值可视化。4类手势共有14个子类，对应变换特征的维度为14维。总共使用了15个数据库，其中14个数据库对应4个手势的14个子类，另外一个为背景样本数据库。
对于每一个数据库，我们计算其中所有样本的变换特征，并计算这些特征的均值。每个均值向量作为矩阵的一行，这样形成一个15×14的矩阵。为可视化方便，矩阵值被归一化到0~255之间，图中显示单元的亮度代表了相应位置矩阵元素值的大小

根据公式(2)中的特征表达方式，每个训练样本都可以被描述为：

$$X_p = \{x_p, c\}, c \in \{0, 1, \cdots, L\} \quad (3)$$

其中c=0表示x是一个背景样本，x=i>0代表x是第i类手势样本。基于此样本描述，我们可以训练一个多类判别分类器F，此处F是一个具有概率输出的多类SVM分类器，使用的是一对多的分类策略，训练代码依旧基于LIBSVM公开库。F对模式x的预测可被表达为如下形式：

$$F(x_p) = (y_0, y_1, \cdots, y_L) \quad (4)$$

其中y_i代表第i类得到的投票分，如果y_0得分最多，那么表示被分类模式最有可能来自于背景图像。此判别模型与前述的表观模型是按照级联的顺序训练的，其训练时所用样本包括，被任一表观模型判定为目标的背景样本、被任一表观模型判定为目标的非定义手势样本、不能与它本身的表观模型实现“优等”匹配的手势样本、可以与其它预定义手势的表观模型实现“优等”匹配的手势样本；这样做既符合实际检测需要、更有助于降低判别模型所遇到的模式分布的复杂性，因此模型总的特征可以被概括为：

$$\{ X = \{x, x_p\}, M = \{\{M_i\}_{i=1}^{L}, F\} \} \quad (5)$$

其中X描述了特征表示，M描述了判别规则，x和x_p之间的联系通过$\{M_i\}_{i=1}^{L}$来建立起来。表观模型中子类模板参数也要调节到一个能允许几乎所有子类手势样本通过的状态，以便总的手势模型能够涵盖尽可能宽泛的手势表观变化，判别模型着重用于排除顽固背景样本，和对相似手势做专门的区分，这样总的检测准确率可以得到提高。

3　基于分层模型的手势分类算法

手势检测的目的是判断图像中有无预定义手势，若有，找出其具体位置，以及所属的手势类别。本文采用多尺度滑动窗口方案[13]，对整幅图像的多个尺度上的几乎所有位置都做分类判断，以查看是否有预定义手势存在。由于待分类图像窗口数目众多，并且梯度方向直方图特征提取过程本身就比较耗时，这导致系统很难应用于实时处理。为了提升算法整体效率，我们引入肤色滤波器，对每个窗口图像，首先使用肤色滤波器做快速分类，被肤色滤波器所接受者才会被传递到分层模型做进一步分类。

3.1　基于肤色线索的算法加速

肤色滤波器由一系列级联起来的AdaBoost[14]分类器构成，所用特征为窗口图像的掩膜图像的局部均值。给定正负样本的掩膜图像（本段以下称之为训练样本）之后，获取单个AdaBoost分类器的方法如下：

（1）首先将正负掩膜图像规范化到标准尺寸；从标准的样本尺寸中，我们可以生成大量的位于不同位置有着不同尺寸的矩形子窗口，如图4(c)所示；提取训练样本中每个子窗口位置的平均亮度，这样，从

每枚样本中我们都能得到大量的局部均值特征；

（2）基于这些均值特征，训练一个AdaBoost分类器，该AdaBoost分类器的学习过程也即是特征选择过程，最终只有少量区分能力强的局部子窗口特征被采用。

在上述方法中，掩膜图像是依据文献[15]中所述方法，对彩色图像直接做肤色分割获得，如图4(b)所示是一组肤色分割的结果示例，其中像素值为1的点代表其为肤色像素点。使用局部均值来过滤窗口图像是基于这样一种认识：即如果一个窗口图像恰好是某个手势的影像，那么它就应该包含大量的肤色像素点，并且这些肤色像素在窗口中的位置分布应该具有一定特征，例如，窗口图像外围的肤色点密度应该比中心区域的要小、上下部分的密度比中间部分的要小等。

(a) (b) (c)

图4 肤色滤波的原理：(a)原始视频图像；(b)原始视频图像经过肤色分割后的掩膜图像；(c)窗口图像内的一些矩形区域，这些矩形区域的亮度均值将会被用于训练肤色滤波器

在训练级联肤色滤波器时，事先要将所有的手势样本和背景图像都分割为掩膜图像，然后参照文献[13]所述的人脸分类器的训练方法做训练，训练过程如下：

（1）首先在背景掩膜图像的随机尺度和随机位置采集子图像作为负样本，和正样本掩膜图像一起，依照前述方法，训练出级联分类器中第一个AdaBoost分类器；

（2）在背景图像的随机位置获取随机尺寸的窗口，使用已经存在的级联分类器对此窗口图像做分类，未被拒绝者将作为顽固样本，收集够一定数量的顽固样本之后，按照前述的AdaBoost分类器训练方法训练下一个阶段性分类器；

（3）重复（2）中方法，直至级联结构中AdaBoost分类器数目达到指定要求。由于肤色滤波器仅仅是被用作预处理，因此参数设置应该使得99%以上的手势样本都能够被接受，此外，级联分类器中AdaBoost分类器个数一般设定为3~5个就能达到比较好的过滤效果。

在检测阶段，首先对全帧图像进行肤色分割得到掩膜图像，并计算掩膜图像的积分图像。在每个检测尺度下，根据检测窗口的尺寸，对肤色滤波器所用的矩形特征的尺寸和位置做调整，这样，对于该尺度下任意窗口图像，其所对应的用于分类的均值亮度特征，能够从前述的积分图像中快速计算获取。这些特征被用于窗口过滤，最终，被肤色滤波器所接受的窗口会被传递到分层模型做进一步的分类。对于一般的场景来说，在此过滤阶段，有90%以上的非肤色窗口都能被排除掉，因此可以显著提高后续识别算法的效率和系统整体的运行效率。

3.2 基于判别模型的手势分类

多尺度滑动窗口方法产生的数以万计的窗口图像，只有少数能够通过肤色滤波，定义这些具有肤色特征的窗口集合为Ω1。对任意窗口$w\in\Omega1$，首先提取其梯度方向直方图特征x，将该特征与每个手势所对应的表观模型做比较，如果该窗口与某个表观模型能实现“优等”匹配，则它被认为包含有该手势的影像，此时不再对该窗口做进一步的分类。反之，如果窗口与模型实现弱匹配的话，需要对其执行进一步的分类。此时，根据x与多个混合模型匹配过程中所产生的实数输出$p_{(i,j)}(x)$，可以按照公式(2)得到变换特征表示x_p，根据表达式(4)获取投票得分向量$F(x_p)$，则最终的类别决策可表达如下：

$$k=\arg_i(\max\{y_i,i=0,1,\cdots,L\}) \tag{6}$$

如果$k\neq0\ \&\ y_k\geq T_p$，我们认为当前窗口对应第k个手势的图像。否则，该窗口被认为是背景图像。在本文所对应的软件系统中，按照经验将阈值T_p设置为0.4。

对于同一帧视频图像来说，当待检测的手势数目增多的时候，候选窗口集合Ω1中的窗口数目基本保持不变。在所述分层模型的分类算法中，有两个比较耗时的步骤，首先是梯度方向直方图特征提取，其次是变换特征的计算（指数操作运算）。但是，对于同一窗口图像，梯度方向直方图特征只需要提取一次，不同手势的表观模型在对窗口分类时可共用此特征。当窗口图像与某个手势模型实现“优等”匹配的时候，它不需要再与剩余的手势表观模型做匹配，也无需使用判别模型F对其做分类；当窗口图像被所有的手势表观模型所拒绝时，无需再使用判别模型F对其做分类；当且仅当该窗口图像与某个/某些表观模型实现弱匹配时，才使用判别模型对其做进一步分类，因此通常每帧中需要使用判别模型F分类的窗口数目不超过100个。根据上述分析，当待检测手势个数适当增加的时候，总的计算时间不会有较多的增长，但为每个手势建立模型却是一项比较繁重的工程，并且一

般需求下，使用过多的静态手势反而为用户带来不便，而是辅以一定的动态手势识别来实现更丰富的操作功能。

根据本节所述的检测方法，对窗口图像的分类过程，是按照从先到后的三个阶段进行的，分别是肤色滤波器分类、多模板表观模型分类、以及判别模型分类。这种级联的方案不仅有助于提高检测效率，而且逐渐降低最终的判别模型所可能遭遇的负类模式的复杂性，从而有助于提高检测的准确性。

4　实验结果与分析

在本文描述的手势检测系统中，共识别如图5所示的四种静态手势。正样本是通过让多个操作者在多种背景下自然地操作手势来获取的，每种手势总共有8000个正样本，负样本中除了背景图像之外，还包括四种预定义手势之外的一些其他手势。软件系统的实现是基于Visual Studio 2008软件平台，测试所用的PC机配置如下：AMD Athlon 2.71G CPU, 2G RAM。样本图像由普通网络摄像头获取，图像分辨率为320×240像素。

图5　手势系统中所预定义的四类静态手势，它们分别是拳头手势、剪刀手势、闭合手掌手势、和张开手掌手势

当有多个待检测手势时，第一个要解决的问题是尽量地降低漏检率，第二个是尽量地降低虚警率和误检率。漏检率主要源于手势倾斜、手势对摄像头视角偏移、手势形变等因素所带来的手势表观变化。虚警率主要是源于背景的复杂性和手势间的相似性。分层模型中的多模板表观模型是为降低漏检率而设置的，图6所示分别是文献[2]和[7]中所述方法和本文所述多模板模型的一些检测结果比较。其中，参与比较的分类器经过参数调整，在负的训练样本集上都有同样的虚警率。第一行所示是文献[2]所述的Haar特征级联分类器的检测结果，第二行所示是[7]所述的基于梯度方向直方图特征的Adaboost分类器的检测结果，第三行所示是本文所述的多模板表观模型的检测结果。从图中可以看出，当手势发生倾斜或者严重形变的时候，[2]和[7]中所述方法将发生漏检。而本文所用的多模

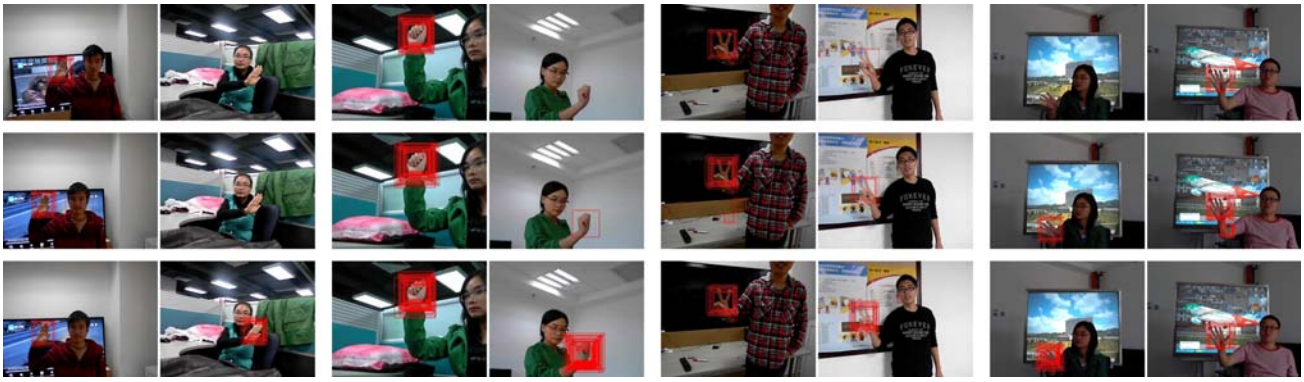

图6　从左到右四列分别对应于图5所示的四种手型的检测结果。每列左半边部分图像中的手势操作较为规范，右半边部分图像中的手势操作存在手势倾斜或者视角偏离。对每张图片只是用与它所含手势相对应的检测器做检测，而不用其它手势的检测器做检测。第一行所示是Haar特征级联分类器的检测结果[2]，第二行所示为梯度方向直方图特征结合AdaBoost分类器检测结果[7]；第三行所示为所述多模板表观模型的检测结果

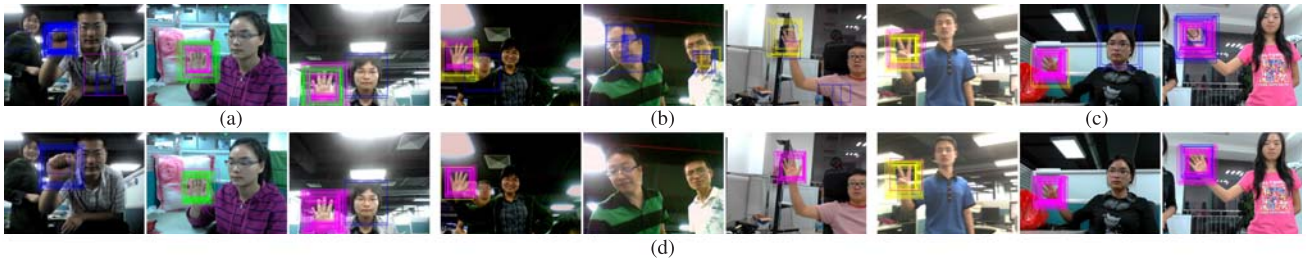

图7　(a-c)三个子图所示分别为Haar方法[2]、单模板梯度方向直方图方法[7]、和本文所述多模板表观模型的检测结果，不同手势对应分类器的检测结果被标记为不同的颜色，从中可以看出，这三种方法都存在不同程度的误检；子图(d)所示是本文所述分层模型的检测结果。四种检测方法的分类器参数设置都调整到使它们在相应的正样本集合上具有相同的正确率

板模型却能够成功地实现检测。这是因为，相比于参考文献[2,7]中的通用模板来说，多模板模型中每个模板都能更加紧致地描述一个子类。

在降低虚警率问题上，本文所述的双阈值判别法、以及基于多类判别模型的二次分类，都发挥了作用。图7所示是文献[2,7]所述方法与本文所述分层模型的检测结果比较。图7(a)是文献[2]所述的基于Haar特征的检测结果，图7(b)中是文献[7]所述的基于梯度方向直方图特征的检测结果，图7(c)是本文所述多模板表观模型的检测结果，图7(d)是本文所述分层模型的检测结果。对图5中所述四种静态手势的检测结果分别被标以“蓝”、“黄”、“绿”、“洋红”色。从检测结果我们可以看出，当手势之间具有较多的相似性或者背景比较复杂的时候，参考文献[2,7]所述方法、以及本文所述多模板表观方法都存在较多的误检，而本文所述分层模型法能够有效地抑制这种误检。这是由于在我们的分层模型中，不仅在表观模型匹配阶段使用双阈值法则，提高检测置信度，而且对存在模糊性的窗口图像做第二次分类，进而排除顽固背景、或者将相似性高的手势区分开来。

为了从量化的角度证明本文所述算法对检测精度的改善，我们计算了虚警率(FP)对正确率(TP)的变换曲线，同时也给出了文献[7]中所述方法的对应曲线作为对比，如图8所示。在测试集合中负样本数目是正样本数目的两倍以上，这是由于检测器的虚警率比较低，当负样本较少时，负样本错分数目过少，导致FP-TP曲线的变化关系不显著。图中所示曲线是虚警率(FP)和正确率(TP)各自经过对数变换之后的变化关系图，这是由于，在滑动窗口目标检测中，单帧图像所产生的非目标类窗口往往数以万计，导致检测器所遇到的非目标类模式远远多于目标类模式，这就需要检测器在分类上要保持极低的虚警率，因此，原始的FP-TP曲线往往会非常靠近x轴，导致变换曲线缺少可视性，FP和TP经对数变换后的变化关系图可视化效果得到改善。从表面上看，图8中所述两种对比曲线间的差别不明显，然而事实上这种差异是非常显著的，例如，对应同一个检测率α，如果两个分类器的虚警率差别是β，一帧图像中总的非手势窗口个数是N，那么从概率的角度来看，两方法所产成的虚惊窗口的个数差别应该是Nβ。若图中两条曲线对应于X轴上同一点的取值分别为－4和－6，N=15000，则相应的虚惊窗口数目分别是275和37，这在实际应用中的差异是非常大的。另外，所检测四个手势的曲线在不同的图形给出，是因为不同手势分类器的性能差异十分显著(这种差异源于手势表观的可鉴别性)，分别显示有助于可视化比较。

在算法实现中，虽然采用了积分图像策略来加速梯度方向直方图特征计算，但总的来说特征提取还是

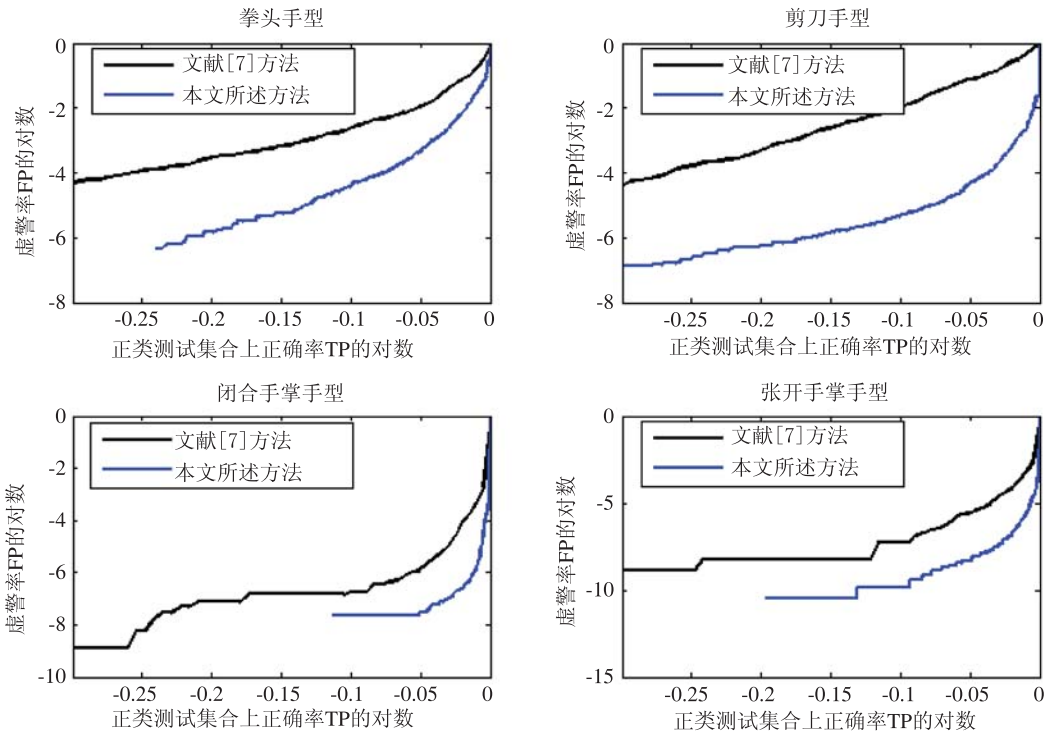

图 8 本文所述方法和文献 [7] 所述方法所生成的 FP-TP 曲线比较，四类手势各自在自己的测试集合上进行

比较耗时。实验结果显示，对于四类ZZ手势检测系统(其中每个表观模型包含3~6个模板)来说，若不使用肤色预处理，处理速度仅仅有5~6帧/秒；若使用肤色预处理，速度可显著提升至20帧/秒以上。需要注意的是，为了降低漏检，肤色滤波器的阈值被调整直到允许几乎所有的手势样本通过，尽管如此，肤色滤波器仍然可以以较小的时间代价排除掉90%以上的非肤色窗口。

5　总结与展望

本文提出了一种新的高效而鲁棒的静态手势识别与分类算法，该方法通过构建通用模板和多个子模板的联合描述方法来有效处理实际应用中手势表观描述的波动性问题，通过构造一组变换特征，提升了顽固背景模式与手势模式、以及相似手势模式之间的可区分性，在基于表观模型的匹配中引入双阈值方法，并在一定情形下使用多类判别模型做二次分类，以此来提高准确性和效率。除此之外，算法通过引入肤色滤波器来排除非肤色窗口从而显著提高了算法的效率。通过与多种现有算法的比较以及在多种复杂背景下的实际测试，表明该方法能够较好地处理复杂背景、视角偏移和手势形变问题，成功识别到大多数以自然方式操作的预定义静态手势。在使用肤色过滤器的情况下，所定义的四类手势的检测系统处理速度可达20帧/秒以上，在手势识别的鲁棒性和计算效率方面均达到了实际应用的水平。后续的研究工作将主要围绕肤色分割中对于环境光照的敏感问题，从而进一步提高系统的稳定性。

参 考 文 献

[1] Premaratne P, Nguyen Q. Consumer electronics control system based on hand gesture moment invariants [J]. IET Computer Vision, 2007, 1(1): 35-41.

[2] Kolsch M, Turk M. Robust hand detection [C] // Sixth IEEE International Conference on Automatic Face and Gesture Recognition, 2004: 614-619.

[3] Kang S K, Nam M Y, Rhee P K. Color based hand and finger detection technology for user interaction [C] // International Conference on Convergence and Hybrid Information Technology, 2008: 229-236.

[4] Xie S P, Pan J. Hand detection using robust color correction and gaussian mixture model [C] // International Conference on Image and Graphics, 2011: 553-557.

[5] Yörük E, Konukoğlu E, Sankur B, et al. Shape-based hand recognition [J]. IEEE Transactions on Image Processing, 2006, 15(7): 1803-1815.

[6] Wilkowski A, Kasprzak W. Hand gesture modeling using dynamic bayesian networks and deformable templates [C] // 7th International Conference on Signal-Image Technology and Internet-Based Systems, 2011: 390-397.

[7] Zondag J A, Gritti T, Jeanne V. Practical study on real-time hand detection [C] // International Conference on Affective Computing and Intelligent Interaction and Workshops, 2009: 1-8.

[8] Zhao X, Song Z, Guo J, et al. Real-time hand gesture detection and recognition by random forest [C] // International Conference on Computational Intelligence and Information, 2012: 747-755.

[9] Dahmani D, Larabi S. User independent system of hand postures recognition using part-based shape representation [C] // International Conference on Signal Image Technology & Internet-Based Systems, 2011: 366-373.

[10] Cinque L, Cupelli M, Sangineto E. Fast viewpoint-invariant articulated hand detection combining curve and graph matching [C] // IEEE International Conference on Automatic Face& Gesture Recognition, 2008: 1-6.

[11] Chang C C, Lin C J. LIBSVM-a library for support vector machines [EB/OL]. http://www.csie.ntu.edu.tw/~cjlin/libsvm/.

[12] Felzenszwalb P F, Girshick R B, McAllester D, et al. Object detection with discriminatively trained part based models [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, 32(9): 1627-1645.

[13] Viola P, Jones M J. Robust real-time face detection [J]. International Journal of Computer Vision, 2004, 57(2): 137-154.

[14] Schapire R E. The Boosting Approach to Machine Learning: An Overview [C] // In MSRI Workshop on Nonlinear Estimation and Classification, CA, USA, 2003.

[15] Aznaveh M M, Mirzae H, Roshan E, et al. A new color based method for skin detection using rgb vector space [C] // IEEE Conference on Human System Interactions, 2008: 932-935.