

(12) 发明专利申请

(10) 申请公布号 CN 101971190 A

(43) 申请公布日 2011. 02. 09

(21) 申请号 200980108017. 6

(51) Int. Cl.

(22) 申请日 2009. 03. 02

G06K 9/00 (2006. 01)

(30) 优先权数据

12/044, 416 2008. 03. 07 US

(85) PCT申请进入国家阶段日

2010. 09. 07

(86) PCT申请的申请数据

PCT/CN2009/070602 2009. 03. 02

(87) PCT申请的公布数据

WO2009/109127 EN 2009. 09. 11

(71) 申请人 香港中文大学

地址 中国香港新界

(72) 发明人 颜庆义 李宏亮

(74) 专利代理机构 北京英赛嘉华知识产权代理

有限责任公司 11204

代理人 余滕 王艳春

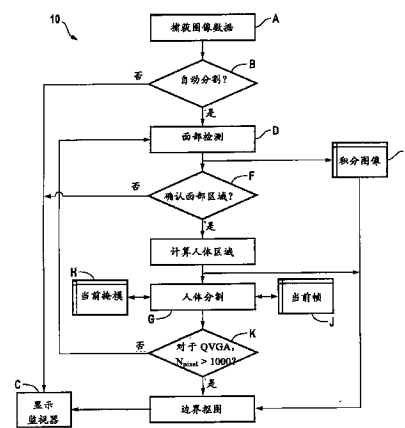
权利要求书 3 页 说明书 13 页 附图 14 页

(54) 发明名称

实时身体分割系统

(57) 摘要

公开了人特征识别系统, 该系统旨在提供基本上实时的人体部分识别, 其中提供了多种方法和结构, 以通过减少的计算需求便于实时识别。人特征识别系统包括: 在混合级联结构上使用了主动提升程序和懒惰提升程序的面部检测模块, 人体分割模块和边界抠图模块。混合级联结构是树形结构, 其中一类分类器是通过低计算量的懒惰提升得到的, 弱分类器是从前面的层得到的。



1. 一种用于在识别人的系统中识别所选的人体部分的特性和特征的方法,包括:  
捕获具有已知人体部分的人体的至少一部分的图像作为数据集;  
通过使所述数据集经过多次基于特征的拒绝测试,从代表所述图像的所述数据集中识别对应于所述已知人体部分的子数据集;  
根据适合于特性分析的、识别的子数据集片段,分割用于所述已知人体部分的所述子数据集;  
使所述子数据集片段与已知特性匹配;以及  
将所述已知特性作为输出报告给输出装置。
2. 如权利要求 1 所述的方法,其中所述已知人体部分是人脸,所述基于特征的拒绝测试使用区域方差特征集作为第一分类器。
3. 如权利要求 2 所述的方法,其中所述区域方差特征集是从在输入扫描期间用于光照修正的积分图像中构造的。
4. 如权利要求 2 所述的方法,其中所述基于特征的拒绝测试进一步使用由仅低-低非归一化 Haar 变换系数之间的差构成的特征集作为第二分类器。
5. 如权利要求 4 所述的方法,其中所述特征集由通过算术运算组合的多个类积木特征形成。
6. 如权利要求 1 所述的方法,其中所述分割步骤使用了主动提升,所述主动提升的特征在于重要性权重采样、Kullback-Leibler 置信图、基于均值漂移的区域分割以及主动特征选择。
7. 如权利要求 6 所述的方法,其中所述特征集包括两个和四个矩形块之间的多个线性组合,所述矩形块用于产生表示 Haar 变换系数的特征。
8. 如权利要求 7 所述的方法,其中特征集分类包括:  
对每个两块特征执行主动提升程序;  
对于每个提升程序记录分类误差;然后  
通过下面的方式选择具有最低误差的两块特征:  
根据误差按次序整理所述两块特征,以及  
选择多个高位的两块特征;然后  
由所述高位的两块特征组成四块特征;然后  
对每个所述四块特征执行所述提升程序;  
对于所述提升程序记录分类误差;然后  
将所述四块特征的所述分类误差与所述高位的两块特征的所述分类误差相比较;以及  
选择具有最低总误差的特征。
9. 如权利要求 6 所述的方法,其中所述主动特征选择包括利用粗略标准搜索特征,然后利用精细标准搜索特征。
10. 如权利要求 1 所述的方法,其中所述子数据集包括混合级联结构,所述混合级联结构是树形结构,其中第一节点类型表示从主动提升学习到的强分类器,第二节点类型是通过低计算量的懒惰提升得到的,并且弱分类器是从前面的层得到的。
11. 如权利要求 10 所述的方法,其中所述分割步骤包括执行懒惰提升程序,以加速匹配。

12. 如权利要求 1 所述的方法,其中所述分割步骤包括将之前的图像帧的前景信息用于当前计算,并且更新前景信息为之后的帧作准备,以促进快速多层分割。

13. 如权利要求 12 所述的方法,其中根据下面的函数更新前景信息:

$$\mu_i^F(n) = \frac{1}{W} \sum_k \mu_k^F(n) \exp\left(-\frac{1}{2} [\mu_k^F(n) - \mu_i^F(1)]^T \Sigma_k^{-1} [\mu_k^F(n) - \mu_i^F(1)]\right)$$

$$\sigma_i^F(n) = \frac{1}{W} \sum_k \sigma_k^F(n) \exp\left(-\frac{1}{2} [\sigma_k^F(n) - \sigma_i^F(1)]^T \Sigma_k^{-1} [\sigma_k^F(n) - \sigma_i^F(1)]\right)$$

其中

$$W = \sum_k \exp\left(-\frac{1}{2} [\sigma_k^F(n) - \sigma_i^F(1)]^T \Sigma_k^{-1} [\sigma_k^F(n) - \sigma_i^F(1)]\right)$$

其中,  $W$  是归一化因数,  $\mu_i^F(n)$  和  $\sigma_i^F(n)$  表示帧  $n$  的前景的第  $i$  个分量的均值和方差。

14. 如权利要求 1 所述的方法,其中所述分割步骤包括利用能量最小化自动产生三分图。

15. 一种通过识别所选人体部分的特性和特征来识别人的系统,包括:

用于捕获的代码,其捕获具有已知人体部分的人体的至少一部分的图像作为数据集;

用于识别的代码,其通过使所述数据集经过多次基于特征的拒绝测试,从代表所述图像的所述数据集中识别对应于所述已知人体部分的子数据集;

用于分割的代码,其根据适合于特性分析的、识别的子数据集片段分割用于所述已知人体部分的所述子数据集;

用于匹配的代码,其使所述子数据集片段与已知特性匹配;以及

用于报告的代码,其将所述已知特性作为输出报告给输出装置。

16. 如权利要求 15 所述的系统,其中所述已知人体部分是人脸,所述基于特征的拒绝测试使用区域方差特征集作为第一分类器,所述特征集是从在输入扫描期间用于光照修正的积分图像中构造的,并且所述特征集包括用于构建的代码,所述用于构建的代码用于将仅低-低非归一化 Haar 变换系数之间的差的特征集构建为第二分类器。

17. 如权利要求 16 所述的系统,包括用于特征集分类的代码,所述用于特征集分类的代码包括:

用于对每个两块特征执行主动提升程序的代码;

用于对于每个提升程序记录分类误差的代码;

用于通过下面的方式选择具有最低误差的两块特征的代码:

根据误差按次序整理所述两块特征,以及

选择多个高位的两块特征;

用于由所述高位的两块特征组成四块特征的代码;然后

用于对每个所述四块特征执行所述提升程序的代码;

用于对于所述提升程序记录分类误差的代码;

用于将所述四块特征的所述分类误差与所述高位的两块特征的所述分类误差相比较的代码;以及

用于选择具有最低总误差的特征的代码。

18. 如权利要求 15 所述的系统,其中所述子数据集包括混合级联结构,所述混合级联

结构是树形结构,其中第一节点类型表示从主动提升学习到的强分类器,并且第二节点类型是通过低计算量的懒惰提升得到的,且弱分类器是从前面的层得到的,其中所述用于分割的代码包括用于执行懒惰提升程序以加速匹配的代码。

19. 如权利要求 18 所述的系统,其中根据下面的函数更新前景信息:

$$\mu_i^F(n) = \frac{1}{W} \sum_k \mu_k^F(n) \exp\left(-\frac{1}{2} [\mu_k^F(n) - \mu_i^F(1)]^T \Sigma_k^{-1} [\mu_k^F(n) - \mu_i^F(1)]\right)$$

$$\sigma_i^F(n) = \frac{1}{W} \sum_k \sigma_k^F(n) \exp\left(-\frac{1}{2} [\sigma_k^F(n) - \sigma_i^F(1)]^T \Sigma_k^{-1} [\sigma_k^F(n) - \sigma_i^F(1)]\right)$$

其中

$$W = \sum_k \exp\left(-\frac{1}{2} [\sigma_k^F(n) - \sigma_i^F(1)]^T \Sigma_k^{-1} [\sigma_k^F(n) - \sigma_i^F(1)]\right)$$

其中,  $W$  是归一化因数,  $\mu_i^F(n)$  和  $\sigma_i^F(n)$  表示帧  $n$  的所述前景的第  $i$  个分量的均值和方差。

20. 如权利要求 15 所述的系统,其中所述分割代码包括利用能量最小化自动产生三分图的代码。

## 实时身体分割系统

[0001] 相关申请的交叉引用

[0002] 本申请基于 2008 年 3 月 7 日提交的第 12/044, 416 号美国专利申请并且要求其权益, 其全部内容通过引用并入本申请。

### 技术领域

[0003] 本发明涉及三维数据分析, 更具体地, 涉及用于应用计算机视觉的图案识别的对象分割。

### 背景技术

[0004] 对象分割是语义对象提取的关键技术, 并且在数字视频处理、图案识别和计算机视觉方面是有用的。分割 / 追踪视频对象形式的三维图像的工作出现在很多应用中, 例如视频监控和监视、视频摘要和索引以及数字娱乐。应用的采样包括:

[0005] - 视频监视, 其中分割结果用于允许识别入侵者或者异常情况, 并且有助于预测和揭示环境中的动作和相互作用的图案, 以确定何时应当把“警报”通知给安全单位。

[0006] - 基于内容的视频摘要, 例如体育赛事摘要, 视频预览, 视频图案挖掘, 即, 需要已分割的语义对象以执行内容分类、表现或理解的工作。

[0007] - 基于内容的编码应用, 其中视频序列的每一帧被分割为具有任意形状的语义上有意义的对象。

[0008] - 计算机视觉, 例如视频抠图、视频“艺术风格化”和渲染, 其中来自输入图像或者视频序列的已分割的二维对象能够用于 3D 场景重建。

[0009] - 视频会议和视频电话应用, 其中分割能够通过以较高质量对最相关对象进行编码而实现更好的质量。

[0010] - 数字娱乐, 其中通过分割能够取代一些具体对象, 如视频游戏。

[0011] 其他可能的应用包括工业检查、环境监控或者元数据与已分割对象的关联等。

[0012] 人体图像对象分割通常被认为是人体识别、行为分析或者人机通信的关键步骤。从图像等取得的、作为所谓人体对象的数据集和特征能够应用于很多领域, 例如视频监控、计算机视觉和视频娱乐。例如, 所提取的人体对象能够用于允许可疑行为的识别, 并且它可以帮助发现有问题的行为并且向安全中心警告可能发生的危险。

[0013] 通常, 对象分割能够分为两个阶段, 即与图案识别相关的期望对象检测和与聚类 (clustering) 技术相关的对象提取。在检测模式中, 对象分割能够以两种方式执行, 有监督的和无监督的。然而, 由于不同的对象特征, 例如颜色、强度、外形和轮廓, 通常难以自动地发现 (无监督的) 期望对象。为了避免对感兴趣对象的分割的错误检测, 已经开发了很多交互式方法, 这些交互式方法需要用户事先限定所期望的对象。由于以用户方面的交互式努力为代价而避免了对象检测的复杂步骤, 所以这些方法通常能够向用户提供比自动方法更好的分割性能。

[0014] 为了满足未来的基于内容的多媒体服务, 在真实世界的场景中迫切地需要以无监

督的方式分割有意义的对象。

[0015] 在文献中能够找到很多视频分割方法,并且通常使用空间和时间信息。空间分割方法关于颜色或强度将每一帧划分成同种区域。典型的划分方法通常被分成基于区域的方法、基于边界的方法和基于分类的方法。

[0016] 包含区域增长、分裂和合并的空间分割方法依赖于局部特征(例如,颜色、纹理、运动和其他像素统计)的同质性。时间分割方法使用主梯度信息来对对象边界进行定位。在基于分类的方法中,首先创建特征空间的划分,然后将其转换为视频信号。这种方法可以是记号(cue)的组合,例如纹理、颜色、运动和深度。空间分割方法能够产生相对准确的对象边界。然而,由于必须对于每一帧在整个图像上进行分割,所以计算复杂度充分高并且限制了用于非实时应用。此外,基于空间方法的主要问题是对于下面的“被破坏的”情况缺少鲁棒性(robustness),例如有噪声的或者模糊的视频图像,在视频图像中区域的边界经常缺失或者与其他区域混合。

[0017] 另一方面,时间分割利用运动而不是空间信息来获得对象的初始位置和边界。所谓的变化检测掩模(mask)是被并入分割处理中的动作信息的最常见的形式。因为感兴趣的对象通常是运动的,所以能够根据帧间或者背景帧基础进行变化检测。由于图像噪声的原因,对象的边界常常是不规则的,必须使用图像的空间信息进行提炼。由于边界微调程序只包含已分割的运动区域而不是整个帧,所以获得了更高的效率。然而,阴影效应、反射和噪声可能被错误地分配给前景对象。而且,通常难以辨别由真实对象运动引起的改变和由噪声、阴影效应等引起的改变。

[0018] 由于感兴趣的对象通常对应于可能具有很多空间-时间变化的多个区域,所以大多数现有的视频图像分割技术不能自动地提取图像中的对象。不使用任何主分割准则难以自动地分割这些对象。“盲分割”算法没有关于被分割对象的环境知识的假设,其内在问题是,相对于低级特性,感兴趣的对象之间可能没有同质性,或者对象可能随着环境因素(例如光照条件等)改变。

[0019] 由于这些和其他原因,需要适用于动态人体形状的改进的对象分割。

## 发明内容

[0020] 在旨在提供基本上实时识别人体部分的人特征识别系统中,提供了多种方法和结构,以通过减少的计算需求便于实时识别,包括面部检测模块、人体检测模块以及边界抠图模块。在具体的实施方式中,面部检测模块使用了主动提升程序。在另一实施方式中,面部检测模块在混合级联结构上使用了懒惰提升程序来加速对象检测。混合级联结构是树形结构,在树形结构中一类节点表示从主动提升学习到的强分类器,另一类分类器是通过低计算量的懒惰提升得到的,弱分类器从前面的层得到。

[0021] 本发明有用的贡献是能够在检测有效面部时拒绝非面部样本的、基于特征的拒绝器。该拒绝器使用了非归一化的 Haar 变换来帮助拒绝处理。其他机制包括主动提升和被动提升,被动提升利用了之前帧的前景特征。在具体的实施方式中提供的是精细和粗略的分割以及利用能量最小化技术自动产生的三分图。其他特征包括根据为实时分割专门开发的算法的方法。

[0022] 在很多的优势之中,本发明能够独立地对对象进行操作,从而使适合的编码算法

能够用于每个对象,致使主观质量的提升。

[0023] 本发明应用于由提供多媒体服务的电信公司以及诸如银行、旅馆和安全机构的服务行业提供的、预期的基于内容的实时多媒体服务。例如,电信公司能够将该技术并入产品中,以对用户提供更好的多媒体服务,其具有建立在基于内容的编码基础之上的更好的视觉质量、视频内容浏览(例如电视或电视频道)和检索以及视频游戏等。本发明还向视频监视应用提供关键技术,该视频监视应用能够为一些特定用户(例如银行、超市和旅馆)提供待识别和追踪的分割的对象(人)。此外,作为图案识别领域的重要技术,本方案的结果还可以用于护照、身份证/信用卡或者其他带有相片的证书中的面部识别。

[0024] 基于本发明的产品能够直接增强实时多媒体服务的竞争力,并且提升当地电信业的技术水平。

[0025] 通过参照以下详细描述及附图,将更好地理解本发明。

### 附图说明

[0026] 图 1A 是根据本发明的对象分割系统的实施方式的结构框图;

[0027] 图 1B 是示出了根据本发明的实施方式的系统工作的流程图;

[0028] 图 2 是人脸检测程序的简化流程图;

[0029] 图 3 是示出了基于特征集的非归一化 Haar 变换(NHT)系数的实例的图;

[0030] 图 4 示出了用作学习特征的基本单元的矩形块的形式;

[0031] 图 5 是示出了给定的归一化权重的采样结果的概率图;

[0032] 图 6 是示出了两块的特征结构的一对相关的三维图;

[0033] 图 7A 是现有技术中已知的级联分类器的图示;

[0034] 图 7B 是根据本发明的混合级联结构的图示;

[0035] 图 8 是权重系数  $w$  的计算的图解说明,其中  $F$  和  $B$  分别表示前景和背景;

[0036] 图 9 是示出了如何将矩形使用在用于粗略分割的初始区域选择中的二维图解说明;

[0037] 图 10 是示出了根据跟踪方案的分割的图;

[0038] 图 11A-11C 是蒙板问题的图示;

[0039] 图 12A-12B 示出了面部检测处理的详细流程图;

[0040] 图 13A-13F 示出了身体分割处理的详细流程图;并且

[0041] 图 14A-14D 示出了对处理进行说明的一组图。

[0042] 下面列出了图 12A-12B 中变量的定义:

[0043] SIZE\_H:帧图像的宽度

[0044] SIZE\_V:帧图像的高度

[0045] IGR(I):计算 I 图像的积分图像

[0046] IGR(I<sup>2</sup>):计算 I<sup>2</sup> 图像的积分图像

[0047] W:样本窗口

[0048] W<sub>s</sub>:样本窗口的尺寸

[0049] 分类器的数目层 Lnum = 12

[0050] 每个层中的节点 NodeNum

- [0051] 训练数据；
- [0052] 窗口  $W_T$ ；变量  $V_T$ ；变量  $V_{T12}$ ；
- [0053] 特征模式  $M[]$ ；变量  $\alpha[]$ ；阈值  $Thrshold$ ；
- [0054] 阈值  $taoT[]$
- [0055]  $V2num = 40$ ；变量分类器的层
- [0056]  $B2num = 50$ ；块分类器的层
- [0057]  $p[]$ ；学习变量的符号
- [0058]  $mv[]$ ；学习变量
- [0059] 下面列出了图 13A-13F 中变量的定义：
- [0060]  $Y, Cb, Cr$ ：表示 YCbCr 颜色空间的分量
- [0061]  $start\_Px$ ：初始窗口的左上点的坐标宽度
- [0062]  $start\_Py$ ：初始窗口的左上点的坐标高度
- [0063]  $W$ ：初始窗口的尺寸
- [0064]  $Square$ ：存储前景像素的结构
- [0065]  $R_i$ ：前景中第  $i$  个对象区域
- [0066]  $CenterNumF$ ：前景的聚类数目
- [0067]  $CenterNumB$ ：背景的聚类数目
- [0068]  $V_{PIXEL}$ ：图像的像素值
- [0069]  $Beta \leftarrow (0, 1)$ ：学习参数的权重
- [0070] Struct square
- [0071] {
- [0072] self；// 当前像素
- [0073] right；// 右邻的像
- [0074] below；// 相邻的像素
- [0075] }；
- [0076]  $\mu_R$ ：区域  $R$  的均值
- [0077]  $\Sigma_R$ ：区域  $R$  的协方差
- [0078]  $d(V_1, V_2)$ ： $V_1$  和  $V_2$  之间的欧几里德距离
- [0079]  $dis(a, b) = \exp\left(-\frac{(a-b)^2}{2\sigma^2}\right)$

### 具体实施方式

[0080] 已经发现，通过基于物理模型或者训练方案来设计适当的检测器，能够检测一些感兴趣的具体对象。图 1A 是对象检测系统 10 的框图，该对象检测系统 10 适合于通过图像元素的分割来识别人体部分。图 1B 是系统 10 的相关流程图。

[0081] 参照图 1A，通过集中在人体的问题上，环境混淆被最小化，并且促进了人体的自动分割，这特别适用于直播视频中有主要人的实时交互系统。本系统 10 具有三个关键处理：集中于例如面部检测 12 的对象检测，集中于人体结构 14 的对象分割，和边界抠图 (boundary matting) 16。输出是描述人体的可用的数据描述 17，该可用的数据描述 17 适合



用于这里被称为艺术风格化 (tooning) 18 的应用。下面将总结每个技术的详细考虑。

[0082] 参见图 1B, 首先捕获视频输入作为图像数据 (步骤 A), 如果不对图像数据进行处理 (步骤 B), 则将其引导至显示监视器 (步骤 C)。如果对图像数据进行处理, 即自动分割 (步骤 B), 那么首先应用面部检测处理 (步骤 D), 构建由面部检测处理的输出 (积分图像) (步骤 E), 并且检测积分图像, 以确认该积分图像是面部区域 (步骤 F)。如果未通过确认, 则显示该输出 (步骤 C)。如果通过确认, 则该数据用于计算身体区域 (步骤 G), 并且应用身体分割处理 (步骤 G), 与当前掩模 (步骤 H) 和当前帧 (步骤 J) 相互作用。如果分割产生像素的数目少于 1000 (步骤 K), 那么再次调用面部检测处理 (步骤 D)。否则, 结合积分图像输入 (步骤 E) 调用边界抠图处理 (步骤 L), 以产生至少用于显示 (步骤 C) 或者其他数据应用的输出结果。

[0083] 人脸检测: 作为非常不定型的对象, 人脸的特征在于尺寸、形状、颜色和肌理的高度易变性。由于面容、光照、头部姿势和表情的变化, 面部与非面部图案之间的边界也是高度非线性的。在通常的工作中, 面部检测器 12 (图 2) 以很多标度扫描图像, 以在缩放的窗口内寻找面部位置。存在三个部分或者子处理器: 肤色滤波器 18, 拒绝级联 20 和提升面部级联 (cascades-of-boosted-face) 分类器 22A、...、22N。肤色滤波器 18 用于在面部检测期间清除有色图像中的非皮肤区域。拒绝元件 20 用于移除大多数非面部的候选区同时允许面部检测的 100% 的准确度。在最后的提升面部分类器 22A-22N 中将检查有希望的类面部位置。

[0084] Cb-Cr 颜色空间中的快速肤色滤波: 已知, 在肤色滤波器 18 中, 通过感知 YCbCr 颜色空间中具有狭窄且一致分布的一定范围的色度值的存在, 能够检测肤色。在元件 18、光学滤波器和 / 或与模拟-数字转换器相连的特定波长光电检测器中, 图像扫描和位置处颜色的分类软件可以用于此目的。所使用的色度值的经验范围通常是  $Cr = [130, 175]$  和  $Cb = [75, 130]$ 。由于面部区域通常显示与不同皮肤类型无关的类似的肤色特性, 所以可能快速地消除或者跳过大多数非皮肤颜色区域, 以节省相当多的计算时间。

[0085] 假设  $S(m, n)$  表示当前图像中位置  $(m, n)$  处滤波后的二进制结果, 即, “1” 用于肤色色度值, 或者相反 “0” 用于肤色色度值。然后我们计算二进制映射的积分图像。如果没有或者只有较少肤色像素出现在这个扫描窗口  $w \times w$  中, 那么我们能够断言在当前窗口中没有发现人脸, 可以跳过该窗口的数据集。这里, 阈值设置为一个较小的值 (例如  $0.06w \times w$ )。

[0086] 基于特征的拒绝器: 基于特征的拒绝器 20 优选地是拒绝模块的级联, 拒绝模块被共同设计, 以拒绝大量的非面部样本, 而同时检测几乎 100% 的有效面部。在访问更复杂的面部分类器以实现低错误肯定率之前, 级联能够显著减少计算时间。简单的特征能够用作标志, 该标志用于构建能够作为级联的一部分的高效拒绝器模块。由于这些特性还用于在拒绝级联 20 之后的提升面部分类器元件 22, 所以对于某些特征产生需要很少或者不需要额外的计算。

[0087] 存在两个特征集: 第一特征集是区域方差, 其能够从两个积分图像中获取, 即 “积分图像” 和 “平方的图像的积分图像”。已知这些积分图像用于在图像扫描处理期间执行光照校正。对于  $24 \times 24$  像素的图像, 存在 76176 个待被考虑的方差特征。假设  $\sigma_i$  表示第  $i$  个区域的方差, 训练处理能够在算法 1 中描述:

[0088] 算法 1: 对使用方差特征的拒绝器的训练

[0089] 1. 输入训练实例  $(x_1, y_1), \dots, (x_n, y_n)$ , 其中对于非面部实例和面部实例分别有  $y_i = 0, 1$ 。

[0090] 2. 对于  $y_i = 0$ , 初始化拒绝标签  $l_i = 0$ 。

[0091] 3. 对于  $t = 1, \dots, T$ :

[0092] 4. 对于来自于面部训练实例中的每个区域  $k$  找到  $\sigma_k^l$  的最小值和最大值, 它们分别由  $\sigma_k^{-l}$  和  $\sigma_k^l$  表示。

[0093] 5. 计算对于非面部训练集的拒绝数  $r_k$ , 即

$$[0094] \quad r^p_k = \sum_{y_2=0, l_1=0} \text{sign}[p\sigma_{i,k} > p\sigma^p_k]$$

[0095] 6. 选择具有最高拒绝数的区域。

[0096] 7. 对于所有被拒绝的样本  $\{i\}$  设置标签  $l_i = 1$ 。

[0097] 8. 从步骤 3 开始重复。

[0098] 第二特征集是两个低-低 (LL) 非归一化 Haar 变换 (NHT) 系数之间的差。由于这些系数将被用于计算高频 NHT 系数作为用于训练阶段 (例如, 容易得到的自适应提升 (AdaBoost) 二进制分类学习算法作为训练阶段) 的输入特征, 所以没有与这个特征集的产生相关联的额外的计算量。基于 Haar 特征的缩放优势, 仅使用 LL 系数, 即, 以  $4 \times 4$  的块大小作为训练集, 其包括了对于  $24 \times 24$  像素图像的 97020 个特征。该训练方法与第一个类似, 从步骤 4 至步骤 6 具有小的修改, 如在算法 2 中:

[0099] 算法 2: 对使用块差异的拒绝器的训练

[0100] 4. 对于来自于面部训练实例中的两个任意系数找到  $D_{(k,j)} = LL_k - LL_j$  的最小值和最大值, 分别用  $D_k^{-l}$  和  $D_k^l$  表示。

[0101] 5. 计算非面部训练集的拒绝数  $r_k$ , 即

$$[0102] \quad r^p_k = \sum_{y_2=0, l_1=0} \text{sign}[pD_{i,(k,j)} \sigma_{i,k} > pD^p_{(k,j)}]$$

[0103] 6. 选择具有最高拒绝数的差异特征。

[0104] 使用分别用于区域方差集和 LL NHT 系数集的 40 到 50 个之间的拒绝器, 系统能够对测试数据集产生 98.269% 的拒绝率和 100% 的检测率。与 LL NHT 系数相比, 区域方差显示出相对高的拒绝率。已经观察到, 对于由 500000 个非面部图像和 12536 个面部图像组成的训练面部集, 第一方差特征拒绝器能够拒绝大约 47.6% 的非面部图像并产生 100% 的检测率。

[0105] 因为使用积分图像能够非常快地计算类 Haar 特征, 所以这些方法中的大多数通过从给定的特征集中选择一个特征来构建弱分类器。图 3 中所示的四类特征是根据不同的 NHT 系数构建的。从某种意义上说, 这些系数更像是能够根据某个风格组成建造的“积木”。每个特征由一块或多块“积木”组成, 这些“积木”通过算术运算 (具体地, 加法、减法和绝对值运算) 组合而成。在图 3 的右下栏中能发现实例。通过分别计算两个 LH 和 HL 系数之间的差值, 获得最初的两个特征。通过四块 HH 积木的加和减能够获得第六个中心环绕特征。对于  $24 \times 24$  的窗口尺寸, 过完备 (over-complete) 特征的数目非常大, 例如, 对于特征 A 是 2646, 对于特征 B 是 69768, 对于特征 C 是 830574, 对于特征 D 是 1045386。

[0106] 主动提升 (active boosting): 主动提升的特征在于特征集、重要性权重采样、

Kullback-Leibler 置信图、基于均值漂移的区域分割和主动特征选择。下面对这些特征中的每一个进行解释。

[0107] 1) 特征集:如图 4 中所示,采用矩形块的形式(栏 A)作为学习特征的基本单元。每个特征通过两个或四个块的线性组合(栏 B、C 和 D)构建。通过适当地调整这些块的组合,能够产生具有不同的类 Haar 特征的系数。假设输入实例尺寸为  $24 \times 24$ 。可能的矩形块的数目将是 76176。因此,由四个块组成的特征的总数会达到  $3.3672e+019$ 。即使对于每个弱分类器用一秒,全部运行时间仍需要  $3.8973e+014$  天,这对于训练处理来说是无法想象的。已知,特征集是过完备的,其中很多特征对于最佳分类器的贡献很小。然而,在整个特征集中进行强力搜索的过程中,它们将占去大多数的训练时间。可以以交互式的方法构建最佳特征集。

[0108] 2) 重要性权重采样:根据本发明的该实施方式的算法的第一步是来自权重分布的重要性采样,也就是消除带有低重要性权重的样本并且增加带有高重要性权重的样本。

[0109] 如本发明上面所述,被错误分类的实例将在接下来的提升阶段被分配更多权重,并且对于那些正确的实例将保持小权重。然而,样本的数目对于具有大权重的样本或者具有小权重的样本是固定的。图 5 示出了给定的归一化权重的采样结果。曲线表示通过提升程序可以获得的权重分布。能够看出,在重新采样步骤之后,那些具有大权重的实例将增加。该处理对于随后的特征选择步骤十分重要,因为该选择是基于训练数据中的信息增益。

[0110] 3) Kullback-Leibler 置信图:用  $\{(x_1^s, y_1^s), (x_2^s, y_2^s), \dots, (x_n^s, y_n^s)\}$  表示来自于原始实例的已重新采样的数据。假设样本  $x^s$  中的每一个能够被写成  $\{x^s(e_1), x^s(e_2), \dots, x^s(e_d)\}$ , 其中  $e_i$  是第  $i$  个元素(例如像素)。用  $\Omega_e$  表示给定样本的元素集。使用重新采样的数据, Kullback-Leibler (KL) 散度用于测量正样本和负样本之差。这里使用了对称的 Lullback-Leibler 距离 (KLD), 其被定义为

$$[0111] \quad KLD(P, Q) = \int q(x) \log \frac{q(x)}{p(x)} dx + \int p(x) \log \frac{p(x)}{q(x)} dx \quad (1)$$

[0112] 规则 1:如果块其中之一具有较大的 KL 散度距离且其他块具有较小的 KL 散度距离,那么能够把特征视为候选的弱分类器。

[0113] 为了根据规则 1 选择适当的特征,首先对于每个元素  $E_i \in \Omega_e$  计算 KL 散度。假设  $C(e_i)$  表示来自正样本和负样本  $x^s$  在元素  $e_i$  处的 KL 测量。这里,集合 C 被称为给定样本的 KL 置信图,该置信图描述了在每个元素处正样本和负样本之差。

[0114] 4) 基于均值漂移的区域分割:为了从 KL 置信图中发现块特征,使用均值漂移滤波来进行聚类,这能够根据置信值将置信图划分成不同区域。均值漂移是在迭代过程中估计密度分布模式的非参数方法。均值漂移程序的核(kernel)在联合密度梯度的最大增加的方向上移动。

[0115] 5) 主动特征选择:用  $(z_i, l_i)$  表示置信图 C 的分割结果,其中  $l_i$  是区域标签。分别用  $z_{\max}$  和  $l_{\max}$  表示  $z_i$  和标签  $l$  的最大值。然后使用由粗略到精细的技术进行特征搜索。

[0116] 由于选择低置信块处于半分辨率,精细阶段的任务是当通过提升阶段发现最佳候选时搜索相邻位置。在图 6 的右侧能够找到图示。此外,应该注意,这个精细处理将强加于每个高置信块上。

[0117] 在特征搜索处理中,使用了两个阈值,即  $\tau_1$  和  $\tau_2$ , 以截短两块特征的搜索范围。

$\tau_2$  的大值或者  $\tau_1$  的小值将增加特征集的数目。在当前的特定实施方式中, 设置  $\tau_1 > 0.5$  且  $\tau_2 < 0.5$ 。不同的阈值将用于不同的级联分类器。

[0118] 除两块特征外, 四块特征也被认为在本发明的考虑范围内。在两块特征提升步骤之后, 选择一些具有更低误差的特征, 并且通过线性组合评估分类误差。所选择的最终最佳弱分类器是两块 / 四块特征之中最好的一个。下面给出了详细的组合过程:

[0119] 对于每个两块特征,

[0120] 执行提升程序,

[0121] 记录分类误差,

[0122] B. 选择具有最低误差的两块特征并且根据误差按照 (上升) 顺序整理两块特征,

[0123] C. 选择最高的  $m$  个两块特征。

[0124] 从这  $m$  个两块特征中组成四块特征,

[0125] 对于每个四块特征执行提升程序,

[0126] 记录分类误差,

[0127] D. 将四块特征与最好的两块特征相比较, 并选择具有最低误差的特征。

[0128] 懒惰提升 (lazy boosting): 与主动提升学习算法不同, 懒惰提升被指定用于分类器的级联, 而且它用于提高检测性能并且显著节省计算时间。该方法将在解释特定的级联结构之后更详细地描述。

[0129] 混合级联结构: 已知直通级联结构的形式用于有效地消除尽可能多的负样本。图 7A 和图 7B 是级联分类器的图示。图 7A 是本领域所公知的级联结构。图 7B 是级联结构, 更具体地是混合级联结构。

[0130] 图 7A 示出了决策树图。级联中的每一层被调整为具有带有可预测“失败”率的较高决策率。仅允许“已接受的”样本进入下一层。在几个层的有价值的检测之后, 大多数的计算集中在少数正样本上, 而不是负样本。

[0131] 根据本发明的实施方式, 提供了混合级联数据结构以加速对象检测和所需的处理。这种树形的级联结构示出在图 7B 中, 其中节点  $C_i$  表示从主动提升中学习到的强分类器。准备好的节点  $C_i'$  表示通过懒惰提升得到的强分类器。弱分类器来自于前面的层, 如在图 7B 中用表示数据路径的虚线所示。由于没有用于在  $C_i'$  中发现弱分类器的额外计算, 所以检测性能将很高效, 并且只需要一些乘法和加法来构造强分类器。下面将描述懒惰提升的详细说明。

[0132] 懒惰提升用于混合级联结构。该方法的目的在于尽可能高效地利用弱分类器。在图 7A 所示的常规分类器的级联中, 可以从几十万个特征中选择在某一层中的每个弱分类器, 其是给定的特征集中最佳的一个。然而, 这些最佳特征仅使用一次。每个层的目的是找到实现所需检测率的最佳强分类器。

[0133] 在根据本发明的实施方式的提升算法 (即懒惰提升算法) 中, 检测效率是焦点, 而不是每个弱分类器的最佳性能。懒惰提升算法中的弱分类器可以不是最好的一个, 因为弱分类器选自前面的层。由于在懒惰提升步骤之前已经计算了这些弱分类器的输出, 所以对于它们不需要计算。计算是最后的分类器的焦点, 其只包括一些乘法和加法。下面给出详细的懒惰提升算法:

[0134] 懒惰提升算法

[0135] 1. (输入)

[0136] 输入训练实例  $(x_1, y_1), \dots, (x_n, y_n)$ , 其中  $x_i$  表示样本, 并且对于正样本和负样本,  $y_i = 1$  或  $-1$ 。

[0137] 2. (初始化)

[0138] 初始化样本权重  $w_i = 1/2p$  和  $1/2q$ , 其中  $p$  和  $q$  是正样本和负样本的数目。

[0139] 3. (懒惰提升)

[0140] 对于  $t = 1, \dots, T$  重复。

[0141] a. 归一化权重  $w_{t,j}$

[0142] b. 从前面的层中获得弱分类器

[0143] c. 对于给定的弱分类器计算分类误差

[0144] d. 选择具有最低误差的最佳弱分类器

[0145] e. 更新权重

[0146] 4. (输出)

[0147] 输出已组合的分类器

[0148] 在人分割中的贡献: 在面部检测之后, 能够快速访问输入图像中的面部位置。下面描述最小切割 (min-cut) 优化方法, 以在获得的面部区域中执行面部分割算法。

[0149] 通常, 当用户的输入明确地限定了前景或背景时, 能够根据图形切割优化来执行分割处理。遗憾地, 在无监督的方式中, 将只存在从提升面部检测器中获得的粗略面部位置。相应的子窗口可以覆盖完整的面部轮廓或者只覆盖面部区域的一些部分。这意味着窗口外的像素可以属于背景, 且窗口内的像素可能是面部的一部分。无法有把握地确定哪个像素应该被标记为背景或前景。下面描述在过不完备标签条件下分割面部区域的技术。

[0150] A. 代价函数: 由一组节点  $V$  (例如像素或区域) 和一组连接这些节点的有向边  $E$  来定义图  $G = \langle V, E \rangle$ 。如果每个像素表示一个节点, 那么图像的分割  $Z = \{z_i\}$  能够通过基于两个代价函数对能量函数进行求解来表示, 两个代价函数是: 用于将每个节点  $i$  分配给前景或背景标签的数据代价  $E_1$ , 以及用于测量两个节点之间相似度的平滑代价  $E_2$ :

$$[0151] \quad E(Z) = \sum_{i \in V} E_1(z_i) + \lambda \sum_{(i,j) \in E} E_2(z_i, z_j) \quad (2)$$

[0152] 数据代价  $E_1$ :  $E_1$  项用于设置将每个像素分配给前景或背景的损失。通常, 在交互式方法中, 两个终端  $F$  和  $B$  或者至少终端  $B$  已由用户的输入来定义。这样, 能够施加一些硬约束, 以通过最小切割程序来保证与用户的绘图一致的标签。这意味着当所分配的标签与用户的绘图笔画相违背时, 可能会应用无穷的代价。然而, 在自动处理中没有预定义的终端。因此, 使用了下述  $E_1$  的新颖的推导。

[0153] 在具体的实施方式中, Gaussian 混合模型 (GMM) 用来根据初始的面部位置对前景或背景的颜色分布建模。对于具有颜色  $z_i$  和标签  $\alpha(i)$  的给定的像素  $i$ , 如下定义像素  $i$  到前景  $F$  和  $B$  的距离:

$$[0154] \quad E_1(\alpha(i) \in F) = \frac{\log(d_i^F)}{\log(d_i^F) + \log(d_i^B)} \quad (3)$$

$$[0155] \quad E_1(\alpha(i) \in B) = \frac{\log(d_i^B)}{\log(d_i^F) + \log(d_i^B)} \quad (4)$$

[0156] 其中

$$[0157] \quad d_i^F = \sum_{k=1}^K w_k^F \left( \frac{1}{1+2 \det \Sigma_k} \right)^{1/2} \exp \left( -\frac{1}{2} [z_i - \mu_k^F]^T \Sigma_k^{-1} [z_i - \mu_k^F] \right) \quad (5)$$

$$[0158] \quad d_i^B = \beta d_i^{B1} + (1-\beta) d_i^{B2} \quad (6)$$

$$[0159] \quad d_i^{B1} = \sum_{k=1}^K w_k^B \left( \frac{1}{1+2 \det \Sigma_k} \right)^{1/2} \exp \left( -\frac{1}{2} [z_i - \mu_k^B]^T \Sigma_k^{-1} [z_i - \mu_k^B] \right) \quad (7)$$

$$[0160] \quad d_i^{B2} = \left( \frac{1}{1+2 \det \Sigma_i^O} \right)^{1/2} \exp \left( -\frac{1}{2} [z_i - \mu_i^O]^T \Sigma_i^{-1} [z_i - \mu_i^O] \right) \quad (8)$$

[0161] 其中,  $w_k$  表示对于 GMM 的第  $k$  个分量与空间样本的百分比相对应的权重,  $\mu_k$  和  $\Sigma_k$  分别表示均值颜色和协方差矩阵。K 表示在前景和背景中 GMM 的分量的数目。图 8 示出了权重计算的实例, 其中, 左边和右边分别表示前景 F 和背景 B。根据画圈区域内聚类的像素的数目能够计算像素的空间样本 (在 F 区域和 B 区域之间)。

[0162] 平滑代价 E2 :E2 项用于设置两个节点 (例如两个像素) 之间的不连续性的损失。当在像素  $i$  和  $j$  之间发现较小的变化时, E2 变得更大, 这意味着在相邻像素之间出现边缘的可能性较小。一般格式是根据局部的强度梯度来定义的。这种新颖的推导允许基于梯度但不通过标签约束采用指数函数。

[0163] 背景学习 :每个节点和背景之间的距离由两部分组成, 该距离通过加权系数  $\beta$  混合。根据当前帧  $i$  中背景的颜色分布计算第一  $d_i^{B1}$ , 且从初始化中学习得到第二  $d_i^{B2}$ 。通过调节权重  $\beta$ , 我们或多或少地能够将先前的背景信息合并到当前的边连接计算中。例如, 如果  $\beta$  设置为 0, 那么到每个节点的背景的连接仅依赖于现有知识, 对于那些静态场景 (例如视频监控) 能够得到现有知识。相反地, 对于动态场景, 尤其是对于那些快速移动背景, 应该考虑更大的权重。

[0164] 背景学习中的每个像素被建模为高斯分布, 其具有均值  $\mu_i^O$  和方差  $\Sigma_i^O$ 。它们是从系统的初始化中学习到的。我们使用 500 个帧来估计模式参数。

[0165] B. 由粗略到精细的分割 :必须从不完整标记的人体区域中建立颜色分布。所推荐的方法工作在两个等级上, 即粗略尺度和精细尺度。

[0166] 最初的分割在粗略等级上进行。如在图 9 中所示 [TLO :新颖], 存在用于估计前景信息的四个区域 (即 a ~ d), 同时存在用于估计背景的九个区域 (即 A ~ J)。例如, 选择七个部分, 即 A、(B+E)、(C+F)、(D+E)、(F+G)、H 和 I, 用于估计背景的颜色分布。采用它们的均值和方差作为背景的初始聚类, 这意味着在背景中存在四个分量用于对 GMM 进行建模。而对于身体区域, 设置具有四个分量的初始高斯模型。初始面部区域设置为小的方形窗口, 该窗口位于检测窗口的中心, 大小为  $W/2 \times W/2$ 。相应的均值和方差被认为是面部区域的初始聚类。对于这个区域中的每个像素, 根据 (3)、(4) 和与其他的相似性计算面部和背景聚类之间的加权距离。最后, 使用最小切割来进行整体优化。

[0167] 第二等级是更精细的分割, 其目的在于改进初始分割结果。相应的前景区域被限定为当前窗口中属于身体端的像素集, 而背景区域由该窗口外被分类为背景端的像素组成。我们分别使用 8 个分量来描述前景颜色和 8 个分量来描述背景颜色。基于 K 均值算法

来估计分量  $k$  的均值和协方差。然后,可使用与粗略尺度类似的方法来计算能量  $E1$  和  $E2$ 。注意,在数据代价函数  $E1$  中,根据位于当前节点中心的、 $20 \times 20$  个像素的限定窗口中的空间样本来估计每个分量的权重。最小切割也被用于最终的优化。

[0168] 基于跟踪的人分割的贡献

[0169] A. 最佳位置预测 :通过追踪技术实现连续帧中的身体分割。在当前帧  $n$  中跟踪身体的第一步是在前面的帧  $(n-1)$  中预测状态空间。这里状态空间指的是身体区域的位置。

[0170] 对于前面的帧,根据运动估计来获得候选身体区域的当前位置。在具体的系统中,运动估计技术用于根据由粗略到精细的精细化策略理念来获得候选焦点区域的投影位置。首先,在一半空间分辨率的采样空间中使用  $15$  的搜索窗执行全搜索运动估计。然后,对于最佳匹配位置,只对八个相邻点执行更精细尺度的搜索。绝对差之和 (SAD) 用作两个区域之间的相似性测量。

[0171] 接着,二阶自动回归 (AR) 模型用于描述状态改变。当对象平滑移动时,AR 模型通常能够提供更好的预测结果。然而,人对象经常因突然移动而改变方向,这导致了对随后分割的不准确预测。为了避免这种恶化的情况发生,在预测位置和原始位置之间进行误差校验,并选择最优的一个。此外,位于预测中心的小窗口用于找到最佳匹配位置,对于运动情况和静止情况,该小窗口被分别优选地设置为  $5 \times 5$  个像素和  $11 \times 11$  个像素。

[0172] B. 多级分割 :在状态预测后,获得了当前帧  $n$  中的粗略掩模。为了减少计算时间,使用三种模式进行身体分割。第一种模式基于三分图 (trimap),其由三个区域组成,即,前景区域、背景区域和未知区域。如图 10 中所示, $O_{n-1}$  表示第  $(n-1)$  帧中的候选面部区域。通过第  $n$  帧中的预测来获得投影区域  $O_n$ 。然后对投影区域  $O_n$  进行腐蚀和膨胀形态操作。所获得的区域分别由  $O_n^o$  和  $O_n^d$  表示。所使用的结构化元素是宽度为  $10$  个像素的正方形结构化元素。使用基于由等式 (2)、(3) 和 (6) 表示的函数的处理,能够确定未知区域中节点的数据代价和平滑代价。然后使用最小分割来将标签分配给每个节点。

[0173] 由于节点集很小,所以第一模式运行的很快,且它还能够得到与当人体没有意外改变时的精细分割类似的分割结果。然而,为了避免在第一模式中由噪声效应引起的可能的累积误差,需要第二模式 (即精细分割) 来修正这些错误。如图 10 所示,在身体区域和未知区域中的像素被认为是所构建的图的节点。然后使用最小分割来将标记分配给每个节点。

[0174] C. 前景模型的更新 [TLO :系统实时特征的新颖和关键] :为了执行快速多级分割,使用了之前帧的前景信息。在当前帧的分割之后,通过下面的函数更新前景 :

$$[0175] \quad \mu_i^F(n) = \frac{1}{W} \sum_k \mu_k^F(n) \exp\left(-\frac{1}{2} [\mu_k^F(n) - \mu_i^F(1)]^T \Sigma_k^{-1} [\mu_k^F(n) - \mu_i^F(1)]\right)$$

$$[0176] \quad \sigma_i^F(n) = \frac{1}{W} \sum_k \sigma_k^F(n) \exp\left(-\frac{1}{2} [\sigma_k^F(n) - \sigma_i^F(1)]^T \Sigma_k^{-1} [\sigma_k^F(n) - \sigma_i^F(1)]\right)$$

[0177] 其中

$$[0178] \quad W = \sum_k \exp\left(-\frac{1}{2} [\sigma_k^F(n) - \sigma_i^F(1)]^T \Sigma_k^{-1} [\sigma_k^F(n) - \sigma_i^F(1)]\right)$$

[0179] 其中  $W$  是归一化因数, $\mu_i^F(n)$  和  $\sigma_i^F(n)$  表示帧  $n$  的前景的第  $i$  个分量的均值和方差。

[0180] 边界抠图的贡献 :抠图是重要操作,其通过估计在每个像素处前景元素的不透明

性而使任意形状的前景元素与图像分离。由于等式 9 中示出的抠图等式具有太多的未知量,所以抠图固有地是欠约束的。

$$[0181] \quad I = \alpha F + (1 - \alpha) B \quad (9)$$

[0182] 从等式 8 中能够看出,当蒙板  $\alpha$  从 1(或 0)变为 0(或 1)时,将存在从前景(或背景)到背景(或前景)的逐渐的颜色过渡。由于  $\alpha$  的值限定在  $[0, 1]$  的范围内,所以在抠图区域中的原始数据  $I$  必须是前景和背景之间的某一个值。

[0183] A. 自适应的三分图:自适应的三分图可以用于解决抠图的问题。图 11A 示出了典型的抠图实例,其中在 F 区域和 B 区域之间存在逐渐改变的前进区。上部区域具有在前景和背景之间的不同的过渡处理。下部区域示出了  $\alpha$  蒙板的曲线。此外,还存在很多能够在背景和前景之间观察到明显边界的情况,这意味着在两个区域之间没有出现混合效应。当在图 11B 示出的未知区域中较好地估计前景和背景颜色分布时,根据在第一种情况中的类似方法,能够求解出相应的  $\alpha$  值(1 或 0)。然而,这种方法在一些情况下会失效,在这些情况中两个区域的分布其中之一不能从相邻的区域中估算出。例如,在图 11C 中所示,存在明显的边界并且在前景和背景之间没有混合区域,虽然在前景区域内能够观察到过渡区域。如果假设两条虚线之间的区域为对该区域执行  $\alpha$  计算的未知区域,那么优化的蒙板将使未知区域中的前景区域平滑,这导致了明显的人工痕迹。这样的病态条件可以通过使用人工限定的三分图来避免,即三个区域: $\Omega_F$ (“明确的前景区域( $\alpha = 1$ )”)、 $\Omega_B$ (“明确的背景区域( $\alpha = 0$ )”)和  $\Omega_U$ (“未知区域( $\alpha \in [0, 1]$ )”),但是其难以通过无监督的方式处理。不幸的是,对于人面部边界通常观察到最后一种情况。

[0184] 为了避免无法管理的情况并且减少可能的估计误差,根据本发明的实施方式,使用自适应的方法来构建三分图。通常,通过使用形态腐蚀和膨胀来根据初始对象边界产生未知区域,能够实现自动地产生三分图。使用了一致的结构尺寸,例如 5 个像素。然而,根据本发明特定的实施方式,三分图的尺寸依赖于居中的像素关于其相邻像素的模糊度。这意味着如果这个区域在前景和背景之间较好地混合,那么将需要较大的尺寸;否则将考虑小的未知区域。由于抠图从闭合的面部轮廓开始,所以在初始边界上对每个像素计算三分图。半径为  $r$  的圆用来通过分别移除和膨胀轮廓的一侧产生未知区域。用  $r_p$  表示面部轮廓上像素  $p$  的半径。这产生了:

$$[0185] \quad r_p = r_{\max} \exp(-k ||I_p - (g * I)_p||) \quad (10)$$

[0186] 其中  $g$  是具有标准差  $\sigma (= 5)$  的高斯函数,其用来通过卷积运算估计平滑度。 $r_{\max}$  表示最大半径,在我们的工作中该最大半径设置为 5。

[0187] 根据初始面部边界的腐蚀和膨胀操作以及未知区域的相应尺寸  $r$ ,产生了新的三分图。该处理对于初始面部边界像素  $p$  的任一侧上的前景和背景腐蚀了和膨胀了少于  $r_p$  个像素。这意味着相应的结构化元素使用了具有计算过的半径  $r$  的圆形结构元素。

[0188] B. 使用能量最小化的 Alpha 估计:为了估计未知区域  $\Omega_U$  中的  $\alpha$ ,定义下面的能量函数用作定义:

$$[0189] \quad E = E_1 + \lambda E_2$$

[0190] 其中  $E_1$  是误差能量,以测量数据估计度, $E_2$  是平滑能量,表示相邻  $\alpha$  值之间的变化。

[0191] 能量  $E_1$  被定义为:



$$[0192] \quad E_1 = \sum_{(m,n) \in \Omega_U} \left[ I_{m,n} - \alpha(m,n) \hat{F}_{m,n} - (1 - \alpha(m,n)) \hat{B}_{m,n} \right]^2 \quad (11)$$

[0193] 其中

$$[0194] \quad \hat{F}_{m,n} = \frac{1}{L} \sum_{(j,k) \in \delta_{m,n}^F, \alpha(j,k) \geq 0.95} w_{j,k} F_{j,k} \text{ 以及} \quad (12)$$

$$[0195] \quad \hat{B}_{m,n} = \frac{1}{L} \sum_{(j,k) \in \delta_{m,n}^B, \alpha(j,k) \leq 0.05} w_{j,k} B_{j,k} \quad (13)$$

[0196] 这里, L 是权重  $w_{j,k}$  的归一化因数, 其对于前景和背景分别被定义为  $\alpha^2 d$  和  $(1 - \alpha)^2 d$ 。d 表示关于像素 (j, k) 的距离。 $\delta_{m,n}^F$  和  $\delta_{m,n}^B$  表示位于像素 (m, n) 中心的区域。在我们的工作中, 使用半径为 11 的圆形区域来执行估计。

[0197] 平滑项 E2 被定义为:

$$[0198] \quad E_2 = \sum_{(m,n) \in \Omega_U} \min \{ (f_k * \alpha)_{m,n}^2, k = 0, 1, \dots \} \quad (14)$$

[0199] 其中,  $f_k$  表示对应于方向 k 的滤波器。E2 测量某一方向上相邻  $\alpha$  的变化, 其促进了  $\alpha$  平滑地改变。在我们的工作中, 基于四个 4 抽头滤波器、关于 0、 $\pi/4$ 、 $\pi/2$  和  $3\pi/4$  方向执行滤波处理。

[0200] 为了使能量函数 E 最小化, 使用梯度下降优化方法。能量 E 的梯度能够写成:

[0201]

$$\nabla E = 2 \sum_{(m,n) \in \Omega_U} \left[ I_{m,n} - \alpha(m,n) \hat{F}_{m,n} - (1 - \alpha(m,n)) \hat{B}_{m,n} \right] \left[ \hat{B}_{m,n} - \hat{F}_{m,n} \right] + 2 \sum_{(m,n) \in \Omega_U} (f_{\min} * (f_{\min} * \alpha))_{m,n}$$

[0202]

$$(15)$$

[0203] 在每次迭代中,  $\alpha$  值被更新为:

$$[0204] \quad \alpha_{n+1}(m,n) = \alpha_n(m,n) - \tau \nabla E_{m,n} \quad (16)$$

[0205] 其中  $\tau$  是步长, 其用于沿着方向使能量最小化并且被设置为 1.5。

[0206] 能量最小化问题是迭代优化过程, 其由三个步骤组成。第一步是根据等式 12 和等式 13 在未知区域  $\Omega_U$  中对 F 和 B 的初始化。第二步是基于梯度下降优化对  $\alpha$  值的更新。初始  $\alpha_0(x, y)$ ,  $(x, y) \in \Omega_U$  被设置为 0.5。然后等式 16 用于执行蒙板值的更新。最后一步是 (F, B) 精细化。如果满足条件  $\alpha(x, y) > 0.99$ ,  $(x, y) \in \Omega_U$ , 将通过将该蒙板值设置为 1 更新面部。对于  $\alpha(x, y) < 0.01$ ,  $(x, y) \in \Omega_U$  的情况, 像素被分类为背景。否则, 在  $0.01 \leq \alpha(x, y) \leq 0.99$  的情况下, 像素 (x, y) 的值将被作为前景  $\Omega_f$  和背景  $\Omega_b$  的混合结果处理。因此我们不把那些像素分类为分割的前景区域。

[0207] 为了增强本发明, 参考了图 12A-12B、图 13A-13F 和图 14A-14D, 其中图 12A-12B 示出了面部检测处理的详细流程图, 图 13A-13F 示出了身体分割处理的详细流程图, 图 14A-14D 示出了图示了处理的一组图形。

[0208] 参照一个或多个具体实施方式解释了本发明。在本领域技术人员审阅本说明书后, 其他实施方式是对他们是显而易见的。因此, 除了权利要求所指出的之外, 本发明不是限制性的。

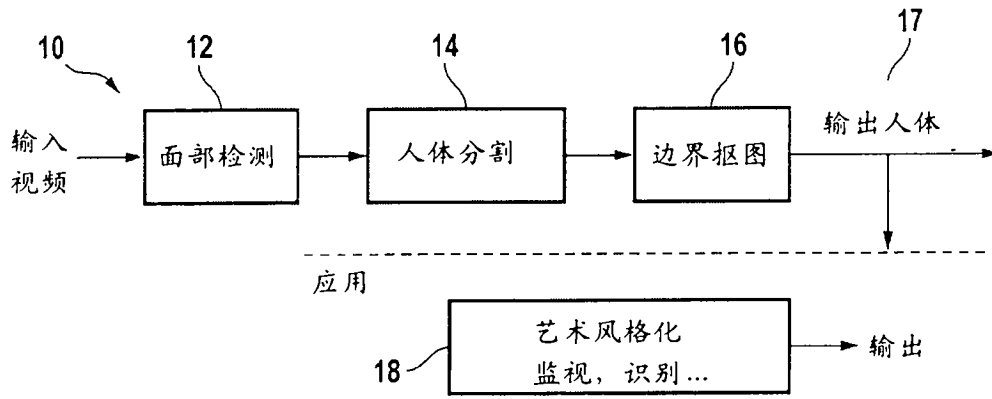


图 1A

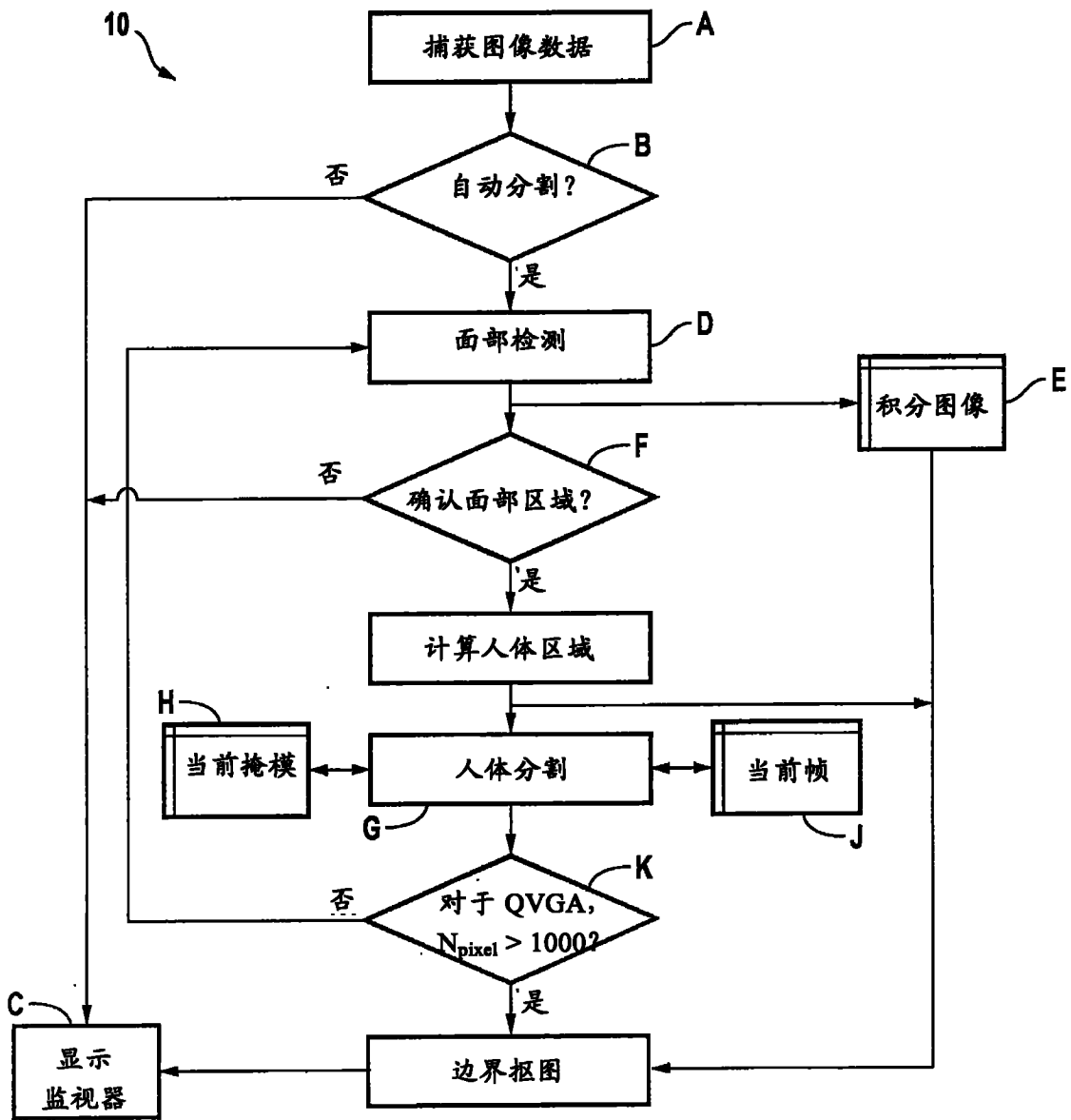


图 1B

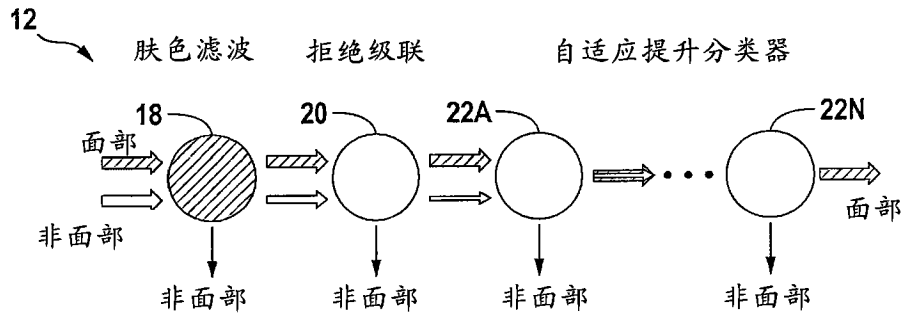


图 2

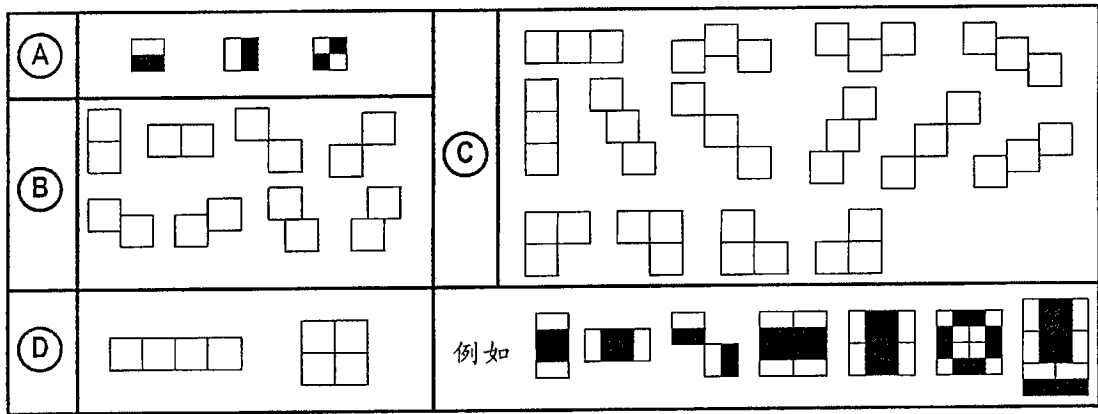


图 3

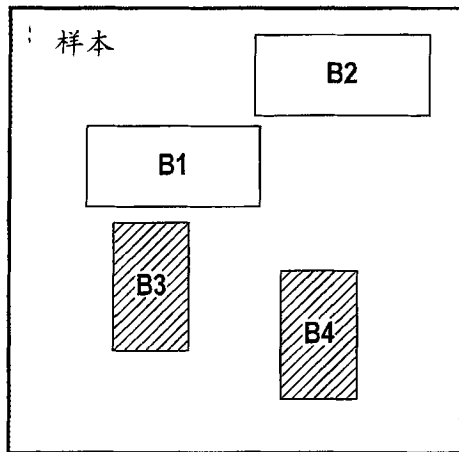


图 4

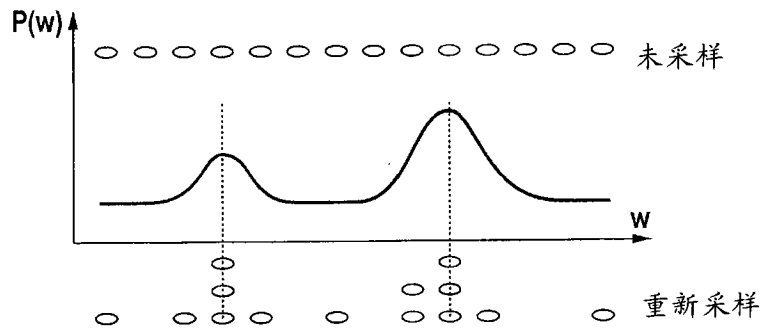


图 5

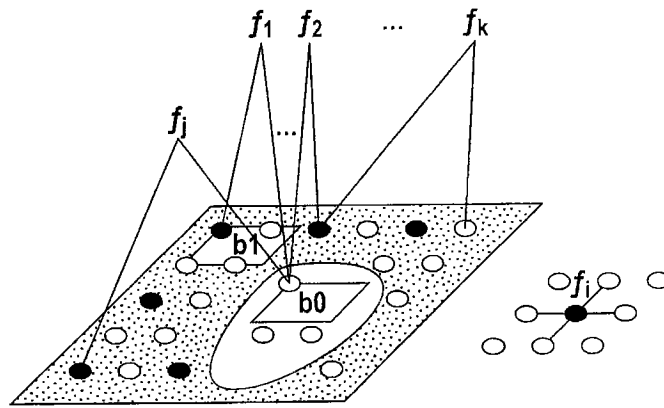


图 6

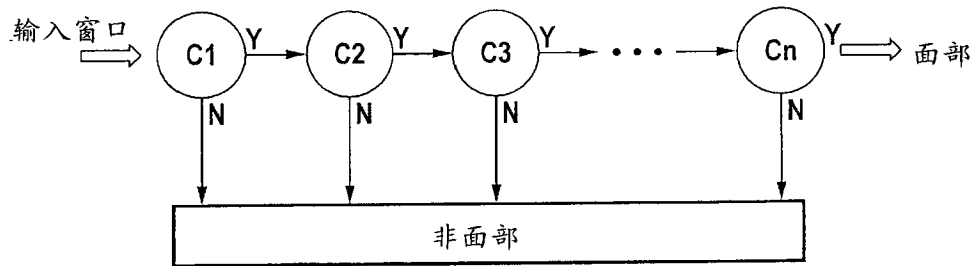


图 7A(现有技术)

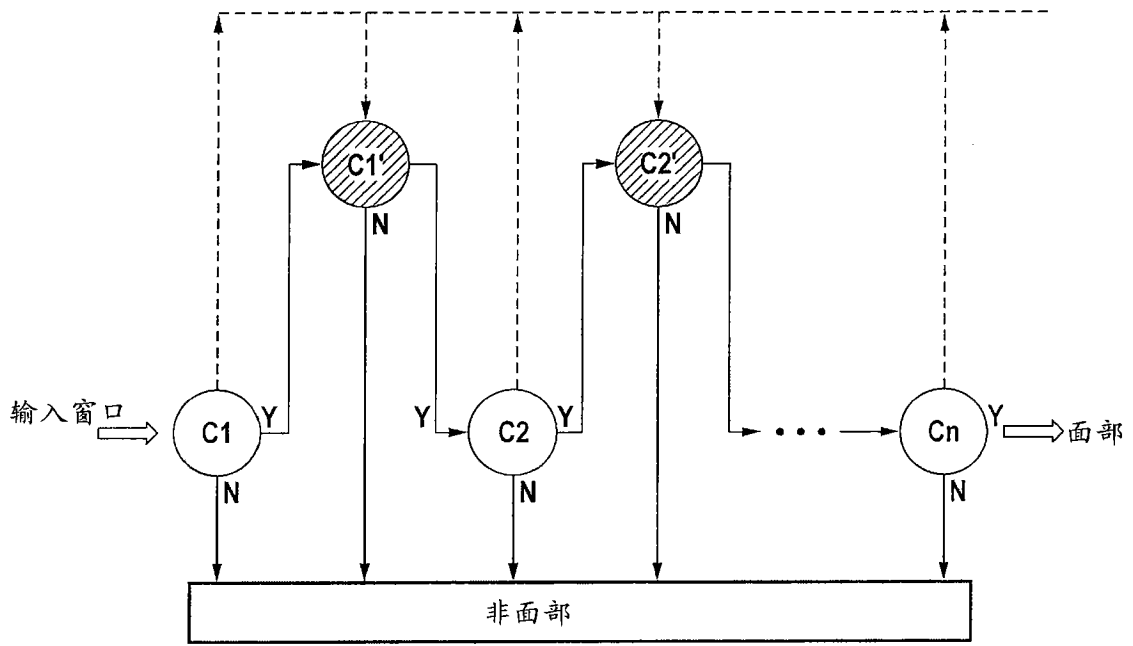


图 7B

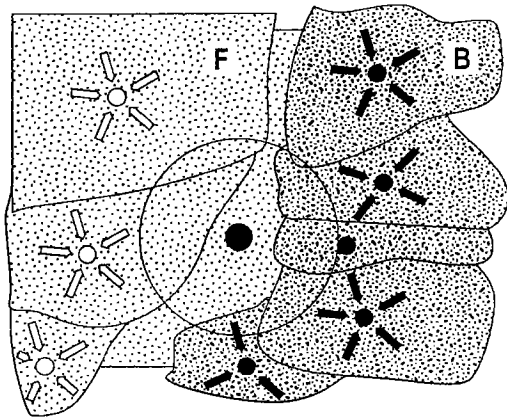


图 8

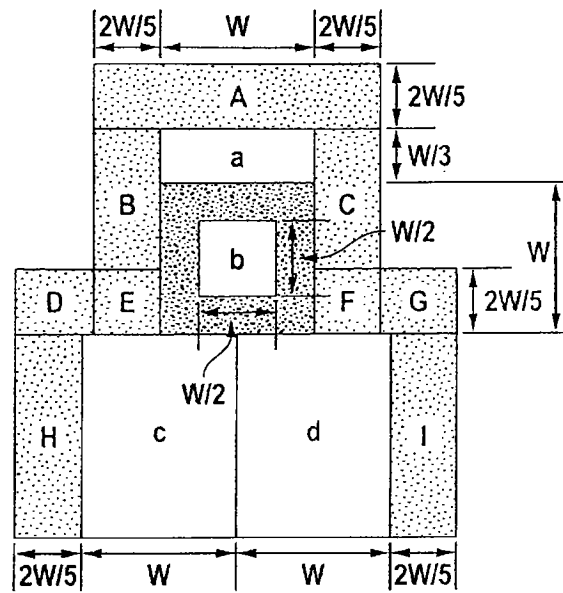


图 9

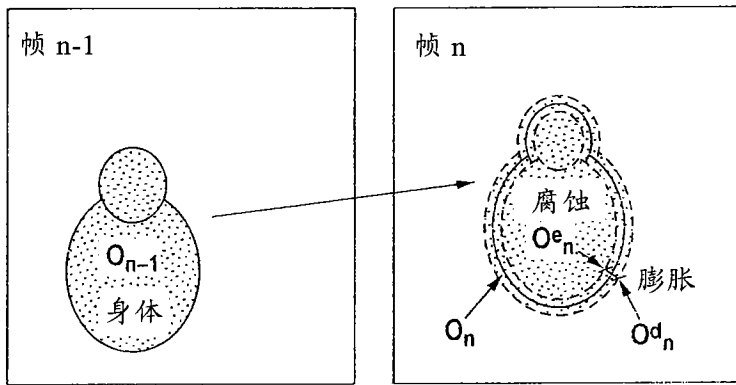


图 10

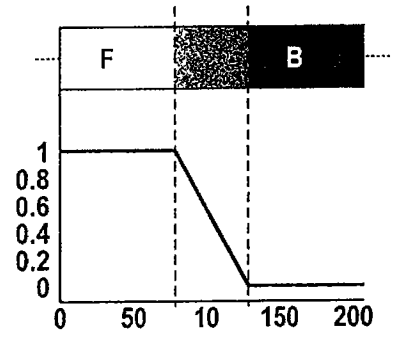


图 11A

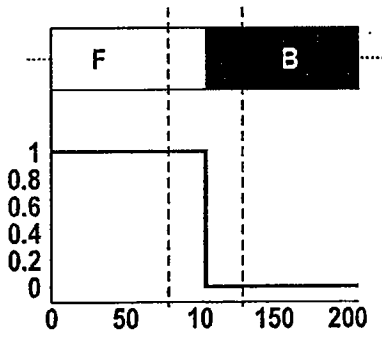


图 11B

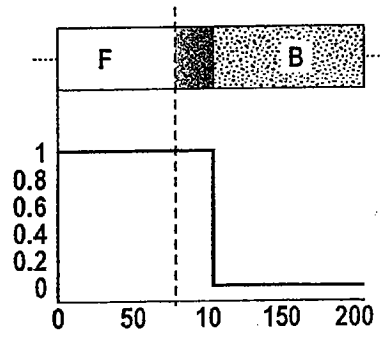


图 11C

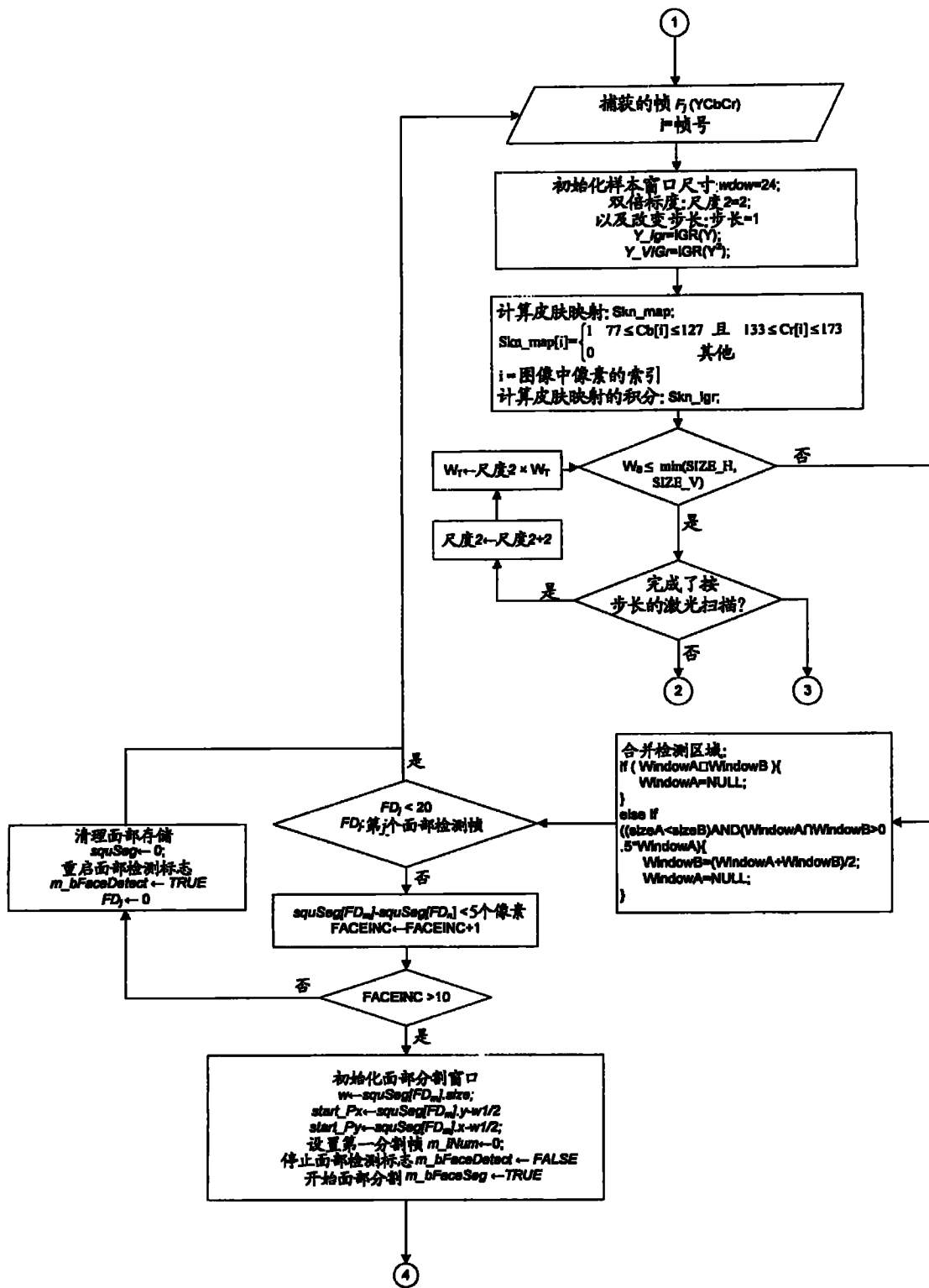


图 12A

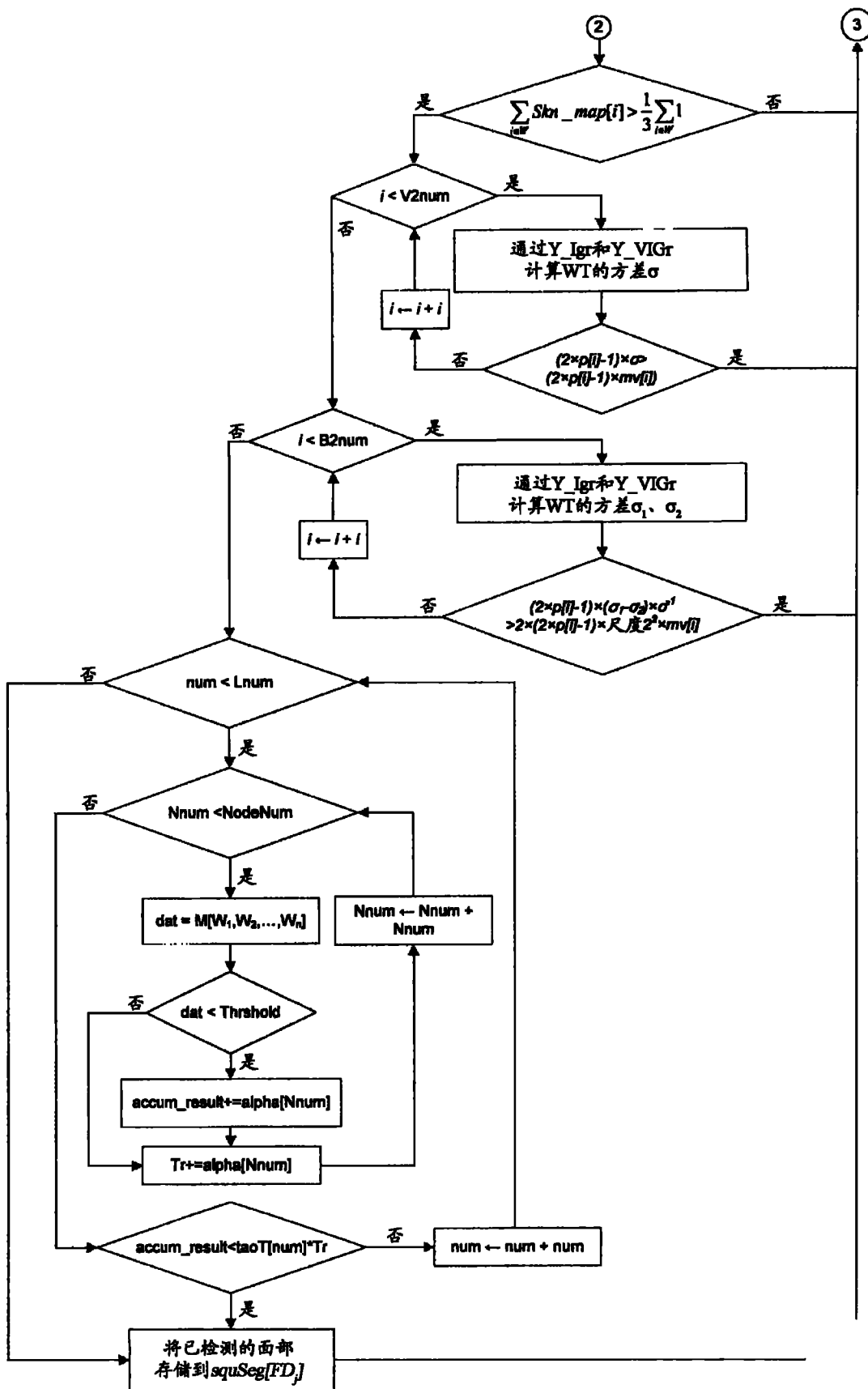


图 12B



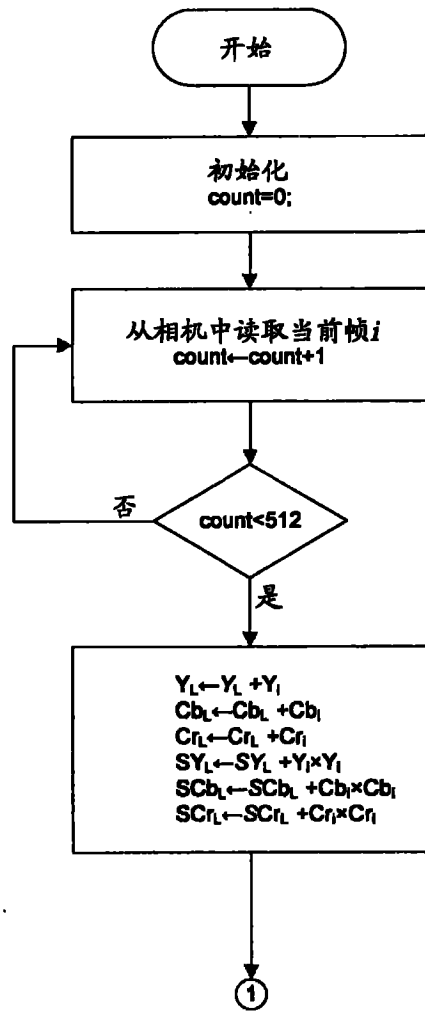


图 13A

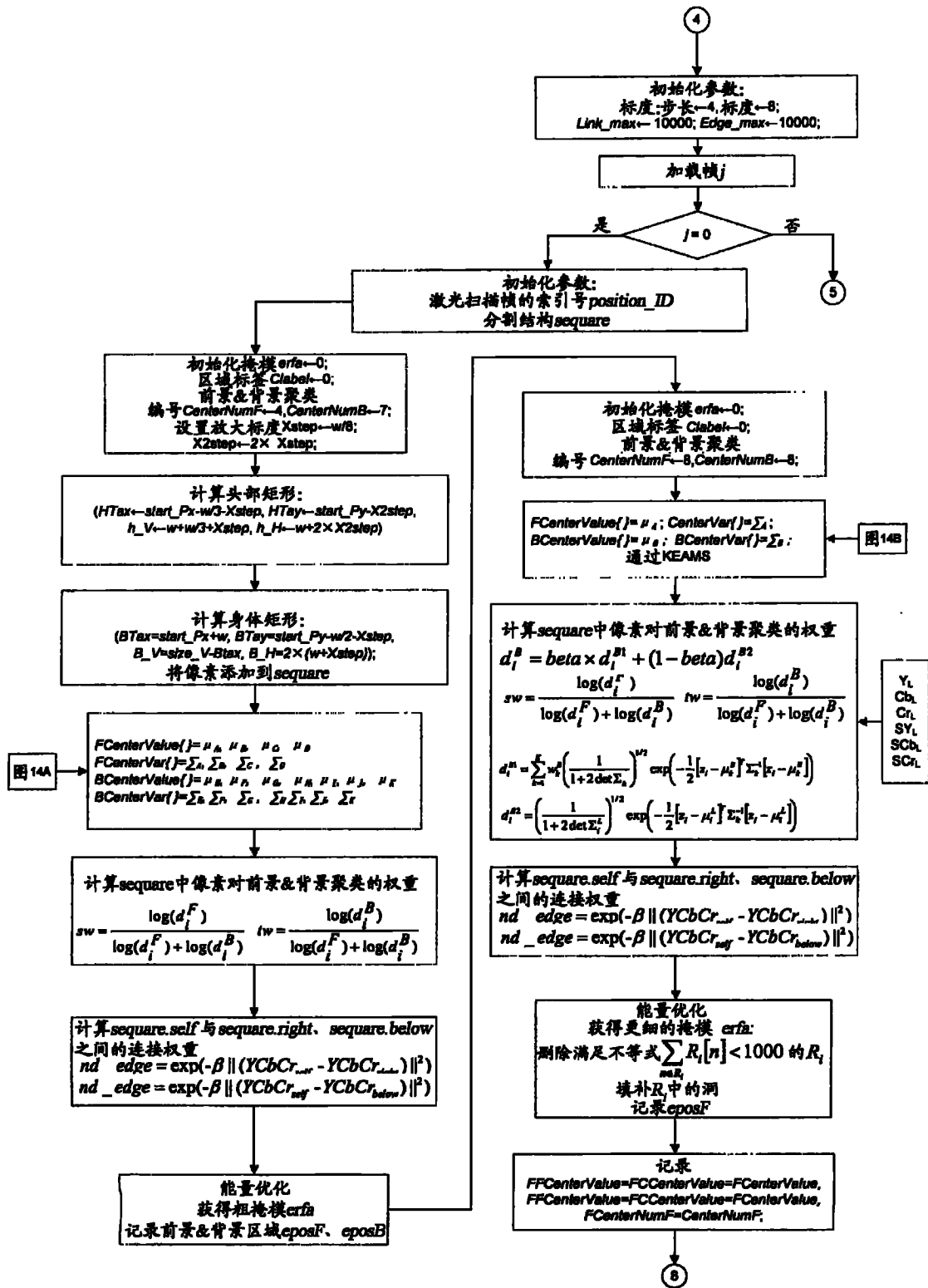


图 13B

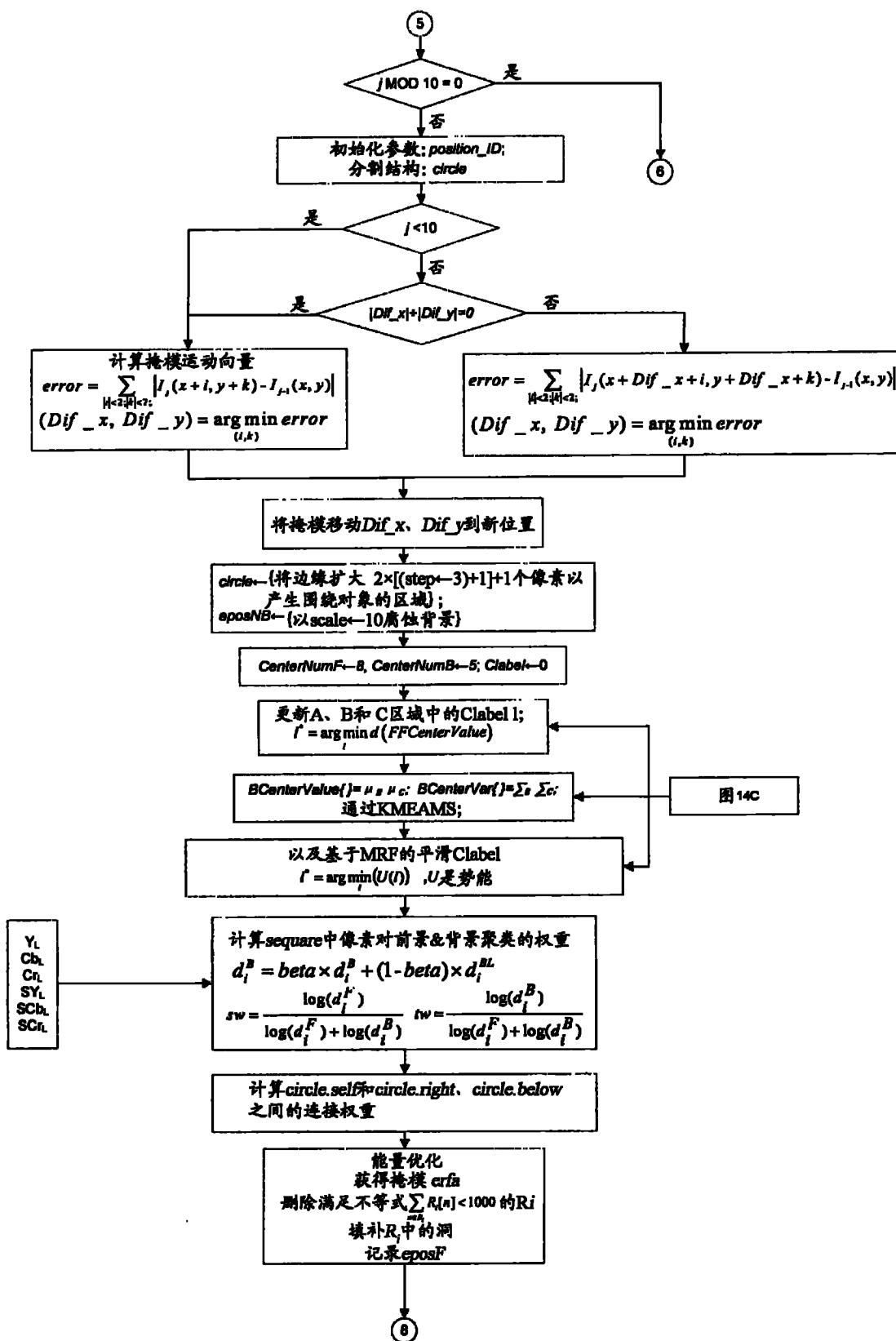


图 13C

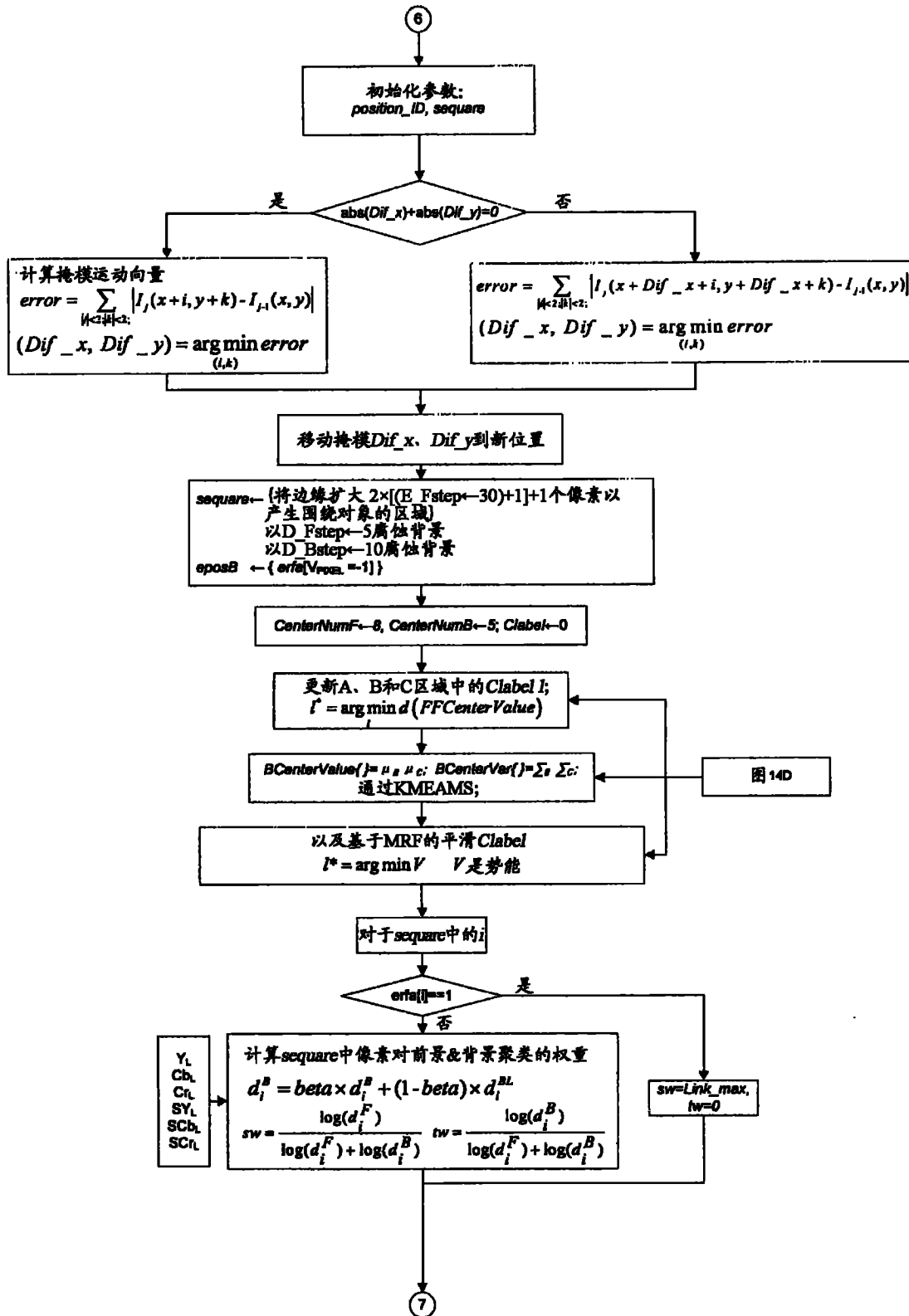


图 13D

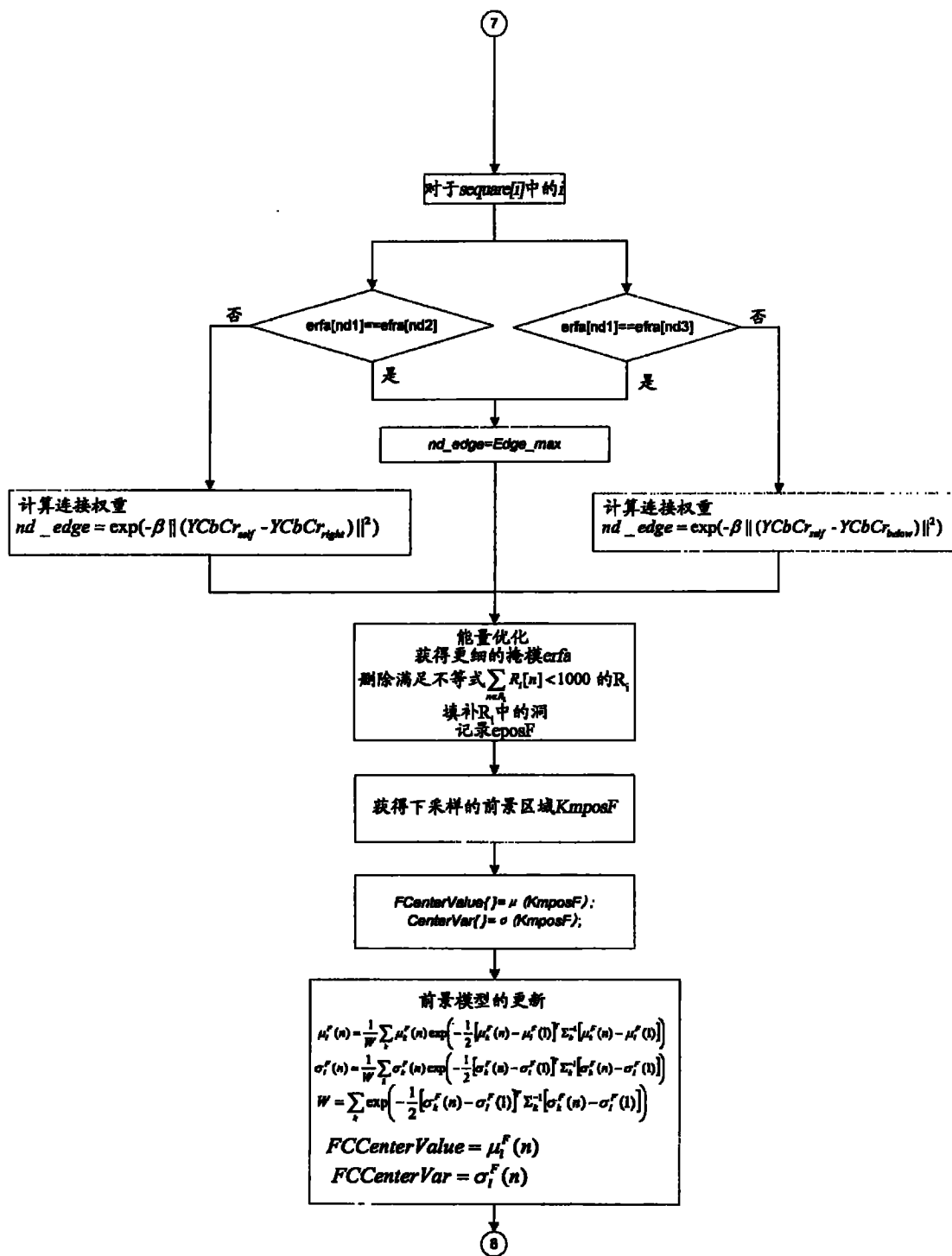


图 13E

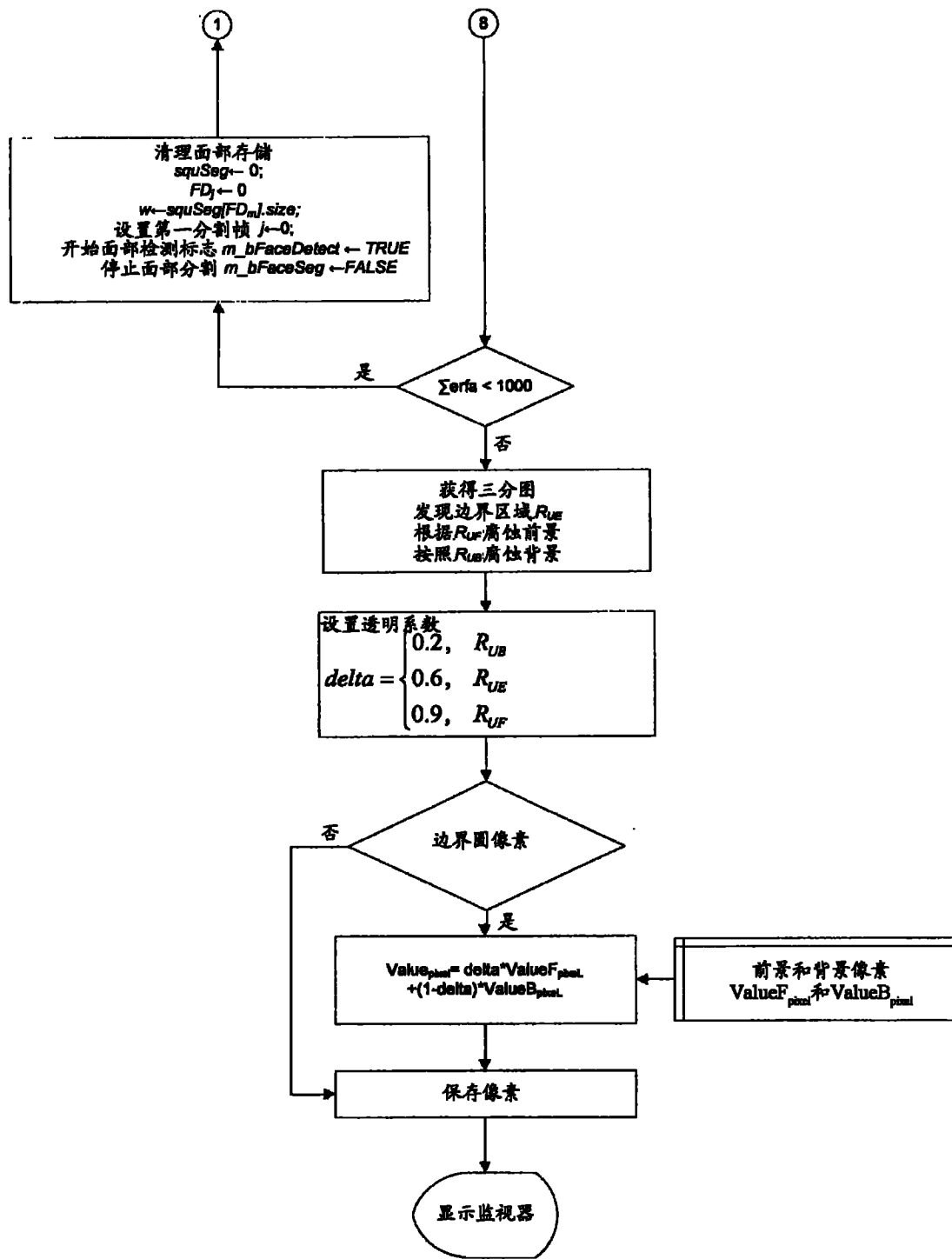


图 13F

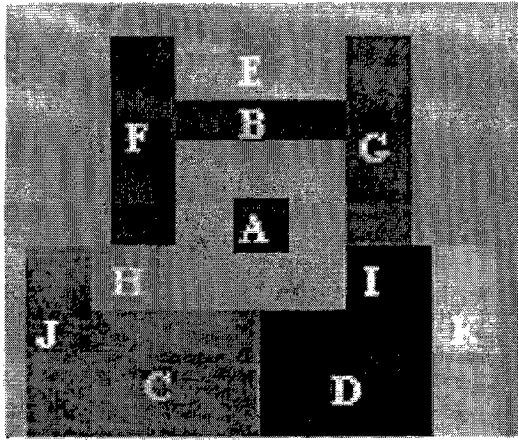


图 14A

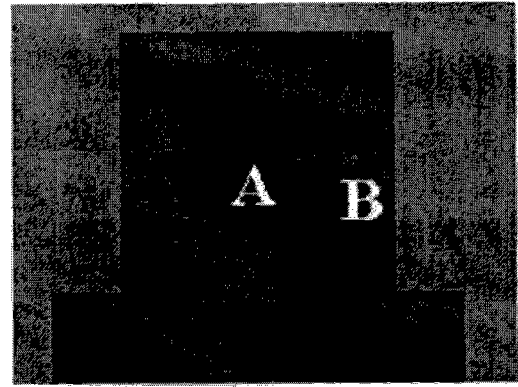


图 14B

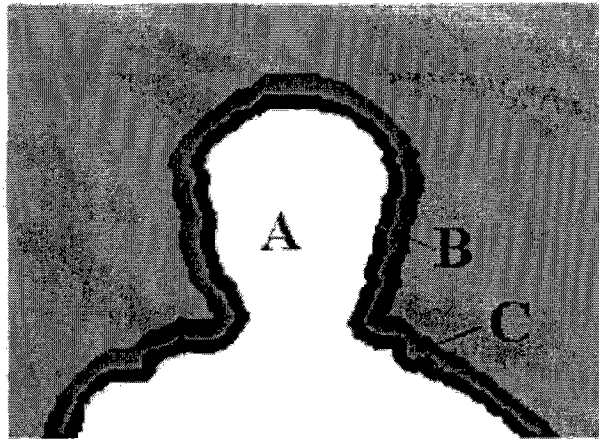


图 14C

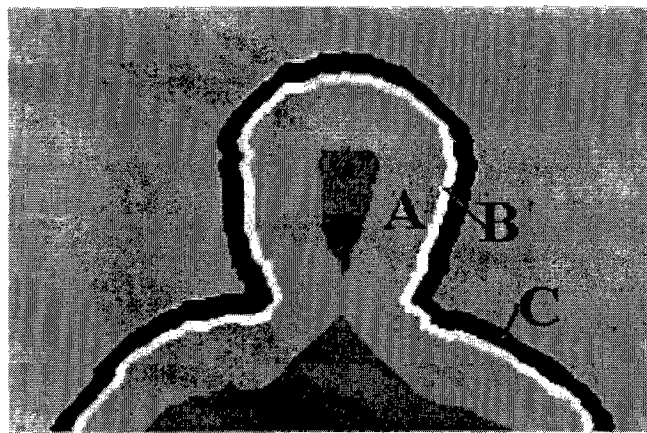


图 14D