视频相册系统

朱才志[1)] 吴秀清[1)] 周 晓[2)] 张 垒[1)]

[1)](中国科学技术大学电子工程与信息科学系,合肥 230027) [2)](合肥工业大学计算机科学与技术系,合肥 230009)

摘 要 为了对视频数据进行有效的管理,提出了一种新的视频检索与浏览系统——视频相册系统。在该系统中,首先用相册生成方案挑选出用户数字视频库的一组代表性的关键帧;接着筛选出的关键帧被预训练的形状模板(如圆形、心形、扇形、邮票形等)所裁剪,最终被打印成册。当用户想浏览视频时,可事先浏览该视频相册,就像浏览普通相册一样,若用户想观看相册中某个关键帧所代表的视频片段,即可首先方便地用摄像手机等设备拍摄该关键帧,并通过无线网络(如蓝牙)把拍摄的图像传输给计算机终端;此后,视频相册系统采用基于自训练与 Snakes 轮廓进化的活动形状模型算法来定位关键帧在拍摄的图像中的轮廓位置,并纠正其成像畸变。最终,系统即可在视频数据库中自动找到与纠正后的关键帧最相似的一幅,并为用户回放其代表的视频片段。实验评测结果表明,该视频相册系统可在数字视频与模拟相册间建立有效的联系。

关键词 视频相册 视频检索 视频概要 自训练 Snakes 模型 活动形状模型

中图法分类号:TP391.3 **文献标识码**:A **文章编号**:1006-8961(2008)08-1544-10

Video Booklet System

ZHU Cai-zhi[1)], WU Xiu-qin[1)], ZHOU Xiao[2)], ZHANG Lei[1)]

[1)] (*Department of Electronic Engineering and Information Science, University of Science and Technology of China, Hefei* 230027)

[2)] (*Department of Computer Science and Technology, Hefei University of Technology, Hefei* 230009)

Abstract This paper proposes a novel system, named Video Booklet, which enables efficient and natural personal video browsing and searching. In the system, firstly representative key-frames of the video segment collection are selected through an elaborate booklet generation approach, and then reshaped by a set of pre-trained personalized shape templates (such as circle, heart, sector, stamp, etc), consequently printed out on a real booklet or album. When users plan to browse the content of their digital video library, they can firstly browse their booklets in a manner as browsing ordinary photo albums. When they want to watch a certain segment indicated by a key-frame in the booklet, they are able to use their camera phones or similar devices to capture the corresponding frame, and send the captured image to a computer via wireless network (such as blue tooth). Thereafter, the target frame is accurately located by a proposed self-training and Snakes evolution based active shape model algorithm, and the distortion of the captured image is corrected. Finally the Video Booklet system will automatically find the most similar key-frame to the corrected one in the video library and begin to play the corresponding segment for the users. Thereby, Video Booklet builds a seamless bridge between digital videos and analog albums.

Keywords video booklet, video retrieval, video summarization, self-training, Snakes model, active shape models

1 引 言

随着数码摄像机的流行,多媒体数据,特别是家庭视频近年来增长迅速。众所周知,与文本数据不同,多媒体数据难以检索与浏览,对普通用户而言,通常检索与浏览他们的个人媒体数据不仅非常耗时,且极为不便。尽管如今介绍基于媒体内容的检

收稿日期:2007-01-05;**改回日期**:2007-03-08

第一作者简介:朱才志(1979~),男。2003 年 7 月获武汉大学工学硕士学位。现为中国科学技术大学信号与信息处理专业博士研究生。主要研究方向为图像处理、计算机视觉与视频分析。E-mail:zhucz@ustc.edu,caizhi.zhu@gmail.com

索与浏览系统的文献屡见不鲜[1,2],但大多数研究均集中于视频分割、视频摘要与视频检索,而用户界面则几乎都基于 PC,这意味着,用户需要用键盘、鼠标或其他遥控设备在计算机上检索、浏览想要的媒体内容。

本文提出一种有效的、自然的视频浏览与检索系统——视频相册系统[3~5](如图 1 所示)。

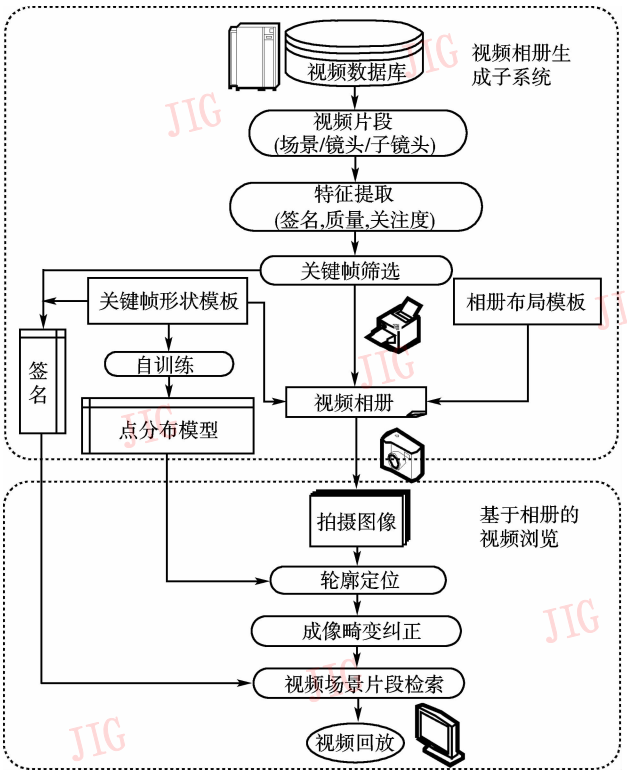


图 1　视频相册系统流程

Fig. 1　Flow chart of video booklet

该系统包括视频相册生成子系统与基于相册的视频浏览子系统两个部分。在视频相册生成子系统中,用户的所有视频数据首先被分割成场景(scene)、镜头(shot)与子镜头(sub-shot)等片段[3],如果从中抽取一组代表性的关键帧,并提取其签名特征[6];此后,再基于选定的关键帧、用户选择的相册布局模板及关键帧形状模板,把关键帧印制成册,那么就生成了模拟的视频相册。在基于相册的视频浏览子系统中,当用户想要浏览其视频数据库时,可事先浏览视频相册,就像浏览普通相册一样。若用户想观看相册中某个关键帧所代表的视频片段,则可方便地用摄像手机或其他相机设备拍摄该关键帧,并通过无线网络(如蓝牙)把拍摄的图像传输给计算机终端。此后,视频相册系统即可在视频数据库中自动找到对应的视频片段,并为用户回放。这样,视频相册系统就在数字视频与模拟相册间建立了无缝链接的桥梁。

视频相册系统的技术难点在于:(1)如何产生相册,包括如何选择恰当的关键帧集合、如何在相册中规划关键帧的布局与外观,以便使得既便于视频检索,又能彰显用户的个性化风格;(2)如何准确定位出关键帧在拍摄的图像中的轮廓位置。由于关键帧在被形状模板裁剪后,外形各异,且常被印制在有背景纹理的图纸中,同时由于拍摄的图像往往质量不高,且存在严重的成像畸变,从而使得轮廓的定位问题非常棘手。针对前者,本文采用视频内容分析算法[3]先从视频库中挑选出最优的关键帧集,以便于视频检索;然后引入相册布局模板与关键帧形状模板,以体现灵活的、吸引人的个性化风格。对于后者,本文采用一种新的基于自训练与 Snakes[7]轮廓进化的活动形状模型(active shape models, ASM)[8]算法。这种自训练方案可自动提取关键帧形状模板的点分布模型(point distribution model, PDM),用以监督 ASM 的轮廓匹配;而基于 Snakes 的轮廓进化策略则可提高 ASM 的鲁棒性与加快其收敛速度,并可最终精确定位关键帧的轮廓。

2　视频相册生成方案

本节主要介绍基于关键帧的筛选方案、相册布局模板及关键帧形状模板,以及系统如何根据数字视频集来生成模拟的视频相册。其中关键帧筛选方案根据视频数据集选择一组最优的关键帧;相册布局模板用于决定相册的整体外观与关键帧布局;而关键帧形状模板则将选定的关键帧裁剪成用户想要的各种形状。

2.1　关键帧的选取

关键帧选取时,首先通过分析视频库的视频时序结构[3]把视频数据分割成 3 层片段,即由小到大分别为子镜头、镜头与场景;接着在每个场景的多个子镜头中选择其一,提取其关键帧,并打印在视频相册上,用于代表该场景片段。

关键帧的选取原则为:(1)最大化关键帧间的签名差异[6],以便降低视频检索的出错几率,关键帧签名的计算过程如下:首先把关键帧划分成 $N = N_x \times N_y$ 个子块(例:$N_x = N_y = 5$);然后计算每个子块的平均灰度级;最后升序排列灰度级,并

把该 N 维的排序数作为关键帧的签名，它是反映图像内在相对灰度分布的一种鲁棒的计序测度，这里签名差异是指对两个关键帧签名的对应位数值间的绝对差求和后的结果；(2)最大化关键帧的画面质量[9]，关键帧的画面质量测度是一种主观性度量，基本上，利用关键帧的对比度、灰度直方图与彩色直方图等多个视觉特征就可以大致描述关键帧的视觉质量；(3)最大化关键帧的代表性，亦即最大化关键帧的关注度(attention)[10]，这里代表性是指用该子镜头的关键帧来代表整个场景时的典型性，亦即子镜头在整个场景中的地位，其也是一种主观度量，可由关键帧所表征的子镜头中的目标运动情况、相机运动情况、颜色与声音信息等多个因素综合描述。

最终关键帧的选择问题可用如下式子表示：

$$\underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}}{\arg\max} F(\boldsymbol{\theta}) = \alpha SD_{\boldsymbol{\theta}} + \beta VQ_{\boldsymbol{\theta}} + \gamma RP_{\boldsymbol{\theta}} \quad (1)$$

式中，SD，VQ，与 RP 分别表示关键帧的签名差异(signature difference)、画面质量(video quality)与代表性度量(representative)，$\boldsymbol{\theta}$ 与 $\boldsymbol{\Theta}$ 分别表示选取的关键帧子集与整个关键帧集合，α，β，$\gamma \geqslant 0$，$\alpha + \beta + \gamma = 1$。该动态规划问题可用遗传算法解出[9]。

2.2 相册布局模板

为使视频相册能像普通相册一样支持个性化的布局风格，在视频相册的产生过程中用户可自由设计相册布局模板，图 2 为 3 种相册布局的样例模板，模板采用 XML 描述。有了相册布局模板，用户就可自由定制视频相册中每页的关键帧的布局。与关键帧形状模板不同的是，相册布局模板仅仅影响相册的外观，而在技术层面上对系统并无影响。

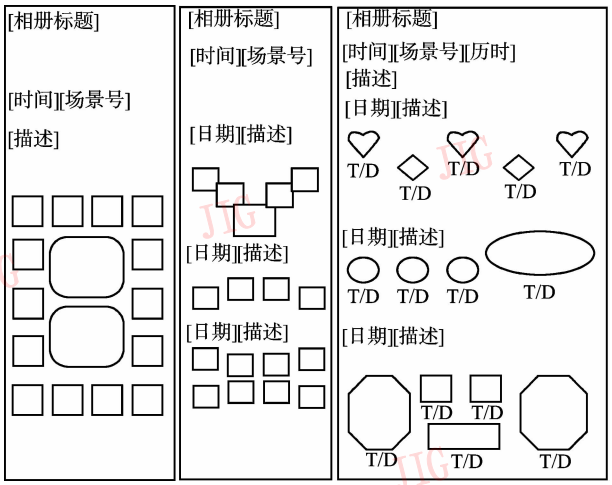


图 2 相册布局模板样图(T/D：时间/历时)

Fig. 2 Sample Booklet layout templates (T/D: time/duration)

2.3 关键帧形状模板

如前所述，为使生成的视频相册在界面上更加友善，该视频相册系统效仿普通相册支持关键帧形状模板的自由设计，使最终显现在相册上的是被形状模板裁剪后的关键帧，图 3 显示了分别被矩形(包括拍摄的图像的初始形状本身)、扇形、椭圆形、心形与邮票形模板所裁剪后的关键帧。该系统同时支持用户自定义的形状模板，如果要添加用户自己的形状模板，则只需用普通画图工具(如 mspaint、photoshop)划一条闭合的形状轮廓线，并将其保存在形状模板库中即可。为使关键帧取得较好的自动裁剪效果，该系统将裁剪区域定为关键帧的关注点(attention view)[11]，亦即关键帧中最引人注目的区域。同时，对裁剪后的每幅关键帧，系统会自动抽取其签名特征[6]，并将其存储于签名数据库中。

图 3 关键帧形状模板样图

Fig. 3 Sample key frame shape templates

2.4 ASM 模型的自训练方案

本小节将详细介绍 ASM 的自训练方法。如前所述，本文采用 ASM 算法来定位关键帧在拍摄的图像中的轮廓。由于经典的 ASM 算法需要用户预先手工标注大量的训练样本，以提取 PDM，用以指导 ASM 的搜索过程，因此为算法在

系统中的应用带来了不便。本文提出的自训练方案不但可自动训练形状模板,并能彻底替代复杂的手工标注过程。若给定一形状模板,则自训练方法就能自动完成所有的训练过程,包括:自动提取标记点、模拟产生训练样本、训练集对齐以及经主成分分析(principal component analysis, PCA)后获得的 PDM 等。本小节以心形模板为例来阐述整个训练过程,其他形状模板的训练亦与此同。

2.4.1　自动提取标记点

给定拟提取的标记点数目,即可用 DP 算法[12]挑选出形状模板的多边形近似点。实验中,对心形模板选取 100 个标记点即可提供足够的轮廓近似精度。提取出标记点后,形状轮廓可用 $2n$ 维向量 $\boldsymbol{X}_0$ 描述($n=100$):

$$\boldsymbol{X}_0 = (x_{0,1},\cdots,x_{0,n},y_{0,1},\cdots,y_{0,n})^{\mathrm{T}} \tag{2}$$

图 4(a)显示了心形模板中自动提取的标记点(用红色标注)。

(a)自动提取的标记点　(b)模拟产生的部分训练样本(在水平旋转方向上)　(c)对齐后的训练样本

图 4　自训练过程示例

Fig. 4　Self-training process

2.4.2　模拟产生训练样本

在 ASM 算法中,一般需要用手工标注大量的样本来训练得到 PDM,以指导 ASM 的轮廓匹配过程。具体到本类情况——关键帧在同一平面,且已知其正面的 2 维形状模板,若要提取其在 3 维空间的成像轮廓,则问题可得到简化,即根据已知的 2 维形状模板,利用透视投影变换就可自动生成训练集,用来模拟关键帧在 3 维空间成像所产生的各种形变(为使 PDM 监督产生的形状与实际拍摄的形状轮廓尽可能匹配,训练集理应囊括由于成像角度与方位所造成的轮廓模型的各种成像畸变)。通常关键帧图像多为用普通焦距相机近焦拍摄所得,由于此时桶形与枕形失真皆可忽略不计,因此关键帧的成像畸变可用孔径模型来模拟,并可看作是由关键帧形状模板在虚拟的针孔相机模型中的先后水平旋转(pan)与竖直倾斜(tilt)后成像所致(如图 5 所示)。

关键帧在成像过程中存在以下两个特征:

(1)通常由于关键帧在拍摄的图像中居中,因此可近似认为相机的光轴穿过图像的中心;

(2)由于拍摄后的关键帧在水平或竖直方向上一般覆盖图像的大部分区域,从而相机视场(field of view,FOV)可近似计算如下:

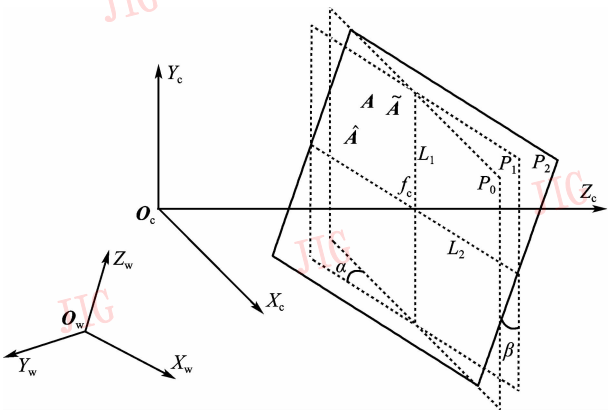


图 5　虚拟孔径模型

Fig. 5　The virtual pinhole model

$$FOV_i = 2\times\arctan(F_i/(2\times f)) \approx 2\times\arctan(B_i/(2\times f)) \tag{3}$$

这里 F,f 与 B 分别代表相机底片长度、镜头焦距与关键帧成像后的物理尺寸。i 代表选择的视场方向:

$$i = \begin{cases} \mathrm{h}, B_{\mathrm{h}}/F_{\mathrm{h}} > B_{\mathrm{v}}/F_{\mathrm{v}} \\ \mathrm{v}, B_{\mathrm{h}}/F_{\mathrm{h}} < B_{\mathrm{v}}/F_{\mathrm{v}} \end{cases} \tag{4}$$

式中,h、v,分别表水平与竖直。由于式(3)中 FOV_i 为已知值,通常在一定范围内,如普通摄像头的视场

一般大于 40°,小于 60°。

由此可在虚拟孔径模型中模拟以上两个特征,即将形状模板垂直置于孔径模型的光轴上,并将模板的中心与虚拟孔径模型的焦点对齐(参见图 5 中平面 P_0)。由式(3)即可推出

$$f_c = M_i/(2 \times \tan(FOV_i/2)) \tag{5}$$

式中,M 表示形状模板像素数,f_c(下角 c 代表 Camera)表示相机虚拟焦距所占像素数,注意式(5)是在式(3)的基础上进行了单位换算。

以上步骤确保了在旋转与倾斜之前,形状模板在孔径模型中的投影与实际模板形状的一致。由于这里只需考虑模板的形状特征,而无需顾及姿态(包括坐标、方向角度与尺度大小等)信息,从而不必考虑世界坐标到相机坐标的变换。

此后,以摄像机焦点为中心,再以一定角度步长来分别旋转和倾斜形状模板,即可获得形状模板在拍摄中投影的模拟。现以模板中的任意标记点 $\boldsymbol{A}=(x,y,z)$ 为例(注意 $z=f_c$)来进行说明,为方便以下坐标变换运算,可首先把原点 $\boldsymbol{O}_c$ 平移至焦距 f_c 处(即在 z 分量上加上偏移量 $-f_c$);接着绕纵轴 L_1 以角度 α 将平面 P_0 水平旋转到平面 P_1,此时点 A 旋转至点 $\hat{\boldsymbol{A}}=(\hat{x},\hat{y},\hat{z})$(图 5),易得

$$\begin{cases}\hat{x} = x\cos\alpha \\ \hat{y} = y \\ \hat{z} = x\sin\alpha\end{cases} \tag{6}$$

随之以角 $\beta°$ 绕横轴 L_2 将平面 P_1 竖直倾斜至平面 P_2,可得 $\hat{\boldsymbol{A}}$ 在 P_2 中的对应位置 $\tilde{\boldsymbol{A}}=(\tilde{x},\tilde{y},\tilde{z})$(这里当步长 $\Delta\alpha$ 与 $\Delta\beta$ 取值较小时,水平旋转与竖直倾斜的先后顺序对最终的训练结果的影响可忽略不计),有

$$\begin{cases}\tilde{x} = \hat{x} \\ \tilde{y} = \hat{y}\cos\beta + \hat{z}\sin\beta \\ \tilde{z} = \hat{z}\cos\beta - \hat{y}\sin\beta\end{cases} \tag{7}$$

将式(6)代入式(7),并加上 f_c 至 $\tilde{z}$ 分量,以恢复相机坐标

$$\begin{cases}\tilde{x} = x\cos\alpha \\ \tilde{y} = x\sin\alpha\sin\beta + y\cos\beta \\ \tilde{z} = x\sin\alpha\cos\beta - y\sin\beta + f_c\end{cases} \tag{8}$$

根据相似三角形理论即可得到点 $\tilde{\boldsymbol{A}}$ 在图像坐标系中的投影坐标(x_i, y_i):

$$\begin{cases}x_i = \tilde{x} \cdot f_c/\tilde{z} \\ y_i = \tilde{y} \cdot f_c/\tilde{z}\end{cases} \tag{9}$$

在忽略常系数 f_c(由于它只与姿态的尺度因子有关,对形状无影响)后,再将式(5)与式(8)代入式(9),即可导出以下投影公式:

$$\begin{cases}x_i = x\cos\alpha/(x\sin\alpha\cos\beta - y\sin\beta + M_i/(2 \times \tan(FOV_{p_i}/2))) \\ y_i = (x\sin\alpha\sin\beta + y\cos\beta)/(x\sin\alpha\cos\beta - y\sin\beta + M_{p_i}/(2 \times \tan(FOV_{p_i}/2)))\end{cases} \tag{10}$$

最终,对任意的标记点 $\boldsymbol{p}_i=(x_i,y_i)$($i \in (0,n]$),通过式(10)即可获得其投影坐标$(x_i, y_i)$。如果用式(10)求出形状模板中的所有标记点的投影坐标,即可模拟得到形状模板在相机前,以角度 α 进行水平旋转、以角度 β 进行竖直倾斜后的投影。当旋转与偏移过程分别以较小的步长 $\Delta\alpha$ 与 $\Delta\beta$ 遍历某个空间 $\boldsymbol{\Omega}$ 时,这里

$$\boldsymbol{\Omega} = \{(\alpha,\beta) \mid -\theta_1^p \leqslant \alpha \leqslant \theta_2^p,\ -\theta_1^t \leqslant \beta \leqslant \theta_2^t\} \tag{11}$$

则可最终得到一个模拟的训练集。为方便观察,本文在图 4(b)中显示了心形模板在模拟水平旋转时生成的部分训练样本,此时旋转角度 $\theta_1^p = \theta_2^p = \pi/6$(上角 p 代表 pan),而倾斜(tilt)角度 $\theta_1^t = \theta_2^t = 0°$(当二者皆取 $\pi/6$ 时,训练样本个数是图 4(b)训练样本数的平方,过于稠密则不便于观察)。这里用形状模板中的同一标记点所模拟生成的所有投影点都被标记为同种颜色。实验参数设置如下:

$$\begin{cases}\Delta\alpha = \Delta\beta = 1/\pi \\ \theta_1^p = \theta_2^p = \theta_1^t = \theta_2^t = \pi/4 \\ FOV_h = \pi/3 \\ FOV_v = \pi/4\end{cases} \tag{12}$$

实验证明,在保证训练样本具有一定数目的同时,自训练性能对以上参数的选择并不敏感。

2.4.3 训练集对齐

实验采用了文献[8]中训练样本对齐算法,以迭代方式实现。主要区别在于,为使对齐后的 PDM 不仅更加紧凑,同时可使对齐后的非线性最小,以便在切空间(tangent space)[13, 14]中实现形状对齐。实验中,形状对齐的迭代步骤不超过 5 次,图 4(c)显示了对图 4(b)训练集对齐后的结果。由图 4 不难

发现,对齐后对应的标记点(以同种颜色表示)比之对齐前更接近于在同一直线上,亦即训练集的对齐减小了 PDM 的非线性。

2.4.4　主成分分析

每个对齐后的训练样本对应于 $2n$ 维(n 代表标记点的个数)切空间中的一点,由 N 个样本组成的训练集组成了 $2n$ 维的合法形状域(ASD)[15]。ASD 呈超椭球体状,该超椭球的轴线方向与长度分别用特征向量 $\boldsymbol{p}_i$ 与特征值 λ_i($1 \leqslant i \leqslant 2n$)描述($\lambda_i$ 为主轴上样本分布的方差)。主成分分析可用于计算 ASD 超椭球的主轴,前 t($t \ll 2n$)个主轴表示训练样本在切空间中最主要的变化模式[8]。实验中可根据 $\sum_{i=1}^{t}\lambda_i = \sum_{i=1}^{2n}\lambda_i \times 98\%$ 来选择 t(这里 $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_{2n}$),意即前 t 个主轴覆盖训练样本在切空间中 98% 的变化模式。一般仅 4 个主轴就足以覆盖 $2n$ 维($n=100$)ASD 的 98% 的变化模式,这亦证明对齐后的训练集的 PDM 非常紧凑。即 PDM 可由 $\bar{\boldsymbol{X}}$,$\boldsymbol{p}_i$ 与 λ_i($1 \leqslant i \leqslant t$)共同描述,$\bar{\boldsymbol{X}}$ 表示训练集的平均形状。

图 6 显示了心形模板在主轴上的形状变化情况,图中每行居中形状表示平均形状,余者从左至右依次为在相应主轴上从 $-3\sqrt{\lambda_i}$ 到 $3\sqrt{\lambda_i}$ 区间内等间距取值所得。黑点代表标记点,图中标记点被浅灰色线段连接成形。PCA 分解后,即可认为形变参数 b(即样本点在各主成分上的分量)互不相关,样本在各主成分上皆服从高斯分布。根据正态分布规律,大约 99.87% 的样本分布在主轴上从 $-3\sqrt{\lambda_i}$ 到 $3\sqrt{\lambda_i}$ 的区间内。

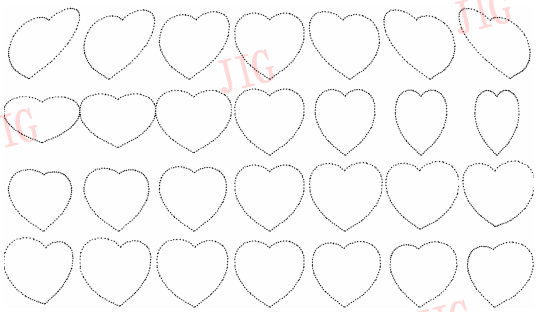


图 6　4 个主成分重建后的形状

Fig. 6　Shapes reconstructed by the first four PCs

对图 6 中形状变化的直观描述是:第 1 行显示了第 1 主成分表示模板在侧倾(混合 pan 与 tilt)后的形变;相应地,第 2、3 主成分分别模拟了模板的垂直倾斜与水平旋转;第 4 主成分则表示更细微的形状变化。

2.4.5　自训练方案的优点与改进方向

由此可归纳自训练方案具有以下主要优点:(1)训练过程完全自动化;(2)可避免手工标注样本所引入的人为误差(主要由于在海量样本的人工标注中,难以确保样本标记点的对应);(3)可按需产生训练样本数,同时训练过程简单高效(以心形为例,在 P4 2.8G PC 上的整个训练时间 ≤ 1s)。

必须指出,由拍摄角度的变化所引起的形变是非线性的,尽管变换到切空间后,这种非线性只能得到有效抑制,但依然存在。由图 7 可见,模拟产生的心形样本的两个主要形变参数 b_1 与 b_2 并非是假定的互不相关,而且从图 4(c)中同色标注的对应标记点的连线并非绝对“直”,亦可直观看出非线性的存在。如果进一步减少由于视角改变所引起的非线性,则一个可行的解决方案[15]是采用基于形变参数的概率密度函数 $p(b)$ 的非线性模型,这里 $p(b)$ 的计算是采用核密度估计的方法。

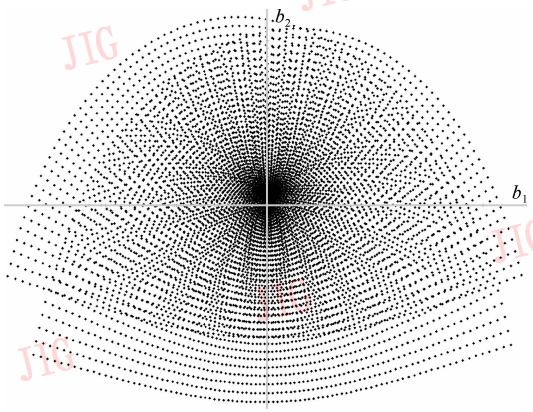


图 7　模拟的心形样本在形变参数 b_1、b_2 坐标系中的分布

Fig. 7　Distribution of simulated heart shape samples in the coordinates of shape parameters b_1 and b_2

3　基于相册的视频浏览

本节阐述基于上述视频相册系统如何实现用户对视频数据的检索与浏览。该子系统的工作流程如图 1 所示,其技术方案如下:首先定位关键帧在拍摄的图像中的轮廓边界;接着通过透视变换纠正关键帧的成像畸变,以恢复关键帧及其模板形状;最后提取纠正后的关键帧的签名,同时在签名数据库中寻

找其最佳匹配,并为用户回放由匹配帧所表征的视频片段。

3.1 基于 ASM 的关键帧轮廓提取

“自下至上”数据驱动的目标提取方法一般用于检测图像中的局部结构,如边界、相似区域等信息,并将其组合后作为判别感兴趣目标的依据。由于没有先验模型的指导,致使上述方法往往普适性差,很不稳定。本文得益于训练所得的 PDM 知识的指导,采用了“自顶向下”模型驱动的 ASM 算法,因为其在噪声图像中更加鲁棒。本节首先以图 8 为例介绍用 ASM 算法定位关键帧轮廓的步骤,再阐述在此基础上改进的基于 Snakes 的 ASM 轮廓进化策略,同时提出可把 ASM 算法的中间结果直接用于关键帧的模板判定。经典 ASM 算法的详细介绍请参见文献[8]。

3.1.1 基于 ASM 算法的关键帧轮廓的定位

根据自训练得到的 PDM,ASM 算法就可采用迭代的轮廓进化策略来定位当前形状模板在拍摄图像上的最佳轮廓的匹配位置。本小节以心形模板裁剪后的关键帧的轮廓定位为例来简要阐述如下 ASM 算法的步骤:

(1)估算关键帧形状模板在拍摄的图像中的位置(见图 8(a))。由于关键帧一般在拍摄的图像的中央位置,因此可用一般的图像处理技术(如 Canny、种子填充等)粗略估算出关键帧的大致所在。此后可将 PDM 中的平均形状 $\overline{\boldsymbol{X}}$ 与之对齐,并将其作为关键帧轮廓 $\boldsymbol{X}$ 的初始估计(如彩图 8(a)所示)。实验发现,最终的定位结果对该初始位置并不敏感。

(2)对 $\boldsymbol{X}$ 中的每个标记点,搜索其在下次迭代中的最佳位置。搜索时,以标记点为中心,沿着标记点组成的轮廓的法线方向,在一定搜索范围内(例如两边各 20pixels 宽),通过搜索在该方向上梯度分量最大的点来作为下次迭代的最佳点。图 8(b)中连接蓝色标记点的绿色多边形轮廓为搜索前的形状,而镶嵌了白色标记点的红色多边形则表示第 1 次迭代后的最佳点组成的轮廓,搜索范围以黄线表示。

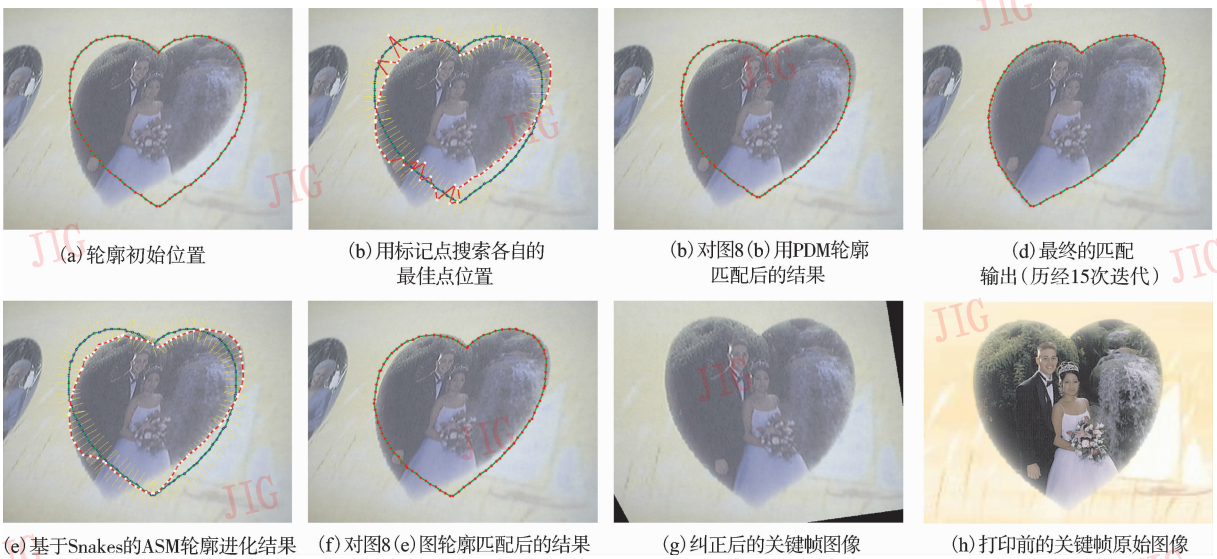


图 8 ASM 搜索步骤

Fig. 8 ASM searching process

(3)在 PDM 的监督下,即可寻找与由最佳位置点组成的轮廓的最佳匹配。实验是采用迭代的匹配算法[13],为确保形变参数的取值在 PDM 的 λ_i 范围内,实验采用 Mahalanobis 距离准则[15],而不是以简单的截断来抑制所有超出范围的形变参数。图 8(c)给出了图 8(b)的匹配结果。

(4)重复步骤(2)、(3),直至收敛。图 8(d)为轮廓收敛结果,当以下条件有一个得到满足时,即认为算法收敛:

①迭代次数超过 50 次;

②迭代中对形变参数的抑制次数超过 20 次;

③迭代前后 $\boldsymbol{X}$ 的变化不明显,亦即标记点的最大偏移量小于阈值(系统中取值为 0.4)。

3.1.2 改进的基于 Snakes 的 ASM 轮廓进化策略

从图 8(b)可见,在搜索得到的由最佳点组成的轮廓中,特别是在心形尖端附近,众多最佳点呈锯齿

状交错,且大部分并非“最佳”。这是由于拍摄图像在心形尖端附近边界梯度丢失严重,从而使标记点在仅以图像的局部信息(法线方向上的梯度)作为最佳度量时,搜索结果不准确。之所以在该不利条件下仍能正确提取轮廓,这要归功于 PDM 在监督匹配过程中的作用,因为通过增加迭代次数,PDM 可在一定程度上弥补上述盲搜索的缺陷,但不能期望在更恶劣的情况下,盲搜索的“过失”总能被 PDM 纠正。

直观上分析,在标记点的最佳搜索中,如果不但考虑了该点的局部图像信息,还考虑到标记点间的内在联系,则必可提高搜索精度,使算法更快收敛。而 Kass 等人提出的 Snakes 正是基于该思想,为此本文提出一种把 Snakes 能量函数直接作为衡量 ASM 标记点搜索的最佳度量的方法。该方法通过合理设定相关控制参数,不仅能提高 ASM 的鲁棒性,还能加快其迭代收敛速度。

Snakes 是在图像分析中应用广泛的一种可变形模板。在图像中,Snakes 可显式地用参数化轮廓表示如下:

$$\boldsymbol{V}(s) = (x(s), y(s))^{\mathrm{T}} \tag{13}$$

式中,x, y 均为坐标函数,$s \in [0,1]$ 表示参数域。上述参数化轮廓在能量泛函

$$E(\boldsymbol{V}) = E_{\mathrm{int}} + E_{\mathrm{ext}} \tag{14}$$

的驱动下搜索,收敛后的轮廓对应于能量极小值。式(14)中的第 1 项为

$$E_{\mathrm{int}} = (\alpha(s) \mid \boldsymbol{V}_s(s) \mid^2 + \beta(s) \mid \boldsymbol{V}_{ss}(s) \mid^2)/2 \tag{15}$$

E_{int} 表示内部形变能量,它描述了可变形轮廓的伸展性与弹性。两个非负的参数函数 $\alpha(s)$,$\beta(s)$,分别决定了轮廓在 Snakes 上任意点 s 的模拟物理特性:$\alpha(s)$ 用于控制轮廓的张力,而 $\beta(s)$ 则决定其刚性。只要 $\alpha(s)$ 与 $\beta(s)$ 有一个为零,则轮廓允许在点 s 处不连续。由于在 ASM 搜索过程中有轮廓的先验形状信息,因此可预先求得在 t 次轮廓匹配后的每个标记点的 1 阶导数 $\boldsymbol{V}_s^{(t)}(s)$ 与 2 阶导数 $\boldsymbol{V}_{ss}^{(t)}(s)$ 的值,在 $t+1$ 次 Snakes 轮廓进化过程中,设定

$$\begin{cases} \alpha^{(t+1)}(s) = a / \mid \boldsymbol{V}_s^{(t)}(s) \mid^2 \\ \beta^{(t+1)}(s) = b / \mid \boldsymbol{V}_{ss}^{(t)}(s) \mid^2 \end{cases} \tag{16}$$

式中,a,b 为常量,它们用于决定张性与弹性形变能量在内部能量函数中的比重,式(16)既归一化了两个内部形变能量,又利用了 PDM 的先验形状信息。实验中设定第 2 项为图像的边界泛函,即

$$E_{\mathrm{ext}} = -c \mid \nabla[G_\sigma * I(x,y)] \mid \tag{17}$$

这里 c 为常量因子,∇为梯度算子,而 $G_\sigma * I(x, y)$ 表示图像与高斯光滑滤波器卷积后的结果,σ 用于控制 E_{ext} 局部极值在空域的扩展,而且算法可方便推广至多尺度空间。图 8(e)给出了在与图 8(b)相同的初始条件下,基于 Snakes 的 ASM 搜索的第 1 次迭代后的效果图,对比两图可发现,图 8(e)中锯齿效应明显得到抑制。图 8(f)是其 PDM 匹配后的结果,对比图 8(c),从轮廓与男士手背的相交部位可见,新的方法收敛得更快。实验结果表明,新搜索算法迭代 8 次后即收敛,少于原算法的 15 次。

3.1.3　形状模板判定

虽然在视频相册生成子系统中可支持多种形状模板,但在用户拍摄关键帧时,系统并不知晓其被哪个形状模板所裁剪,这就需要一个有效的形状模板判定准则,用于选择最匹配的形状模板。由于在 ASM 搜索中形变参数的被抑制,即预示着形状模板的不匹配,因此实验中首先应排除那些迭代过程被 3.1.1 节步骤(4)中条件②强行终止的候选形状模板。最终认定的具有最小平均匹配误差(最后一次的迭代中第(2)步与第(3)步输出轮廓间的平均误差)的形状模板即是最佳匹配。

实验中,当拍摄的图像尺寸是 640 × 480,有 5 个候选的轮廓模板时,则在 P4 2.8G 计算机上整个轮廓定位与模板判定过程的计算时间少于 0.2s。

3.2　关键帧成像畸变纠正

通常拍摄关键帧时,由于相册平面与相机光轴并不垂直,这就导致了关键帧的成像畸变,因此在提取关键帧的签名前,应该用透视投影变换来恢复关键帧的形状,拍摄的图像 $I(x_i, y_i)$ 到纠正图像 $I_0(u_i, v_i)$ 的变换公式如下:

$$\rho \begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix} = \begin{bmatrix} m_{00} & m_{01} & m_{02} \\ m_{10} & m_{11} & m_{12} \\ m_{20} & m_{21} & m_{22} \end{bmatrix} \begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} \tag{18}$$

这里 ρ 为比例因子。为计算透视变换系数 $m_{i,j}$ ($0 \leqslant i,j \leqslant 2$,其中已知 $m_{2,2} = 1$),需要 4 对对应点的坐标。由于 ASM 搜索确保了定位的轮廓与其形状模板间标记点的对应关系,因此实验可先采用 DP 算法[12]从 $2n$ 个标记点中挑选出 4 对顶点,然后用奇异值分解(SVD)方法求解式(18);最后把解出的

透视变换应用于拍摄的图像,即可纠正其成像畸变(如图 8(g)所示)。通过与图 8(h)的比对可发现,纠正后的图像与原始图像非常接近。

最终,通过提取纠正后的关键帧的签名,并在签名数据库中找到其最佳匹配,则系统就能自动播放由匹配的关键帧所表征的场景片段。

4 实验评测

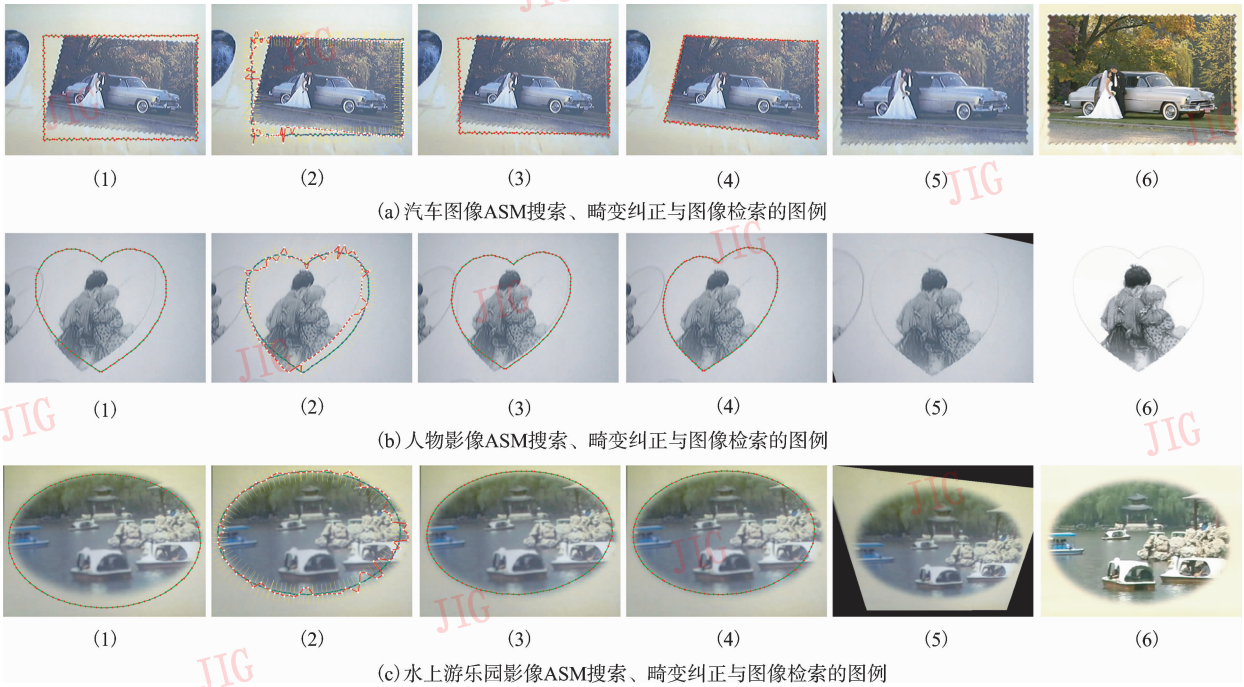
为验证本视频相册系统的应用效果,利用一些实际拍摄的视频图像进行了检索实验。实验用的图像取自包括 20 个家庭摄影剪辑的视频数据库。数据库共包括 1 850 个镜头与 5 098 个子镜头。实验时,首先把每个摄影剪辑分割成 10 个场景[3],共得到 10 × 20 = 200 个场景,接着用 2.1 小节的关键帧选择方案提取 200 个关键帧及其签名,并存于数据库中。实验时随机选择其中的 50 个关键帧用来对系统进行评估,每个选择的关键帧均被 5 个形状模板(如图 3 所示)所裁剪,首先得到 50 × 5 = 250 个关键帧,并用激光打印机打印出来(彩印与黑白打印各一半);然后对每个打印出的关键帧,用摄像头每隔 30s 连续拍摄,且在拍摄过程中始终保持关键帧处于拍摄图像中的居中位置,通过不断改变摄像头的姿态与视角,共拍摄到 100 幅包括不同分辨率的图像(从 160 × 120 到 640 × 480);最终共得到 250 × 100 = 25 000 幅拍摄的图像。实验时,先用 ASM 算法定位关键帧在拍摄的图像中的轮廓边界,并在纠正其成像畸变后,提取其签名,若根据提取的签名能从数据库中检索到正确的视频片段,即认为测试成功,反之失败。表 1 列出了对 5 个形状模板的检索成功率,显示了视频相册系统具有很高的精度。

表 1 形状模板的检索精度

Tab. 1 The searching precision of shape templates

形状模板	矩形	椭圆形	心形	扇形	邮票形
图像数目	5 000	5 000	5 000	5 000	5 000
成功比率	1.00	0.98	0.95	0.97	0.93

图 9 显示基于了自训练的 ASM 算法定位关键帧轮廓的其他例子。尽管图 9(c)的定位结果并不准确(由于关键帧边界羽化严重),但得益于关键帧签名很强的区分性能,实验仍能检索到对应的视频。



(a)汽车图像ASM搜索、畸变纠正与图像检索的图例

(b)人物影像ASM搜索、畸变纠正与图像检索的图例

(c)水上游乐园影像ASM搜索、畸变纠正与图像检索的图例

(每行自左至右分别表示:(1)轮廓初始位置;(2)标记点搜索各自的最佳点位置(以第 5 次迭代搜索为例);(3)对上图用 PDM 轮廓匹配后的结果;(4)最终的匹配输出;(5)纠正后的关键帧图像;(6)打印前的关键帧原始图像)

图 9 其他 ASM 搜索与成像畸变纠正的例子

Fig. 9 More ASM searching and shape restoration examples

5 结　论

本文提出了一种新的视频检索与浏览系统,该系统能为用户,特别是不熟悉电脑使用的用户提供一种新的视频管理与浏览方式。应用基于自训练与 Snakes 进化的 ASM 算法,视频相册系统即能根据模拟相册中的关键帧,为用户检索并回放对应的视频片段。

本系统还可做如下扩展:可以基于颜色或灰度特征的相似性,根据拍摄的关键帧来检索内容相近的视频片段;同时系统亦可用于管理照片数据库,并可根据拍摄的图像来检索其所在的照片集;当数据库容量过大,且关键帧签名的区分能力不足以区分时,还可把关键帧分成多册,用户只需事先拍摄相册的 ID 图像,即可把关键帧签名的比对限制在该相册范围内。同时,基于 ASM 的形状模板判定算法虽然简单有效,但由于其计算量会随着形状模板数量的增加呈线性增长,因此系统还需考虑新的形状模板预判断机制。上述扩展与改进之处亦是将来工作的研究重点。

致　谢　感谢微软亚洲研究院华先胜研究员对论文的指导工作!

参考文献(References)

1 Ekin A, Tekalp A M, Mehrotra R. Integrated semantic-syntactic video modeling for search and browsing [J]. IEEE Transactions on Multimedia, 2004, **6**(6): 839 ~ 851.

2 Dagtas S, Al-Khatib W, Ghafoor A, *et al*. Models for motion-based video indexing and retrieval [J]. IEEE Transactions on Image Processing, 2000, **9**(1):88 ~ 101.

3 Hua Xian-sheng, Li Shi-peng, Zhang Hong-jiang. Video booklet [A]. In: Proceedings of IEEE International Conference on Multimedia & Expo [C], Amsterdam, the Netherland, 2005: 189 ~ 192.

4 Zhu Cai-zhi, Hua Xian-sheng, Mei Tao, *et al*. Video booklet-a natural video searching and browsing interface [A]. In: Proceedings of the 7th ACM SIG Multimedia International Workshop on Multimedia Information Retrieval [C], Singapore, 2005: 113 ~ 120.

5 Zhu Cai-zhi, Mei Tao, Hua xian-sheng. Natural video browsing [A]. In: Proceedings of ACM SIG Multimedia [C], Singapore, 2005: 265 ~ 267.

6 Hua Xian-sheng, Chen Xian, Zhang Hong-jiang. Robust video signature based on ordinal measure [A]. In: Proceedings of International Conference on Image Processing [C], Singapore, 2004: 685 ~ 688.

7 Kass M, Witkin A, Terzopoulos D. Snake: active contour model [J]. International Journal of Computer Vision, 1987, **1**(4):321 ~ 332.

8 Cootes T, Taylor C, Cooper D, *et al*. Active shape models-their training and their applications [J]. Computer Vision and Image Understanding, 1995, **61**(1): 38 ~ 59.

9 Hua Xian-sheng, Lu Lie, Zhang Hong-jiang. AVE-automated home video editing [A]. In: Proceedings of ACM Multimedia [C], Berkeley, CA, USA, 2003, 490 ~ 497.

10 Ma Yu-fei, Lu Lie, Zhang Hong-jiang, *et al*. A user attention model for video summarization [A]. In: Proceedings of ACM Multimedia [C], Juan-les-Pins, France, 2002: 533 ~ 542.

11 Ma Yu-fei, Zhang Hong-jiang. Contrast-based image attention analysis by using fuzzy growing [A]. In: Proceedings of ACM Multimedia [C], Berkeley, CA, USA, 2003: 374 ~ 381.

12 Douglas D, Peucker T. Algorithms for the reduction of the number of points required to represent a digitized line of its caricature [J]. The Canadian Cartographer, 1973, **10**(2):112 ~ 122.

13 Baldock R, Graham J. Image Processing and Analysis-A Practical Approach [M]. London, UK: Oxford University Press, 2000.

14 Zhou Yi, Gu Lie, Zhang Hong-jiang. Bayesian tangent shape template: estimating shape and pose parameters via Bayesian inference [A]. In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition [C], Madison, Wisconsin, USA, 2003: 109 ~ 118.

15 Cootes T, Taylor C. Statistical Models of Appearance for Computer Vision[EB/OL]. http://www.isbe.man.ac.uk/~bim/Models/app_model.ps.gz. 2005-7-1.