

脱机手写女书字符图像多方向文本行提取*

黑光月, 王江晴, 孙阳光

(中南民族大学 计算机科学学院, 武汉 430074)

摘要: 针对传统文本行提取方法的不足, 提出一种脱机手写女书字符图像多方向文本行提取方法。该方法先将预处理后的女书字符图像通过形态学膨胀和最小外接矩形来提取出女书字符块, 然后用 Delaunay 三角剖分对字符块构建三角网, 再根据三个给定的准则计算三角网中每条边的权重构建链接模型, 最后根据依权重排序后的边, 利用互斥原则搜索出最优文本行。实验结果表明, 该方法能有效提取扇面、布帕等不同载体上女书字符图像中的多方向文本行。

关键词: 女书; 多方向文本行; 脱机手写; 最小外接矩形; Delaunay 三角剖分; 链接模型

中图分类号: TP391.1 **文献标志码:** A **文章编号:** 1001-3695(2013)02-0627-04

doi: 10.3969/j.issn.1001-3695.2013.02.086

Multi-oriented text lines extraction from offline Nǚshu characters image

HEI Guang-yue, WANG Jiang-qing, SUN Yang-guang

(College of Computer Science, South-Central University for Nationalities, Wuhan 430074, China)

Abstract: For the problems of the traditional text lines extractions, this paper proposed a new method to extract multi-oriented text lines from offline handwritten Nǚshu characters image. Firstly, by morphological dilating and minimum enclosing rectangle, it extracted blocks of Nǚshu characters from the preprocessed Nǚshu characters image. It built a triangulation network by Delaunay triangulation on these blocks. According to three given rules, it constructed a link model by calculating the weight of each edge of the triangulation network. Finally, it extracted the optimal text lines by the ordered edges and mutual exclusion. The experimental results show that this method can effectively extract the multi-oriented text lines from Nǚshu characters images on different carriers, such as fans and handkerchiefs.

Key words: Nǚshu; multi-oriented text lines; offline handwriting; minimum enclosing rectangle; Delaunay triangulation; link model

0 引言

随着计算机性能的不不断提升以及 OCR 软件的快速发展, 人们希望通过计算机自动识别具有复杂图文结构的文献图像。计算机在识别复杂文档图像之前, 必须首先提取出其中的文本行, 然后对文本行进行字符切分, 再将切分出来的字符输入计算机进行识别。可见文本行提取是字符切分的前提, 其提取的好坏间接影响字符的识别效果。一般的印刷体文档图像和比较规范的手写体文档图像中行与行之间相互平行, 其文本行结构比较规范, 用投影直方图^[1]的方法便可提取出文本行。纸扇上记录的女书^[2]文献, 行与行之间有一定角度, 文本行呈多方向发散结构, 并且有些扇面上存在图案, 这些都加大了文本行提取的难度。

国内外学者对文档图像文本行提取已经作了多方面的研究。文献[3]采用基于游程连通域分析的方法对文本行进行提取, 但其只能提取印刷体中水平方向的文本行。文献[4]采用多尺度的边缘检测法提取出文本信息, 但其不能提取出文本行的结构信息。文献[5]针对联机手写笔迹文本行, 采用基于感知的链接模型有效地提取出多方向的笔迹结构, 但该方法不

适用于脱机文档图像的多方向文本行提取。文献[6]用每个英文字符的候选特征点计算出英文文本行的基线, 再根据基线信息来提取多方向文本行, 该方法无法处理基线不齐的英文文本行, 也不适合处理复杂的女书字符。文献[7]首先将阿拉伯文献图像分为若干子窗口, 根据每个子窗口的投影轮廓信息合并文本行书写方向相同的窗口, 进而提取出多方向文本行, 该方法适合书写比较密集的阿拉伯文献, 但不能解决字符稀疏的文献和带图案的文献, 同样不适合扇面上的女书文本行提取。

本文针对脱机手写体女书多方向文本行的特点, 提出一种脱机手写女书图像多方向文本行提取方法。该方法先将女书字符图像进行预处理, 然后通过形态学膨胀和最小外接矩形提取女书字符块, 采用 Delaunay 三角剖分^[8]对字符块构建三角网, 再根据三个准则计算三角网中每条边的权重, 最后根据权重排序后的边, 利用互斥原则搜索出最优文本行。

1 女书图像多方向文本行提取

1.1 女书多方向文本行简介

女书是世界上唯一的女性文字, 它起源于湖南省永州市江永县, 因此又叫江永女书^[2]。女书字符为斜体, 呈“多”字形,

收稿日期: 2012-07-01; **修回日期:** 2012-08-14 **基金项目:** 国家自然科学基金资助项目(60975021)

作者简介: 黑光月(1987-), 男, 河南郑州人, 硕士, 主要研究方向为模式识别、数字图像处理(heiguangyue@126.com); 王江晴(1964-), 女, 教授, 博士, 主要研究方向为模式识别、数字图像处理; 孙阳光(1978-), 男, 讲师, 博士, 主要研究方向为计算机视觉、图像处理与分析、目标识别与跟踪等。

其书写格式为自上而下、自右向左,其文本行为竖行(即女书的文本行书写方向同汉字的列的书写方向),书写载体有纸质手抄本、纸片、扇面、布帕、花带等。其中以扇面女书文献最为特别,其字符文本行是呈多方向的,扇子边缘存在边框,并且有些扇面中存在图案,这些因素大大增加了扇面女书文献中文本行提取的难度。

1.2 算法流程

图 1 为本文采用的多方向文本行提取算法流程示例,步骤如下:

- a) 预处理。将输入的彩色扇面女书图像灰度化,然后二值化,再对其去噪,得到原始二值图像。
- b) 将原始二值图像中的笔画进行合并,即将邻近的笔画合并为类似于单个字体的字符块,原则上只要不同行的字不合并在一起即可。
- c) 统计合并后字符块的面积,只保留与字符块平均面积相差不大的字符块,删除非字符块,其中非字符块可能包括扇子边框、图案等。
- d) 对每个字符块求解最小外接矩形。
- e) 以字符块的重心为顶点,利用 Delaunay 三角剖分建立平面三角网结构。
- f) 根据三个重要参数计算出三角网中每条边的权重,该权重值反映了相邻两个字符块处于同一文本行的可能性。
- g) 对三角网中边的权重值从大到小排列,再根据贪心策略,利用互斥原则搜索出最优文本行。

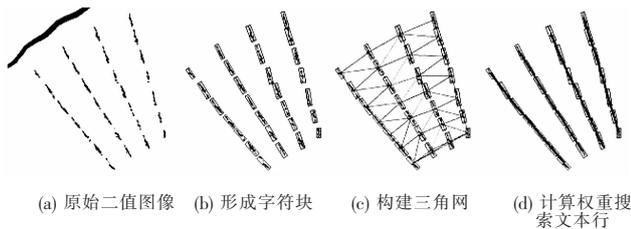


图 1 算法流程示例

1.3 字符块的形成与选择

字符块是构成文本行结构的最基本单元,由一个或多个笔画构成,确定每个字符块由哪些笔画构成是文本行提取的基础与前提。由于扇面女书字符行与行之间以及同一行中上下字符之间的间隔相对于同一个字符的笔画之间的间隔较大,为使同一字符的笔画聚集在一起,采用半径为 R 的圆形结构元素对预处理后的二值图像进行膨胀处理^[9],以得到字符块的位置和边缘,原则上只要处于不同行的笔画不合并在一起即可。如图 2(a)所示,对图 1(a)中的二值女书文献进行形态学膨胀操作,同一字符内的笔画膨胀在一起,构成字符块。

扇子图像中有扇形边框,即扇子中扇面区域的边缘部分,并且有些扇子中存在图案,这些边缘和图案给字符块的选择带来了巨大的干扰,因此应将边缘和图案去掉。假定膨胀后的二值图像中的连通域集合为 $C = \{c_1, c_2, c_3, \dots, c_n\}$,其中 c_i 表示第 i 个连通域,连通域总数为 n ,且 c_i 对应连通域的面积记为 a_i ,若满足如下条件:

$$a_i > k \times a_{avg} \quad (1)$$

则 c_i 为边缘或者图案,否则 c_i 为字符。其中: a_{avg} 为膨胀后字符块面积的平均值, k 为常数。图 2(b)为利用式(1)将图 2(a)中左上角的扇面边缘去掉,仅保留字符块区域。

1.4 字符块最小外接矩形

为了近似表示每个字符块的面积以及确定每个字符的书写方向,为链接模型的建立提供基础,需要对每个字符块求解最小外接矩形。在求解字符块最小外接矩形之前,先使用格雷厄姆法^[10]求出字符块的凸多边形,然后使用旋转卡壳^[11]求解凸多边形的最小外接矩形。该算法的实现基于一个重要的结论,即最小外接矩形与凸多边形至少有一条公共边。具体实现步骤如下:

- a) 计算凸多边形所有顶点中的最左点、最右点、最下点、最上点,分别记为 P_{xmin} 、 P_{xmax} 、 P_{ymin} 、 P_{ymax} 。
- b) 分别过 P_{xmin} 和 P_{xmax} 作平行于 y 轴的平行线,再分别过 P_{ymin} 和 P_{ymax} 作平行于 x 轴的平行线,如图 3(a)所示,两组平行线构成了一个封闭的矩形,并且与凸多边形的边形成了四个夹角,分别记做 θ_i 、 θ_j 、 θ_k 、 θ_l 。
- c) 设封闭矩形的面积为 S ,并将最小外接矩形的面积 S_{min} 赋值为无穷大。
- d) 令 $\theta = \min(\theta_i, \theta_j, \theta_k, \theta_l)$ 。
- e) 将封闭矩形顺时针旋转 θ ,使得封闭矩形至少有一条边与凸多边形的一边重合,如图 3(b)所示。
- f) 计算当前封闭矩形的面积 S ,如果 $S < S_{min}$,则令 $S_{min} = S$,并记录当前封闭矩形的四个顶点。
- g) 重新计算 θ_i 、 θ_j 、 θ_k 、 θ_l 。
- h) 重复 d) ~ g),直至平行线已旋转角度超过 90° 。

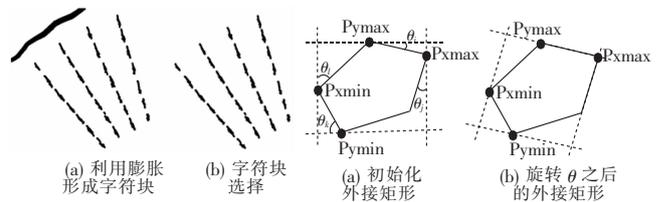


图 2 字符块形成与选择示例

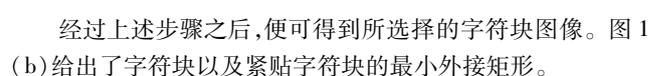


图 3 求解最小外接矩形示意图

1.5 链接模型

链接模型^[12]是对字符块邻近关系进行描述的一种模型,并且是最优文本行提取的前提。本文利用 Delaunay 三角剖分将字符块连接成三角网结构,然后根据三个重要参数计算出每条边的权重,建立字符块的链接模型。

1.5.1 建立三角网

本文采用 Delaunay 三角剖分对字符块进行三角网的构造。Delaunay 三角剖分的基本定义为:假设 V 是二维实数域上的有限点集,每一点 P_i 对应一个区域:

$$\{V_i\} = \{P: \|P - P_i\| \leq \|P - P_j\|, \forall j \neq i\} \quad (2)$$

如果 V_i 内的点到 P_i 的距离不大于到 P_j 的距离,连接全部相邻的 Voronoi 多边形生成点所构成的三角网格称为 Delaunay 三角剖分^[13]。其中, V_i 是 P_i 的 Voronoi 多边形, P_i 是 Voronoi 多边形 V_i 的生成点。Delaunay 三角剖分除了具有空外接圆与最大最小角性质以外,还具有最接近和唯一性的优点。最接近表示以最近邻的三点形成三角形,该特性保证了只有存在近邻关系的字符块之间才会建立链接,而距离较远的字符之间不会建立链接,这样在模型建立阶段就可避免较远字符间产生链接关系。唯一性表示不论从区域何处开始构建,最终得到的剖分结果一致,该特性保证了无论从哪一个字符块重心开始构造,最

终的三角网结构始终是唯一。本文采用逐点插入的方法对字符块的重心坐标进行 Delaunay 三角剖分,剖分后三角网图像如图 1(c) 所示,其中的三角网比较规则,满足本文所需链接模型中只有近邻字符之间建立链接的条件。

1.5.2 计算链接权重

链接权重反映了相邻两个字符块处于同一文本行的可能性。对于女书字符而言,一般同一方向文本行近邻字符的大小相差不大;同一方向近邻字符最小外接矩形的长边与该方向基线的夹角相差不大;同一方向近邻字符的间距较近,而不同方向行之间的字符间距较大。综合上述三个准则,在计算链接权重时主要考虑以下三个因素:链接两端字符块的大小差异、字符块最小外接矩形长边与链接之间的夹角、链接长度。

a) 链接两端字符块的大小差异。处于同一方向行近邻字符块的大小应该相差不过大,相差太大说明不是同一文本行的字符块。为了较准确地描述字符块的面积,本文采用字符块最小外接矩形的面积近似表示字符块的面积。设 S 为大小差异对链接权重的影响参数,则

$$S = 1 - \frac{|s_1 - s_2|}{\max(s_1, s_2)} \quad (3)$$

其中: s_1, s_2 分别表示链接两端字符块的面积。

b) 字符块最小外接矩形长边与链接之间的夹角。任何文字行在书写时都会保持一定的方向,虽然不同载体上的女书字符在整体上是呈多方向的,但是同一文本行中文字的方向是一致的。因此本文采用字符块方向与链接方向的夹角来刻画每一字符块与文本行书写方向上的偏差性。女书字符是一种长菱形的字体,其最小外接矩形有明显的长边与短边,因此用长边的方向表示女书字符的书写方向。设 A 为夹角对链接权重的影响参数,则有

$$A = \prod_{i=1}^2 \frac{w_i \times \cos^2 \alpha_i + h_i \times \sin^2 \alpha_i}{w_i + h_i} \quad (4)$$

其中: w_i, h_i 分别为字符块最小外接矩形的短边和长边; $\alpha_i (i = 1, 2)$ 为链接方向与链接相连字符块方向的夹角,如图 4 所示。

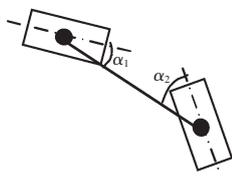


图 4 字符块与链接夹角

c) 链接长度。链接长度是相邻字符块重心连线在其最小外接矩形外部的线段长度,是计算链接权重时最重要的一个因素。设链接 l 对链接权重的影响参数为 L ,则

$$L = \exp\left(-\left(\frac{l}{\bar{l}}\right)^2\right) \quad (5)$$

其中: \bar{l} 为链接长度的平均值, l 为当前链接的长度。

考虑到三个因素对权重计算的影响不同,因此用式(6)计算链接权重:

$$W = p \times S + q \times A + r \times L \quad (6)$$

其中: p, q, r 分别为三个影响参数的系数,且 $p + q + r = 1$ 。由于链接长度 L 对权重的影响最大,链接两端字符块的大小差异以及字符块最小外接矩形长边与链接之间的夹角对权重的影响相对较小,所以本文 p, q 取值较小, r 取值较大。

1.6 最优文本行提取

链接模型为最优文本行提取提供了解空间。最优文本行

提取可以看做是从链接模型中提取出一组具有最大链接强度和的非冲突链接。冲突链接表示属于同一个字符块的两个链接,且它们不满足近似共线的条件。此处近似共线的条件是指这两个链接的夹角小于阈值 α_T ,而非冲突链接表示夹角较小的链接。这样就把最优文本行提取问题转换为求解一个具有最大链接强度和的非冲突链接子集。

最优文本行提取采取贪心策略,其思想是选择当前权重最大的非冲突链接,过程示例如图 5 所示。设 $G = (V, E)$ 为带权重的三角网结构,则它是一个无向连通图,其中 V 为字符块重心的集合, E 为字符块之间的链接集合。设 B 为最优行集合, e_c 为与集合 B 中边相互冲突的边,则搜索算法步骤如下:

- a) 将集合 B 初始化置空。
- b) 当链接集合 E 不为空时,从 E 中选取当前权重最大的边 e 。
- c) 在集合 E 中删除 e 以及与 e 相互冲突的边,再将边 e 添加到集合 B 中并跳至步骤 b)。
- d) 根据 B 中元素之间是否存在链接对集合 B 进行子集的划分,使每一个子集中仅包含同一方向行的字符。

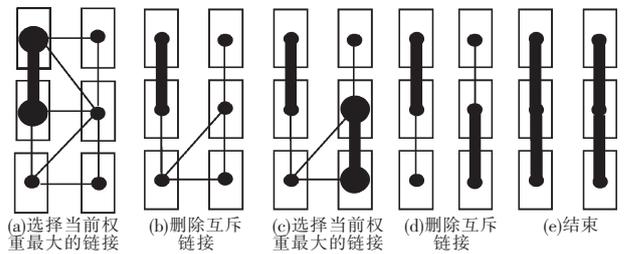


图 5 最优文本行排列过程

2 实验结果及分析

本文算法在 Windows 7 操作系统, MATLAB2010a 环境下编程实现,将其应用于两组脱机女书字符图像中进行测试。一组为 30 个扇面女书字符图像,另一组为 20 个布帕女书字符图像,共计 395 行。在实验中,参数 $R = 5, k = 5, \alpha_T = 0.125\pi, p, q, r$ 分别取 0.2、0.2、0.6。

图 6 所示为本文算法对扇面女书字符图像文本行提取结果示例。从图 6(a)(b) 中可以看出,本文利用字符形成与选择能正确地复杂扇面背景中提取出女书字符部分,并根据链接模型正确地提取出向四周发散的多方向文本行。从图 6(c) 可以看出,提取出的文本行每行只包含字符块和相邻两字符块之间的链接,同一竖行中相邻两个字符符合在一起不影响最终文本行提取结果。图 7(a) 为用针线绣在布帕中的女书字符图像,由于布帕材料柔软,图中字体难免有少许扭曲,并且图中同时存在横行、竖行,加之拍摄角度不正导致图像中的文本行倾斜较大,从图 7(c) 可以看出,本文算法依然能够正确提取出文本行。图 8 给出了弯曲女书图像文本行提取示例,由于弯曲文本行满足局部的线性条件,本文算法依然能正确提取出弯曲行女书字符图像的文本行结构。

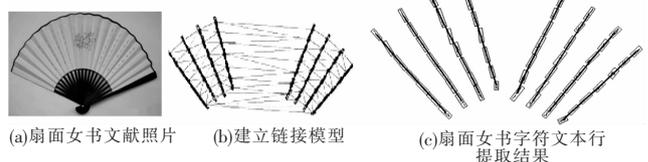


图 6 扇面女书字符图像文本行提取结果

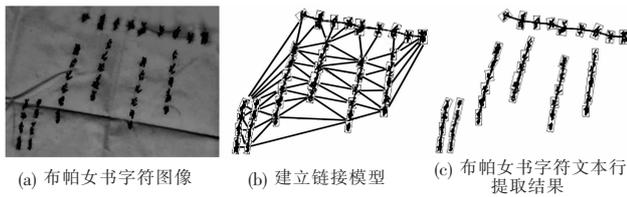


图7 布帕女书字符图像文本行提取结果

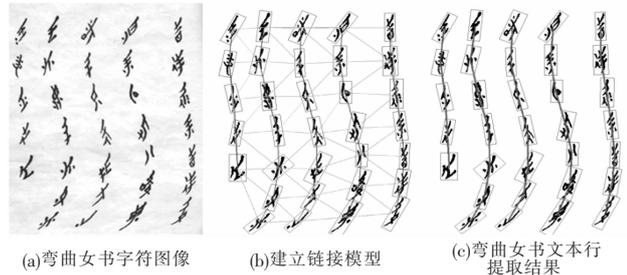


图8 弯曲女书字符图像文本行提取结果

实验结果表明,本文算法能够有效地提取出扇面、布帕等不同载体上女书字符图像中多方向的文本行结构以及少许弯曲的女书文本行结构。

3 结束语

本文实现了一种用带权重的三角网结构自动提取脱机手写女书多方向文本行的方法。实验结果表明,该方法能够有效提取出扇面、布帕等不同载体上女书字符图像中的多方向女书文本行,并且能够提取出布帕图像中存在弯曲的女书文本行,为女书字符切分以及识别奠定了基础。

参考文献:

[1] DOS SANTOS R P, CLEMENTE G S, REN T I, *et al.* Text line segmentation based on morphology and histogram projection [C]//

Proc of the 10th International Conference on Document Analysis and Recognition. 2009:651-655.

[2] 李庆福. 女书文化研究[M]. 北京:人民出版社,2009.

[3] 肖人岳, 秦慕婷. 一种复杂文本图像中快速文本行检测算法[J]. 科学技术与工程, 2008, 8(23):6253-6257.

[4] LIU Xiao-qing, SAMARABANDU J. Multiscale edge-based text extraction from complex images [C]//Proc of IEEE International Conference on Multimedia and Expo. 2006:1721-1724.

[5] 敖翔, 戴国忠, 王宏安. 基于感知的多方向在线手写笔迹文本行提取[J]. 计算机辅助设计与图形学学报, 2007, 19(1):14-19.

[6] PAL U, SINHA S, CHAUDHURI B B. Multi-oriented text lines detection and their skew estimation [C]//Proc of Indian Conference on Computer Vision, Graphics and Image Processing. 2002:270-275.

[7] OUWAYED N, BELAD A. Multi-oriented text line extraction from handwritten Arabic documents [C]//Proc of the 8th IAPR International Workshop on Document Analysis Systems. 2008:339-346.

[8] 周培德. 平面点集三角剖分的算法[J]. 计算机辅助设计与图形学学报, 1996, 8(4):259-264.

[9] 冈萨雷斯. 数字图像处理[M]. 阮秋琦, 阮宇智, 译. 2版. 北京:电子工业出版社, 2007.

[10] 周培德. 计算几何——算法分析与设计[M]. 北京:清华大学出版社, 2000:61-62.

[11] PIRZADEH H. Computational geometry with the rotating calipers [D]. Montréal: School of Computer Science, McGill University, 1999.

[12] AO Xiang, LI Jun-feng, WANG Xu-gang, *et al.* Structuralizing digital ink for efficient selection [C]//Proc of the 11th International Conference on Intelligent User Interfaces. 2006:148-154.

[13] 胡勇, 王国胤, 杨勇. 人脸特征约束点的三维表情合成[J]. 计算机应用研究, 2012, 29(2):754-756.

(上接第609页)

参考文献:

[1] HAUPTMANN A G. Lessons for the future from a decade of informedia video analysis research [C]//Proc of the 4th International Conference on Image and Video Retrieval. 2005:1-10.

[2] SONG Y, HUA Xian-sheng, DAI Li-rong, *et al.* Semi-automatic video annotation based on active learning with multiple complementary predictors [C]//Proc of Workshop on Multimedia Information Retrieval. 2005:97-104.

[3] YAN R, NAPHADÉ M. Semi-supervised cross feature learning for semantic concept detection in videos [C]//Proc of IEEE Conference on Computer Vision and Pattern Recognition. 2005:657-663.

[4] SONG Y, HUA Xian-sheng, DAI Li-rong, *et al.* Semi-automatic video annotation based on active learning with multiple complementary predictors [C]//Proc of ACM International Workshop on Multimedia Information Retrieval. 2005:97-104.

[5] HE Jing-ni, LI Ming-jing, ZHANG Hong-jiang, *et al.* Generalized manifold-ranking based image retrieval [J]. IEEE Trans on Image Processing, 2006, 15(10):3170-3177.

[6] YUAN Xun, HUA Xian-sheng, WANG Meng, *et al.* Manifold-ranking based video concept detection on large database and feature pool [C]//Proc of ACM Multimedia Conference. 2006.

[7] WANG Fei, WANG Jing-dong, ZHANG Chang-shui, *et al.* Semi-supervised classification using linear neighborhood propagation [C]//Proc of IEEE Conference on Computer Vision and Pattern Recognition.

2006:160-167.

[8] 卢汉清, 刘静. 基于图学习的自动图像标注[J]. 计算机学报, 2008, 31(9):1630-1639.

[9] SAUL L K, ROWEIS S T. Think globally, fit locally: unsupervised learning of low dimensional manifolds [J]. Journal of Machine Learning Research, 2003, 4:119-155.

[10] GENG Xin, ZHAN De-chuan, ZHOU Zhi-hua. Supervised nonlinear dimensionality reduction for visualization and classification [J]. IEEE Trans on Systems, Man, and Cybernetics, Part B: Cybernetics, 2005, 35(6):1098-1107.

[11] YANG Gao-ming, YANG Jing, ZHANG Jian-pei. Semi-supervised clustering-based anonymous data publishing [J]. Journal of Harbin Engineering University, 2011, 32(11):1489-1494.

[12] QI Guo-jun, HUA Xian-sheng, RUI Yong, *et al.* Correlative multi-label video annotation [C]//Proc of ACM Multimedia Conference. 2007:17-26.

[13] ZHA Zheng-jun, MEI Tao, WANG Jing-dong, *et al.* Graph-based semi-supervised learning with multi-label [C]//Proc of IEEE International Conference on Multimedia and Expo. 2008:1321-1324.

[14] TANG J, HUA Xian-sheng, MEI Tao, *et al.* Video annotation based on temporally consistent Gaussian random field [J]. Electron Lett, 2007, 43(8):448-449.

[15] TONG S, CHANG S F. Support vector machine active learning for image retrieval [C]//Proc of the 9th ACM International Conference on Multimedia. New York: ACM Press, 2001:107-118.