一种将羽毛球比赛的2D视频转换到3D视频的算法

刘 杨 齐 春 杨静怡
(西安交通大学图像处理与模式识别研究所 西安 710049)

摘 要 文中提出一种羽毛球比赛的2D视频转换到3D视频的算法。在这类视频中,前景是最受关注的部分,准确地从背景中提取出前景对象是获取深度图的关键。文中采用一种改进的图割算法来获取前景,并根据场景结构构建背景深度模型,获取背景深度图;在背景深度图的基础上,根据前景与镜头之间的距离关系为前景对象进行深度赋值,从而得到前景深度图。然后,融合背景深度图和前景深度图,得到完整的深度图。最后,通过基于深度图像的虚拟视点绘制技术DIBR来获取用于3D显示的立体图像对。实验结果表明,最终生成的立体图像对具有较好的3D效果。

关键词 羽毛球比赛视频,3D转换,深度图的提取,改进的图割算法,DIBR
中图法分类号 TN911.73 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2018.08.011

2D-to-3D Conversion Algorithm for Badminton Video

LIU Yang QI Chun YANG Jing-yi
(Institute of Image Processing and Pattern Recognition, Xi'an Jiaotong University, Xi'an 710049, China)

Abstract This paper proposed a 2D-to-3D conversion algorithm for badminton video. The most attractive part of badminton video is the foreground. The core of the depth map extraction is to separate the foreground objects accurately from the background. The improved grab cut segmentation algorithm is used to extract foreground regions. A model for the background depth is constructed based on the structure of scene. The depth value is assigned for foreground based on the distance of scene objects from the viewpoint and the background depth map. Then the depth of foreground and background are merged. Finally, the synthesized stereo pairs of images for 3D display are obtained by DIBR formula. The experimental results show that the generated stereo images have good 3D perception performance.

Keywords Badminton video, 3D conversion, Depth map extraction, Improved grab cut segmentation, DIBR

1 引言

尽管目前具有3D功能的硬件设备众多,但是可用于显示的3D素材一直存在着片源不足的问题,这极大地制约了3D技术的推广。现有的2D视频资源种类繁多、获取方式简单,2D技术也相对比较成熟,将这些已有的2D视频转换成具有立体效果的3D视频是解决片源不足问题的一个很好的途径。普通的2D视频转3D视频的流程中主要包含两个关键步骤:1)从输入的2D视频中恢复出已经丢失的深度信息,生成对应的深度图;2)合成虚拟的立体图像对[1]。由于实际场景复杂多样,从这些不同的场景中恢复出深度信息的难度较大,而立体图像对的获取流程相对比较固定[2],受场景影响不大,因此有更多的学者将重心放在了前一步的深度图提取上。已有的深度图提取方法大致可以划分为两类:半自动方法和全自动方法[3]。半自动的深度图提取方法[4-7]需要有专业的技术人员参与深度分配,虽然得到的深度图的细节部分更加准确,但是需要消耗大量的时间和人力。相比之下,全自动方法不需要人工指配深度,因具有高效性和便捷性而得到广泛使用。

一些全自动方法使用了场景中所蕴含的深度线索来获取深度图,如线性透视线索[8]、纹理线索[9]、聚焦/散焦线索[10-11]、相对高度线索[12]等,但是这些方法只在包含有对应线索的场景中有效。为了增强算法的场景适应性,文献[13-14]提出了基于混合深度线索的深度图提取方法。基于机器学习的方法[15-16]也被用于深度图的提取中,但是目前所提出的方法仅适用于少数几种简单场景,比如包含建筑物、天空、陆地等的户外自然场景以及普通的室内场景。另外,基于机器学习的方法还存在两个方面的缺陷:1)非常依赖于大量的RGBD训练图像集[3];2)每当输入新的2D图像,都需要重新训练网络参数。

Ji等[17]使用分块的方法把图像划分成具有不同深度的块区域,然后以场景中天空与陆地交汇处的地平线作为深度最大的位置来获取深度分布梯度图,并在此基础上为各个块区域进行深度赋值。Cheng等[18]使用基于区域的聚类方法,按照颜色相似性以及像素间的空间距离关系把输入图像划分成大量的同质区域,然后结合先验深度模型为这些区域进行深度赋值。这两种方法都比较适用于自然风光类等静态场景的深度图提取,对含有运动目标的场景不太适用,容易将具有同一深度的目标划分成不同的区域。Cheng等[19]将由运动

到稿日期:2017-10-24 返修日期:2017-12-20 本文受国家自然科学基金(61572395,61133008)资助。
刘 杨(1993-),女,硕士生,主要研究方向为2D/3D视频转换,E-mail:yangliu5040@163.com;齐 春(1955-),男,教授,博士生导师,主要研究方向为图像超分辨率增强、图像检测与跟踪、图像分析与识别,E-mail:qichun@mail.xjtu.edu.cn(通信作者);杨静怡(1993-),女,硕士生,主要研究方向为机器学习和深度学习。

视差获得的深度信息和由几何透视获得的深度信息按照一定的权重比例相融合，以获取深度图。Freiling 等[20]提出了运动目标检测与背景深度获取相结合的深度图提取方法。Nam 等[21]使用颜色直方图和累计帧差法来分离前景和背景，然后使用先验深度模型进行深度赋值。这几种方法都比较适用于静止背景中含有运动目标的场景，但它们都存在前景检测不够准确的缺陷，从而降低了深度图的质量和观看的舒适度。

目前，所提出的众多研究方法只有在某些简单场景下才能够获得较好的效果，还不存在一种通用的、适用于所有场景的转换算法。羽毛球比赛是当今世界上比较受欢迎的体育项目之一，属于众多 2D 视频中观众比较喜欢的含有运动目标的场景，这类视频场景的内容比较丰富，既含有多个运动对象的场景，又含有人物特写镜头和场外的一些普通场景；另外，目前也鲜有专门将羽毛球比赛的 2D 视频转换成 3D 视频的算法的相关研究，因此，本文选取 2D 羽毛球比赛视频作为研究对象，通过充分地分析场景、深入地挖掘其中有用的先验信息，来做出更好的立体效果。

羽毛球比赛视频中主要包含 3 类镜头：1)摄像机正对着整个球场区域、运动员、裁判以及观众所拍摄下来的全局场景；2)对运动员聚焦的近景镜头；3)场外的观众席镜头。因为第一类镜头在整个视频中所占的比例较大，并且是观众最感兴趣的部分，所以本文将针对第一类镜头进行深入研究。观众在观看这类视频镜头时，注意力会主要集中在运动的前景对象上，因此对该类场景的深度图提取应该关注于前景与背景的分开处理。本文首先根据场景的深度分布结构来构建相应的深度模型；其次为背景区域进行深度赋值，以获取背景深度图；然后采用改进的图割算法[22]来提取前景对象，并根据前景对象与镜头间的距离关系及其在场内的初始深度值大小为其分配合理的深度值，从而得到前景深度图；最后通过融合前/背景深度获取完整的深度图，使用基于深度图像的虚拟视点绘制技术(Depth Image Based Rendering)来获取用于在 3D 播放设备上显示的立体图像对。

本文第 2 节详细介绍了深度图的提取过程；第 3 节介绍了基于深度图像的虚拟视点绘制技术(DIBR)；第 4 节对提出的算法进行测试，并展示了部分实验结果；最后总结全文。

$$I(i,j)=\begin{cases}0, & \\ |G(i,j)-R(i,j)|+|G(i,j)-B(i,j)|+G(i,j), & \text{otherwise}\end{cases}$$

$$I(i,j)=\begin{cases}0, & I(i,j)<T\\ 1, & I(i,j)\geqslant T\end{cases} \tag{3}$$

3)找到 I 中所有值为 1 的像素中的最大 8 连通区域作为初始球场区域，但此时得到的初始球场区域内部含有较多空洞部分，需要对其进行填补。采用的填补方法是，找出初始球场区域中每一行位于左右两端的两个边界像素的位置，即 P_L 和 P_R，将 P_L 和 P_R 之间的所有像素点的值都标记为 1，其他标记为 0，从而得到水平填补后的结果。类似地，再对初始球场区域做垂直方向上的填补，然后取这两次填补结果的交集。

4)找到经过填补后的球场区域中最上方和最下方 4 个顶点像素的位置，将这 4 个顶点连接起来构成一个四边形，找出四边形内像素值为 0 的像素点，并将它们的像素值更新为 1，从而得到最终的球场区域。

5)经过以上步骤的处理之后，利用 sobel 边缘检测算子

2 深度图提取

深度图通常用一幅 8 位的灰度图像来表示，图像中取值在[0,255]内的像素值大小反映了真实场景中物体距离观察者或摄像机的远近，0 表示距离观察者最远，255 表示距离观察者最近[23]。

2.1 背景深度图的获取

在羽毛球比赛视频中，全局场景的整体深度分布趋势是自底向顶逐渐增加的，属于水平型分布的多层次场景，即位于同一平面上的物体中距离观察者越近的呈现在二维平面图像中越靠近底部的位置，因此处于同一水平线上的深度值应相同。据此，按照式(1)来构建全局场景下的深度模型，如图 1 所示。

$$depth(I(i,j))=\frac{i-1}{m-1}\times 255,1\leqslant i\leqslant m \tag{1}$$

其中，$depth(I(i,j))$表示在 $m\times n$ 大小的深度模型中位于像素点 $I(i,j)$处的深度值。

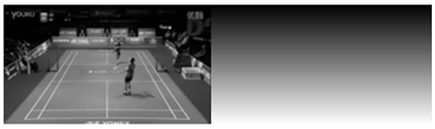


图 1 全局场景图像及对应的深度模型

Fig. 1 Global scene image and corresponding depth model

通常，赛场四周会被一些赞助商的广告牌包围，这些广告牌将整个全局场景划分成了赛场内区域和赛场外区域，是场内区域和场外区域的边界部分。考虑到观众在观看比赛时对这两部分区域的关注度不均衡，为了能够做出更加符合人眼视觉感知的立体效果，我们为场内区域和场外区域单独分配了深度。

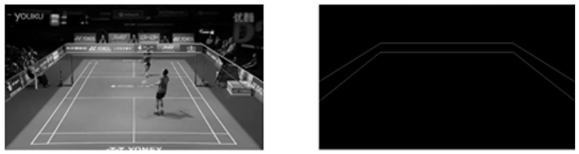
首先，需要找到边界区域的位置，也就是要检测出矩形广告牌的边界线。边界线检测的步骤如下：

1)计算输入图像中每个像素位置处的 RGB 分量值。

2)根据式(2)和式(3)对图像中的每个像素进行阈值化处理，其中 $T=\frac{1}{N}\sum_i\sum_j I(i,j)$是阈值，$N$ 表示输入图像的像素总数。

$$G(i,j)<R(i,j)\ \text{or}\ G(i,j)<B(i,j) \tag{2}$$

以及形态学处理中的膨胀运算对除去球场区域后的余下图像进行边缘检测，获取可供筛选的有用直线段。然后，使用 Hough 变换直线检测来提取包围球场的广告牌区域的直线段，从而找到边界线的位置。以原始视频中的第 5 帧图像为例，得到的检测结果如图 2 所示。



(a)标注后的边界线 (b)边界线二值图像

图 2 检测结果

Fig. 2 Detection results

接着，对赛场内区域和赛场外区域进行深度赋值。图 3 给出了深度分配的原理图。把赛场内区域连同正前方、右侧

及左侧的边界区域依次标注为 A，B，C，D，对应区域的深度赋值方式如下：

$$A: depth(I_A(i,j)) = \frac{255-d}{m-r_{\text{B\_bottom}}} \times (i-r_{\text{B\_bottom}}) + d \quad (4)$$

$$B: depth(I_B(i,j)) = d \quad (5)$$

$$C: depth(I_C(i,j)) = depth(I_A(r_1,j)) \quad (6)$$

$$D: depth(I_D(i,j)) = depth(I_A(r_2,j)) \quad (7)$$

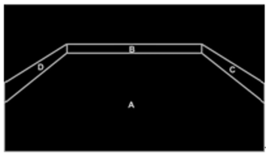


图 3　深度分配的原理图

Fig. 3　Principle diagram of depth value assignment

由于观众对赛场内区域的关注度明显大于对赛场外区域的关注度，为了能够更加突显赛场内区域的立体效果，我们为赛场内的深度分布设置了一个动态取值范围$[d,255]$，d 表示赛场内距离摄像机最远位置处的深度值，对应于图 3 中 B 区域最底端直线段所在位置处的深度；为了保持场内区域和场外区域的深度一致性，B 区域位置后的场外区域的深度分布范围为$[0,d]$。d 的大小可以根据观看效果进行调整，本文中取 $d=35$。式(4)中的 m 表示图像高度，$r_{\text{B\_bottom}}$ 表示图 3 中 B 区域最底端直线段所在的行数，$depth(I_A(i,j))$ 表示 A 区域内坐标(i,j)处像素点的深度值。我们假定边界位置处的广告牌都是垂直于地面的，因此 B 区域中的每个像素点都具有相同的深度值 d。式(5)中的 $depth(I_B(i,j))$ 表示 B 区域内坐标(i,j)处像素点的深度值。式(6)中的 $depth(I_C(i,j))$ 表示 C 区域内坐标(i,j)处像素点的深度值，其大小由 C 区域内与该点列数相同的所有像素点中的最底端像素点(r_1,j)处的深度值 $depth(I_A(r_1,j))$来决定。D 区域内的深度赋值与 C 区域类似。式(7)中的 $depth(I_D(i,j))$表示 D 区域内坐标(i,j)处像素点的深度值，其大小由 D 区域内与该点列数相同的所有像素点中的最底端像素点(r_2,j)处的深度值 $depth(I_A(r_2,j))$决定。据此，可得到包含边界区域的赛场内的深度图，如图 4 所示。

图 4　赛场内(包含边界位置)的深度图

Fig. 4　Depth map inside playfield (including boundary)

采用与式(1)类似的深度赋值方式对场外区域进行深度赋值，可以得到如图 5 所示的赛场外区域的初始深度分布图，其中每一行上的所有像素点处的深度值都相同。考虑到场外区域中还有诸多人物目标存在，为了进一步细化场外区域的深度图，本文采用阈值分割法将这些对象区域划分出来，然后为各个区域分配相应的深度值，文中该阈值取场外区域图像的灰度均值 T_1，计算可得 $T_1=25$。对这些目标进行深度赋值时所采用的两个准则分别为：1)深度一致性，即同一目标区域内的深度值相同；2)每个区域的深度值大小由该区域中的最底端像素在初始深度图中对应位置处的深度值决定。最后，把赛场内区域的深度图与赛场外区域的深度图叠加，便可得到完整的背景深度图，如图 6 所示。

图 5　赛场外区域的初始深度图

Fig. 5　Initial depth map outside playfield

(a)阈值分割后的二值图像　(b)场外区域深度分布图　(c)完整的背景深度图

图 6　背景深度图的获取

Fig. 6　Depth map acquisition of background

2.2　前景深度图的获取

为了能够更加真实地再现原始场景中各个对象在三维空间中的位置关系，本文所指的前景对象除了比赛的运动员外，还包含场内距离摄像机较近的其他对象，例如左右两侧的裁判以及位于右侧裁判下方的木箱。我们使用由 Tang 等[22]提出的改进图割算法(One cut 算法)来获取更加精确的前景对象。

Tang 等首次提出的使用快速全局最优二切分割技术的 One cut 算法是对传统的基于梯度下降方法如 Grab cut[24]算法的改进。Grab cut 算法采用多通道混合高斯模型对图像中的前背景目标像素建模，然后结合用户的初始标记信息来估计混合高斯模型中的各项参数，通过迭代的图切割来达到最优分割。而 One cut 算法通过一次图切割就可实现较好的分割效果，无论是在分割效率还是分割结果上都比 Grab cut 算法更优[25]。One cut 算法中使用的能量函数如式(8)所示：

$$E(S) = |\bar{S} \cap R| - \beta \| \theta^S - \theta^{\bar{S}} \|_{L1} + \lambda |\partial S| \quad (8)$$

其中：

$$|\partial S| = \sum w_{pq} |s_p - s_q|, w_{pq} = \frac{1}{\|p-q\|} \cdot e^{\frac{-\Delta I^2}{2\sigma^2}}, \sigma^2 = \frac{1}{N_{n(p,q)\in n(I)}} \sum \| I_p - I_q \|^2, \Delta I^2 = \| I_p - I_q \|^2, \beta = \frac{|R|}{-\| \theta^R - \theta^{\bar{R}} \|_{L1} + |\Omega|/2} \cdot \beta'$$

其中，S 表示前景像素点的集合，$\bar{S}=\Omega \backslash S$ 表示背景像素点的集合。第一项表示包围盒 R 的一个标准膨胀项。第二项是外观重叠项惩罚项，其中 θ^S 和 $\theta^{\bar{S}}$ 分别表示前景目标 S 和背景 $\bar{S}$ 对应的直方图。第三项是对比敏感惩罚项。$-\beta \| \theta^S - \theta^{\bar{S}} \|_{L1}$ 表示基于 L1 范数的外观重叠惩罚项。$R \subseteq \Omega$ 表示包围盒对应的二元码，$S \subseteq \Omega$ 是一个分割段，$l_s=\{s_p \mid p \in \Omega\}$ 是 $S \subseteq \Omega$ 的特征函数。N_n 指 $n(I)$ 的元素个数，$n(I)$ 则是图像 I 中相邻的像素对集合，β' 是一个全局参数。由于 One cut 算法对矩形框比较敏感，用户可以事先在图像上标注一些种子点来标记前/背景像素点，因此式(8)又可以进一步简化为式(9)所示的能量函数：

$$E_{\text{seeds}}(S) = -\beta \| \theta^S - \theta^{\bar{S}} \|_{L1} + \lambda |\partial S| \quad (9)$$

通过最小化这个能量函数，便可实现最佳的前背景分离。图 7 给出了采用该算法得到的前景分割结果。

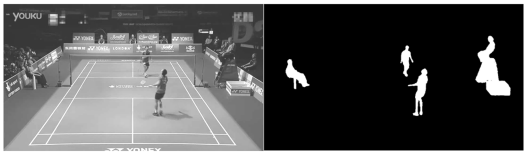


图 7 原始视频中的第 5 帧图像和对应的前景二值图像

Fig. 7 The 5th frame and corresponding foreground binary image in original video

接下来根据式(10)为前景对象进行深度赋值。

$$\begin{cases} d_{f1}=d_{g1}+d_1, & \text{closer player} \\ d_{f2}=d_{g2}+d_2, & \text{farther player} \\ d_{f3}=d_{g3}+d_3, & \text{referees and box} \end{cases} \tag{10}$$

其中，$d_{fi}(i=1,2,3)$表示各前景对象在前景深度图中的深度值，它由两部分组成：1)$d_{gi}(i=1,2,3)$，表示前景对象区域中的最底部像素在赛场内深度图中相同位置处的深度值；2)$d_i(i=1,2,3)$，即为了能够更加凸显出前景对象的立体效果而为其额外分配的深度值。仍然以原始视频中的第 5 帧图像为例，参数 d_1,d_2,d_3 的取值分别为 30，5，15，得到的前景深度图如图 8 所示。

图 8 前景深度图

Fig. 8 Foreground depth map

2.3 前/背景深度融合

根据融合式(11)可以获取完整的深度图像。

$$depth(i,j)=\begin{cases} Bdepth(i,j), & F(i,j)=0 \\ Fdepth(i,j), & \text{otherwise} \end{cases} \tag{11}$$

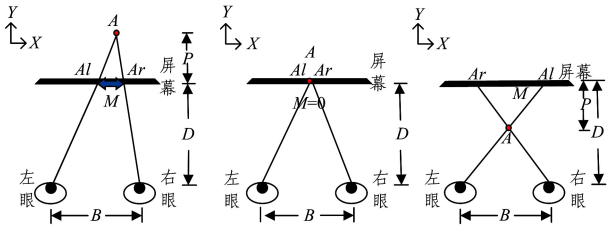
其中，$depth(i,j)$是最终的深度图中位于(i,j)位置处的像素的深度值，$Bdepth(i,j)$ 是背景深度图中位于(i,j)位置处的像素的深度值，$Fdepth(i,j)$是前景深度图中位于(i,j)位置处的像素的深度值。在最终的深度图像中，如果位于(i,j)位置处的像素值不为前景像素点 $F(i,j)$，那么该点处的深度值就为背景深度图中对应位置处的深度值，否则为前景深度图中对应位置处的深度值。输入的 2D 视频帧图像中多种噪声干扰的存在也会影响到深度图的质量，为了能够降低这些干扰，同时又能较好地保留图像中的细节部分，本文采用具有保边去噪特性的双边滤波器来对原始深度图作进一步处理。经过深度融合与双边滤波处理后得到的完整深度图如图 9 所示。

图 9 完整深度图

Fig. 9 Complete depth map

3 基于深度图像的虚拟视点绘制技术(DIBR)

本文采用一种简化的 DIBR 算法[26]来获取用于在 3D 播放设备上显示的虚拟立体图像对。该算法主要包含 3 个关键步骤：1)根据深度值求取视差值；2)根据视差值的大小对原始图像中的像素进行左右平移，得到具有差异的立体图像对；3)修复经过像素平移后的立体图像对中出现的空洞区域[27]。图 10 给出了 3 种不同的视差模型，分别是正视差、零视差和负视差。



(a)正视差 (b)零视差 (c)负视差

图 10 3 种视差模型

Fig. 10 Three different parallax models

图 10 中，B 是双眼间距，D 是观察者与屏幕之间的距离，视差 M 表示左右眼图像之间的水平距离，即 $M=A_r-A_l$，P 表示人眼所能感受到的景物深度。当 A 点的左眼图像 Al 位于右眼图像 Ar 的左侧时，视差 $M>0$，此时观众所看到的景物 A 具有进入屏幕的效果，如图 10(a)所示。根据三角形相似定理，可以得到最大视差 $\max M$ 与最大深度 $\max P$ 之间的关系，如式(12)所示。由于式(12)中的 M,B 等符号的单位均为米制单位，而在图像中则是以像素为最小单位，为了便于后续的图像处理，需要按照式(13)做一个单位转换。其中，$Horizontal\_resolution$ 表示屏幕的水平分辨率，$screen\_width$ 表示屏幕的水平宽度，$\max PM$ 是经过单位转换后的最大正视差。同理，可按照式(14)、式(15)计算得到图 10(c)所示模型中经过单位转换后的最大负视差 $\max NM$，此时观众所看到的景物 A 具有出屏的效果。而在图 10(b)中，$M=0$，观众所看到的对象 A 正好出现在屏幕上。

$$\frac{M}{B}=\frac{Ar-Al}{B}=\frac{P}{P+D}$$

$$\Downarrow \tag{12}$$

$$\max M=\max P\,\frac{B}{\max P+D}$$

$$\max PM=\max M\times\frac{Horizontal\_resolution}{screen\_width} \tag{13}$$

$$\frac{|M|}{B}=\frac{|Ar-Al|}{B}=\frac{P}{D-P}$$

$$\Downarrow \tag{14}$$

$$\max|M|=\max P\,\frac{B}{D-\max P}$$

$$\max NM=-\max|M|\times\frac{Horizontal\_resoltion}{screen\_width} \tag{15}$$

接着，按照式(16)在深度与视差之间做一个映射，并根据深度图获取对应的视差图。

$$parallax(i,j)=\frac{\max NM-\max PM}{255}\times depth(i,j)+\max PM \tag{16}$$

其中，$parallax(i,j)$表示由深度图中(i,j)位置像素处的深度值$depth(i,j)$计算得到的视差值大小。当深度值为 0 时，对应正视差 max PM；当深度值为 255 时，对应负视差 max NM。

最后，对原始视频图像中的每个像素点左右平移视差图中相同位置处视差值的一半，即可获得具有差异的左右眼立体图像对，计算方式如式(17)和式(18)所示。

$$I_{\text{left}}\left(i,j-\frac{parallax(i,j)}{2}\right)=I_{\text{orginal}}(i,j) \tag{17}$$

$$I_{\text{right}}\left(i,j+\frac{parallax(i,j)}{2}\right)=I_{\text{orginal}}(i,j) \tag{18}$$

实际场景中物体之间存在的遮挡、覆盖等现象，使得遮挡物体与被遮挡物体之间的深度值不连续。经过上述步骤的处理之后，由于原始图像中的像素移动的幅度和方向不一样，在左右眼图像中有的像素点位置上没有与其相对应的原始图像的像素点平移到该位置，因此便产生了如图 11 所示的空洞区域。

(a)含有空洞的左眼图像

(b)含有空洞的右眼图像

图 11　含有空洞的图像

Fig. 11　Images with holes

为了提高左右眼立体图像对的质量，以获取较好的立体效果，本文将需要修复的空洞区域划分为 3 种类型，并针对每种类型分别提出对应的空洞修复方案。1)对于分布在球场内呈颗粒状的小面积空洞点，采用中值滤波予以消除；2)对于分布在左眼图像左下端边缘处与右眼图像右下端边缘处的空洞，在确定空洞像素的位置后，用原始图像中相同位置处的像素来填补；3)对于分布在前景对象边缘处的空洞，提出的填补方法是(这里以左眼图像为例)：获取经过中值滤波处理后的左眼图像，由于空洞区域主要分布在前景对象的左侧部分，因此以前景对象左侧轮廓线为起始线，每次以一个像素为单位向左平移扩张，直到平移所经过的区域刚好能把空洞区域内的最后一个空洞像素点包含进去为止，此时计算总共平移的像素个数 L。同样地，在原始图像中以前景运动员左侧轮廓线为起始线，向左平移 L 个像素，截取原始图像中轮廓线在平移中所经过的像素区域来填补空洞区域；对于余下的空洞像素，遍历每个空洞像素点邻域内的像素，如果是背景像素，就用背景像素来填充。右眼图像的填补过程与左眼图像的填补过程类似。经过对以上 3 种类型的空洞区域进行修复处理后得到的左右眼图像如图 12 所示，由此可见，经过修复后的图像的质量得到了大幅提高。

(a)修复后的左眼图像

(b)修复后的右眼图像

图 12　经修复处理后的图像

Fig. 12　Images after restoring treatment

4　实验

选取视频帧大小为 1104×622、帧频为 25 帧/秒的羽毛球比赛视频作为测试视频，实验环境为 Intel Core i5，CPU2.4 GHz，8GB 内存。所使用的 3D 播放设备是屏幕长为 733 mm、水平分辨率为 1366 像素的 32 寸电视。双眼间距 B 取值为 65 mm；通过实验发现，选取景物出屏与入屏比例为 1∶5 时得到的观看效果良好，这里设置观看距离 D 为 500 mm，人眼所能感受到的最大景物深度假定为 max P=100 mm(D 与 max P 所对应的具体数值可以根据这一比例进行更改)。按照式(12)－式(15)可计算得到由米制单位转换为像素单位后的最大正视差 max $PM\approx20$ 和最大负视差 max $NM\approx-30$。为了能够更加突显出赛场内区域的立体效果，我们将赛场内的深度取值映射到$[d,255]$范围内，其中 d 表示赛场内距离摄像机最远位置处的深度值，这样既能够突出球场的立体感，又能保持场内和场外深度的一致性。通过实验发现，当 d 的取值在$[15,40]$范围内时，可以得到较好的观看效果，如果 d 的取值过大，观众会发现球场部分过于突出屏幕，从而显得比较突兀而缺乏真实感；而当 d 的取值过小时，人眼所感受到的立体冲击感则不够强烈(文中展示的实验结果是在参数 d 的取值为 35 时得到的)。另外，在 2.2 节中提到，为了让人眼关注度较高的前景对象具有更加强烈的出屏效果，为前景中的两个运动员以及左右两侧的裁判和木箱都分配了额外的深度值。通过实验发现，这些额外分配的深度值既不能过大也不能过小，过大会不符合实际场景深度，同时会在获取左右眼图像中出现面积较大的空洞问题，过小则出屏效果不够强烈。当为距离镜头较近的运动员分配的额外深度值 d_1 的取值在$(0,30]$内，为左右两侧的裁判以及右侧的木箱额外增加的深度值 d_3 的取值在$(0,15]$内，另一个运动员额外增加的深度值 $d_2<d_3$ 时，可以获得良好的观看效果(文中展示的实验结果是在参数 d_1,d_2,d_3 的取值分别为 30,5,15 时得到的)。利用本文所提出的转换算法对视频中的每帧图像进行处理，限于篇幅，这里仅给出部分帧图像的深度图及经过空洞修复后的左右眼立体图像对，如图 13－图 19 所示。

图 13　第 15 帧图像、对应的深度图及左右眼立体图像对

Fig. 13　The 15th frame, corresponding depth map and stereo image pairs

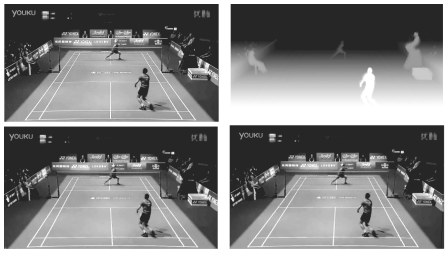


图 14　第 35 帧图像、对应的深度图及左右眼立体图像对

Fig. 14　The 35th frame, corresponding depth map and stereo image pairs

图 15 第 55 帧图像、对应的深度图及左右眼立体图像对

Fig. 15 The 55th frame, corresponding depth map and stereo image pairs

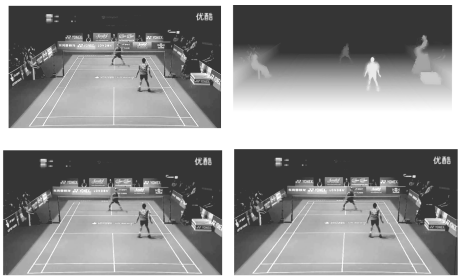


图 16 第 75 帧图像、对应的深度图及左右眼立体图像对

Fig. 16 The 75th frame, corresponding depth map and stereo image pairs

图 17 第 95 帧图像、对应的深度图及左右眼立体图像对

Fig. 17 The 95th frame, corresponding depth map and stereo image pairs

图 18 第 125 帧图像、对应的深度图及左右眼立体图像对

Fig. 18 The 125th frame, corresponding depth map and stereo image pairs

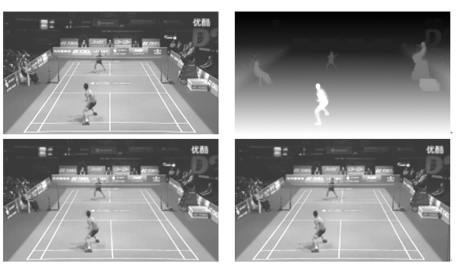


图 19 第 235 帧图像、对应的深度图及左右眼立体图像对

Fig. 19 The 235th frame, corresponding depth map and stereo image pairs

从以上得到的深度图序列可以看出，真实场景的深度变化趋势以及各个对象之间的位置关系都得到了较好的体现。直观上来看，生成的左右眼图像和原始图像的差别不大，图像质量良好。将每帧图像对应的左、右眼图像对序列合成帧频为 25 帧/秒的左右格式视频片段后在 3D 电视上播放，可以看到明显的立体效果。

此外，为了能够定量地评估实验结果，引入了图像的客观质量评价。图像的客观评价是指在分析人眼视觉系统的基础上建立相关模型，然后利用具体的公式计算数值来评价图像质量的好坏。目前，人们应用的客观评价指标主要为均方误差(Mean Squared Error，MSE)和峰值信噪比(Peak Signal to Noise Rate，PSNR)。利用这两个指标对本文算法得到的实验结果进行客观评价。其中，用 MSE 来判断生成的左右眼图像相对于原始视频帧图像的失真程度，数值越大表示失真度越大，然后通过峰值信噪比 PSNR 来估计图像噪声的大小。MSE 与 PSNR 的计算公式如式(19)和式(20)所示。

$$MSE=\frac{1}{M\times N}\sum_{0<i\leqslant M}\sum_{0<j\leqslant N}(f(i,j)-f'(i,j))^2 \tag{19}$$

$$PSNR=10\times\log_{10}\left(\frac{255^2}{MSE}\right) \tag{20}$$

其中，M 和 N 表示图像的宽度和高度，$f(i,j)$ 表示原始图像中 (i,j) 位置处的像素值，$f'(i,j)$ 表示生成的左右眼图像中 (i,j) 位置处的像素值。从原始视频中截取部分帧，按照式(19)和式(20)计算得到的左眼图像和右眼图像的均方误差 *MSE* 和峰值信噪比 PSNR 的大小如表 1 所列。

表 1 原始视频中部分帧图像所对应的 *MSE* 和 *PSNR*

Table 1 Corresponding *MSE* and *PSNR* of some frames in original video

帧序号	左眼图像		右眼图像	
	MSE	*PSNR*	*MSE*	*PSNR*
5	28.0899	33.6453	28.2510	33.6205
15	28.3021	33.6126	28.4746	33.5862
35	28.3731	33.6017	28.5849	33.5694
55	28.2436	33.6216	28.3086	33.6116
75	27.5766	33.7254	27.8414	33.6839
95	27.8381	33.6844	27.9397	33.6686
125	28.4446	33.5908	28.2540	33.6200
155	28.3344	33.6077	28.3826	33.6003
195	27.9126	33.6728	27.9819	33.6620
235	27.9649	33.6647	28.4415	33.5913

从表 1 中的实验结果可以看出，利用本文提出的转换算法所得到的左右眼图像的整体质量较好，这进一步从客观上验证了所提算法的可行性。

结束语 本文提出了一种将羽毛球比赛的 2D 视频转换成 3D 视频的算法，该算法使用了改进的图割算法从背景中较准确地提取出受关注度较高、距离摄像机较近的前景对象。根据场景的深度分布结构构建相应的背景深度模型，然后在此基础上为背景进行深度赋值，以获取背景深度图。为了能够更加突显出前景对象的立体效果，在每个前景对象区域初始深度值的基础上再为其分配额外的深度值从而获取前景深度图。接着，融合前/背景深度图以生成最终的深度图像，借助基于深度图像的虚拟视点绘制技术来获取用于在 3D 播放设备上显示的立体图像对，通过对测试视频进行测试可以观察到较好的立体效果。与此同时，本文算法也为诸如羽毛球

类的比赛视频的 2D 格式到 3D 格式的转换提供了一种可行的思路，后续工作中我们也将继续完善视频中其他镜头的 3D 转换。

参 考 文 献

[1] PATIL S, CHARLES P. Review on 2D to 3D image and video conversion methods[C]// Proceedings of the International Conference on Computing Communication Control and Automation. Paris, France, 2015: 728-732.

[2] EIGEN D, PUHRSCH C. Depth map prediction from a single image using a multi-scale deep network[J]. Computer Science, arXiv: 1406. 2283. 2014: 2366-2374.

[3] DOMINIC J M. Recent trends in 2D to 3D conversion: A Survey [J]. International Journal for Research in Applied Science and Engineering Technology, 2014, 2: 388-395.

[4] JU K, LI Y, XIONG H. Depth propagation with tensor voting for 2D-to-3D video conversion[C]// Proceedings of the International Conference on Acoustics, Speech and Signal Processing. Shanghai, China, 2016: 1696-1700.

[5] TSAI T H, FAN C S. Monocular vision-based depth map extraction method for 2D to 3D video conversion[J]. EURASIP Journal on Image and Video Processing, 2016, 2016(1): 1-12.

[6] JUNG C, CAI J. Superpixel matching-based depth propagation for 2D-to-3D conversion with joint bilateral filtering[C]// Proceedings of the International Conference on Image Processing. Quebec, Canada, 2015: 3515-3519.

[7] YAN X, YANG Y, ER G, et al. Depth map generation for 2D-to-3D conversion by limited user inputs and depth propagation[C]// Proceedings of the 3dtv Conference: the True Vision-Capture, Transmission and Display of 3d Video. Antalya, Turkey, 2011: 1-4.

[8] FAN Y C, CHEN P W, CHEN W S. Low computing complexity architecture design of 2D-to-3D image converter[C]// Proceedings of the IEEE International Conference on Consumer Electronics. Taiwan, 2014: 99-100.

[9] HAN K, HONG K. Geometric and texture cue based depth-map estimation for 2D to 3D image conversion[C]// Proceedings of the IEEE International Conference on Consumer Electronics. Berlin, 2011: 651-652.

[10] CHEN Y C, WU Y C, LIU C H, et al. Depth map generation based on depth from focus[C]// Electronic Devices, Systems and Applications. IEEE, 2010: 59-63.

[11] LIN J, JI X, XU W, et al. Absolute depth estimation from a single defocused image[J]. IEEE Transactions on Image Processing, 2013, 22(11): 4545-4550.

[12] YONG J J, BAIK A, PARK D. A novel 2D-to-3D conversion technique based on relative height-depth cue[J]. Proc Spie, 2009, 7237: 72371U-72371U-8.

[13] LAI Y K, LAI Y F, CHEN Y C. An Effective Hybrid Depth-Generation Algorithm for 2D-to-3D Conversion in 3D Displays [J]. Journal of Display Technology, 2013, 9(3): 154-161.

[14] SAXENA A, SCHULTE J, NG A Y. Depth estimation using monocular and stereo cues[C]// Proceedings of the International Joint Conference on Artifical Intelligence. Hyderabad, India, 2007: 2197-2203.

[15] MOHAGHEGH H, SAMAVI S, KARIMI N, et al. Depth eatimation from single images using modified stacked generalization [C]// Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing. Shanghai, China, 2016: 1621-1625.

[16] KONRAD J, WANG M, ISHWAR P, et al. Learning-Based, Automatic 2D-to-3D image and video conversion[J]. IEEE Transactions on Image Processing, 2013, 22(9): 3485-3496.

[17] JI P, WANG L, LI D X, et al. A novel 2D-to-3D conversion method based on blocks world[C]// Proceedings of the IEEE International Conference on Audio, Language and Image Processing. 2012: 541-543.

[18] CHENG C C, LI C T, CHEN L G. A novel 2D-to-3D conversion system using edge information[J]. IEEE Transactions on Consumer Electronics, 2010, 56(3): 1739-1745.

[19] CHENG C C, LI C T, HUANG P S, et al. A block-based 2D-to-3D conversion system with bilateral filter[C]// Proceedings of the International Conference on Consumer Electronics. Kyoto, Japan, 2009: 1-2.

[20] FREILING B, SCHUMANN T, LAI Y K, et al. Block based depth map estimation algorithm for 2D-to-3D conversion[C]// Proceedings of the IEEE International Symposium on Consumer Electronics. Hsinchu, Taiwan, 2013: 53-54.

[21] NAM S W, KIM H S, BAN Y J, et al. Real-Time 2D-to-3D conversion for 3DTV using time-coherent depth-map generation method[J]. International Journal of Contents, 2014, 10(3): 187-188.

[22] TANG M, GORELICK L, VEKSLER O, et al. GrabCut in one cut[C]// Proceedings of the International Conference on Computer Vision. Sydney, Australia, 2013: 1-8.

[23] YIN S, DONG H, JIANG G, et al. A novel 2D-to-3D video conversion method using time-coherent depth maps[J]. Sensors, 2015, 15(7): 15246-15264.

[24] ROTHER C, KOLMOGOROV V, BLAKE A. GrabCut: interactive foreground extraction using iterated graph cuts[J]. Acm Transactions on Graphics, 2004, 23(3): 309-314.

[25] CHEN X, LI J. Summary of Segmentation Algorithm Based on Graph Theory[J]. Computer and Digital Engineering, 2016, 44(10): 2043-2047. (in Chinese)
陈杏，李军. 基于图论的分割算法研究综述[J]. 计算机与数字工程，2016，44(10)：2043-2047.

[26] FEHN C. Depth-image-based rendering (DIBR), compression, and transmission for a new approach on 3D-TV[C]// Proceedings of SPIE-The International Society for Optical Engineering. 2004: 93-104.

[27] ZHU C, ZHAO Y, YU L, et al. 3D-TV System with Depth-Image-Based Rendering: Architectures, Techniques and Challenges [M]. New York: Springer, 2013.