

A Superpixel-Based Voting Scheme For Removing Moving Objects In Photos

Su Tao^{1,a,*}

¹College of Computer Science, Beihang University, Beijing, China

^a237111904@qq.com

*Corresponding author

Keywords: Superpixel, Voting scheme, Moving object, Detection and removing.

Abstract. When photographing scenic photos, there will be some tourists, cars or other moving objects appearing in photos, which will affect the effect to some extent. Although inpainting technology provides a solution to remove unwanted objects in a single image, it requires a lot of interactions to indicate unwanted objects, and in some cases, there are still some noticeable artifacts in results. With development of smartphones and cameras, people always take a series of photos in almost the same location to capture a good picture (without other unwanted objects). In this paper, a superpixel-based voting scheme is introduced to remove unwanted moving objects in a photo with multiple related photos. First, we adopt image registration to align the image sequence. User can chose one of these photos as the target image, and the portrait are segmented from background of these photos. Then, superpixel blocks are generated in the target background, and we propose the superpixel-based voting algorithm to detect and remove moving objects automatically. At last, an interactive operation is provided to refine the result. Experimental results show that our method can remove moving objects in photos effectively.

一种基于超像素投票模型的运动物体去除方法

苏涛^{1,a,*}

¹北京航空航天大学计算机系, 北京, 中国

^a237111904@qq.com

*通讯作者

关键词: 超像素; 投票模型; 运动物体; 检测与去除

中文摘要. 在拍摄人物风景照时, 特别是一些著名的景点, 总会有其他游客或者汽车之类的物体出现在照片中, 这在一定程度上影响了照片的效果。虽然抠图以及图像补全技术可以去除照片中多余物体, 但是其需要繁杂的交互操作, 并且效果不尽人意。随着智能手机和相机的发展, 人们会在几乎同一位置拍摄的一组照片, 本文则利用了照片序列的相似特征, 提出了一种基于超像素的投票方案, 实现了目标照片中运动物体的去除以及背景恢复。首先, 对照片序列进行图像配准, 用户选择一张需要修缮的照片作为目标图像, 并对目标图像进行前景人物与背景的分隔。然后, 在目标背景图像上生成超像素块, 并采用基于超像素的投票算法自动检测并去除目标图像中运动物体。最后, 本文提供了交互式操作来完善结果。实验结果表明, 本文的方法能有效地去除图像中的运动物体。

1. 引言

现实生活中，经常会遇到这样的问题：游客会在某一地点拍摄多张照片，其目的是为了从中筛选出一张效果最佳的照片，即使这样，也会出现诸如以下的问题：有其他游客、车辆或别的物体出现在照片背景中，照片中人物表情不好等等，这在一定程度上影响了照片的效果。因此，从多张照片中去除背景中的多余对象，从而得到效果最佳的照片，是非常必要的。

去除照片中的运动物体在图像处理领域具有挑战性，主要难点在于如何自动地检测运动物体以及如何准确地将检测到的物体去除。目前主流的方法是先将要去除物体人为标识出来，然后利用纹理合成^[1]、图像补全^[2-7]等技术完成图像恢复。然而这些方法存在如下缺陷：交互操作繁杂；对于复杂图像处理效果不佳；图像恢复的效果不够真实，有明显的修补痕迹。所以，本文利用照片序列信息的相关性，提出了一种基于超像素的投票算法，可以自动检测并去除目标图像中的运动物体。首先要对图像序列进行配准，并利用图割算法^[8]对用户选定的图像（目标图像）进行前景人物与背景的分离，其次，对分离后的目标背景图像进行超像素分割，产生超像素块，并将改数据结构复制给其余图像（候选图像）；然后通过投票算法将目标背景图像中的检测到的超像素块替换为候选图像中相应的超像素块，并将目标背景图像与前景人物图像进行融合；最后本文提供了一种交互式操作来完善结果。实验结果表明，本文的方法能有效地去除图像中的运动物体。

2. 准备工作

图像中运动物体的检测可以大致分为两类：人工检测和自动检测。人工检测方法需要用户提取运动的对象，这种方法的实现简单，然而它的适应性并不强，它需要很多复杂的交互操作，不适用于处理复杂场景的图像。而自动检测方法可以产生准确真实的结果，并且该方法不需要用户进行复杂的交互操作。目前对静态相机的研究已经非常广泛，并且很多相关算法被先后提出，如自适应背景建模方法、统计背景建模方法^[9]以及光流技术^[10]等。本文提出的方法也属于此类范畴，本文使用基于超像素的投票算法自动检测运动物体，并且产生了准确的检测结果。

在完成图像中与运动物体的检测后，我们的下一个任务就是图像恢复。图像完成方法可以分为基于结构的图像恢复和基于纹理的图像恢复两种方法。其中基于结构的图像恢复技术从边界向待恢复区域的内部平滑进行图像恢复，该方法利用待恢复区域边界信息来填充待恢复区域内部信息。虽然这类方法在图像待恢复区域较小的时候可以产生比较真实的恢复效果，但是它却无法处理待恢复区域面积较大的图像。而基于纹理的图像恢复技术可以实现对图像中较大的区域的进行恢复。这类方法通过在图像中寻找与待恢复区域特征信息相近的块来填补了待恢复区域的信息，需要进行大量的块匹配运算。相比基于结构的方法，基于纹理的方法可以产生更真实的效果，尤其是在待恢复区域面积较大的场景中。然而，对于基于纹理的图像恢复技术，寻找匹配块的计算量庞大，耗时长，是该方法的瓶颈。Criminisi et al.^[2]的论文中介绍了一种利用纹理合成进行图像恢复的方法，该方法表示可以通过利用局部图像结构来优化图像信息补全的顺序，从而达到减小计算量的目的。而对于包含复杂场景和有大面积待恢复区域的图像，相邻的局部区域没有足够的信息去完成准确的匹配。Huang et al.^[11]在论文中指出可以利用图像中的平面信息以及规则化去完成匹配。Sun et al.^[12]的论文提出了一种图像恢复方法，允许让用户指出结构化的信息。Whyte et al.^[13]在论文中提出了利用SIFT特征点进行单应性分组，并对分割区域进行不同的单应性变换，因为将图像自动分割为多个区域是很难的，所以他们的的方法需要用户干预才能获得正确的分割，并不能自动完成图像恢复。本文的方法采用超像素分割方法可以自动、准确地进行图像分割，然后将目标图像中检测到的超像素块替换为候选图像中相应的超像素块，因此本文的方法可以更加有效地产生真实的结果。

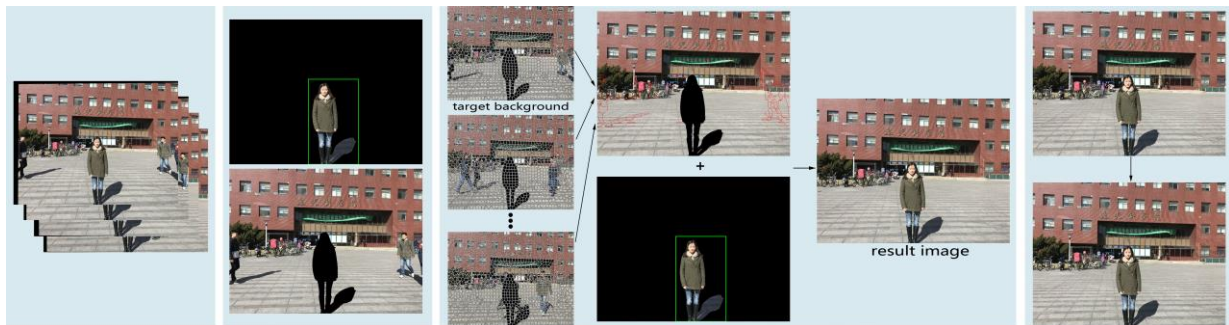
3. 基于超像素的投票模型

在几乎相同的位置拍摄一系列人物风景照片作为模型的输入，如果照片中有其它影响照片效果的物体（如游客或者车辆等），就会在照片序列中留下相应的痕迹。因此，我们假设这些运动物体在照片序列中的位置都有所不同。

对于给定的图像序列 I_1, I_2, \dots, I_n ，我们可以通过以下步骤得到去除运动物体后的结果图像。

- 1) 对图像序列进行配准；
- 2) 对目标图像 I_t 以及其它候选图像进行前景人物与背景的分离；
 - a) 用户选择一张照片作为目标图像 I_t ，并将其分割为前景图像 F_t 和背景图像 B_t ；
 - b) 对其余候选图像 $I_{c1}, I_{c2}, \dots, I_{cn-1}$ 进行如上操作，得到前景图像 F_{ci} 和背景图像 B_{ci} ；
- 3) 检测并去除目标图像中的运动物体；
 - a) 对目标背景图像 B_t 进行超像素分割，产生超像素集 S_t ；
 - b) 将 S_t 复制到其余候选背景图像 B_{ci} ，产生相应的超像素集 S_{ci} ；
 - c) 利用基于超像素的投票算法，对背景图像序列进行运动物体的检测和去除；
 - d) 将目标前景图像 F_t 与结果背景图像融合；
- 4) 用户对图像进行完善。

如图1所示，本文的方法流程主要有四个步骤。首先，提取图像序列的SIFT特征，利用基于SIFT的图像配准技术^[14]对图像序列进行对齐操作。其次，选择一张图像作为目标图像，利用图割算法对目标图像进行前景人物与背景的分离，并采用模板匹配算法^[15]对其余候选图像进行前、背景分离。然后，本文提出基于超像素的投票算法，来自动检测并去除图像中的运动物体。其中步骤3.a和3.b产生了目标图像超像素集和候选图像超像素集，步骤3.c通过投票算法将目标图像中包含运动物体信息的超像素替换为候选图像中相应的超像素。最后，将之前的目标前景图像与去除了运动物体的背景图像融合即可得到结果图像。本文最后提供了一种对图像进行完善的操作。



(1) 图像配准 (2) 图像分割 (3) 运动物体检测与去除 (4) 用户辅助操作

图1 基于超像素投票算法的基本步骤

4. 运动物体的自动检测与去除

本文采用基于超像素块的投票算法来对运动物体进行检测和去除。基于块的投票算法比基于像素的投票算法计算更加高效与合理，也可以减少图像配准所产生的错误和图像噪声。然而基于超像素块的投票算法存在两个问题，一个是如何生成目标背景的超像素集以及如何候选背景中生成相应的超像素集。另一个是要定义什么样的图像特征，以及如何利用这些特征进行投票算法。下面将详细介绍基于超像素的投票算法。

4.1 背景图像的超像素生成算法

超像素分割算法^[16]可以对图像像素进行局部聚类，该算法可以产生形状近似均匀、颜色一致的超像素块，因此本文选择用超像素块结构来替代普通的块结构。

首先对分离后的目标背景图像 B_i 进行超像素分割，先利用中值滤波算法^[17]对图像进行平滑操作，然后用简单聚类迭代线性算法（SLIC）^[18]产生 M 个超像素块（ $S_{t1}, S_{t2}, \dots, S_{tm}$ ），其中超像素初始边长为 α ，本文中 α 的取值为20像素。然后按照目标背景图像 B_i 的超像素结构，对候选背景图像中的像素进行聚类。

4.2 基于超像素的投票算法

基于超像素的投票算法用于比较两张图像中相对应的超像素的相似程度，从图像序列中找出得票数最多的超像素，也就是与其它超像素相似度最大的超像素，来替换掉目标背景图像中的相应超像素。两张图像中相应超像素 S_{ti}, S_{ci} 的相似性计算公式如下所示：

$$Sim(S_{ti}, S_{ci}) = \frac{C}{m} \quad (1)$$

其中 m 是超像素中的像素总数（两个超像素中的像素总数相等）， C 为两个超像素中的相似像素数目。如果相似度 Sim 大于一个阈值，就认为超像素 S_{ti} 和 S_{ci} 相似。

然后，图像序列中相对应的超像素两两计算相似度，统计出每个超像素获得的票数，并选择得票数最多的超像素去替换目标背景图像中相应的超像素。如果有多个超像素获得了最大票数，那么就对这些超像素进行平滑度 Smo 计算，计算公式如下所示：

$$Smo(S_c) = -(\sum |r(i, j) - avgr(S_c)| + \sum |g(i, j) - avgg(S_c)| + \sum |b(i, j) - avgb(S_c)|) \quad (2)$$

其中 $r(i, j), g(i, j), b(i, j)$ 分别为像素点 $P(i, j)$ 的颜色值分量，而 $avgr(S_c), avgg(S_c), avgb(S_c)$ 分别为该超像素的平均颜色值分量。由于超像素块几乎是颜色一致的，所以我们选择平滑度较高的超像素去替换目标背景图像中相应的超像素。

4.3 用户辅助操作

在完成目标背景图像中的超像素块的替换后，我们将之前的前景图像与之融合，即可得到去除了运动物体的结果图像。然而，本文还给用户提供了辅助操作，可以帮助用户完善照片。如图2所示，左边是去除了运动物体后的结果图，图中左下角有黑影，需要用户去除，右图是通过辅助操作去掉黑影后的图像。我们为用户提供了一种辅助操作，用户选定该区域后，通过对比该区域在各个图像中的相似度，并采用投票算法，选出得票数最多的区域去替换当前区域。然而，如果当前区域的得票数最多，我们采用图像修复算法^[19]，利用该区域周围的纹理信息去填补该区域。

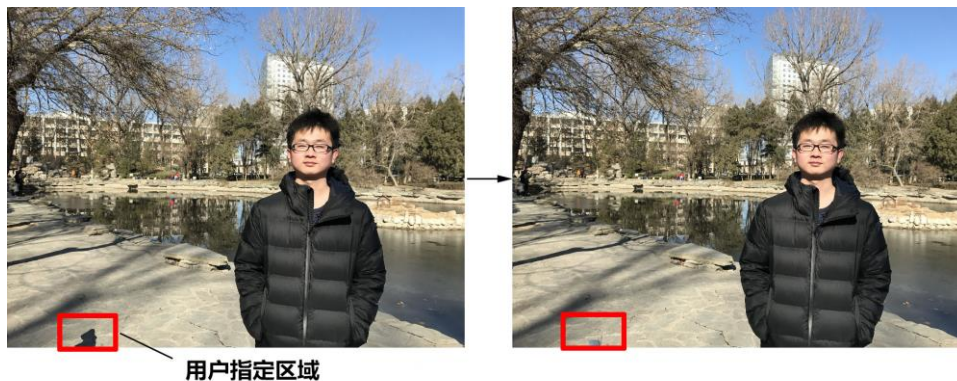


图2 用户辅助操作示意图

5. 实验与结果分析

本文所有的实验性能指标都在如下的平台上进行的：Inter Core i7 3.4GHz的CPU，16GB的内存。实验使用OpenCV开发环境，在不到一分钟的运行时间，可以完成1200×900（其中托雷多城的图像为820×1200）像素的图像处理工作。所有实验中输入的图像序列包含4幅图。

5.1 实验结果及效果对比

图3所示为部分实验结果图，其中左边为输入的图像序列（4幅），中间为用户选择的目标图像，我们用红色矩形框标出了不想要的物体，右边是自动检测并去除运动物体后的结果图像。其中前两行图片的场景分别为鸟巢和西班牙的托雷多城，它们都具有简单空旷的背景以及流动的人群或车辆，甚至有的行人比照片中的前景人物还要大，通过本文的方法，目标图像中的运动物体都被完全地去除了。第三行的树林场景图片包含了复杂的背景以及细小的树枝，从结果图像中可以看出本文的方法可以很好地处理包含复杂背景的图像。在第四行的操场场景中，存在有许多运动物体的阴影，从图中可以看出，我们的方法在处理阴影时与处理运动物体一致，都能够达到真实良好的效果。

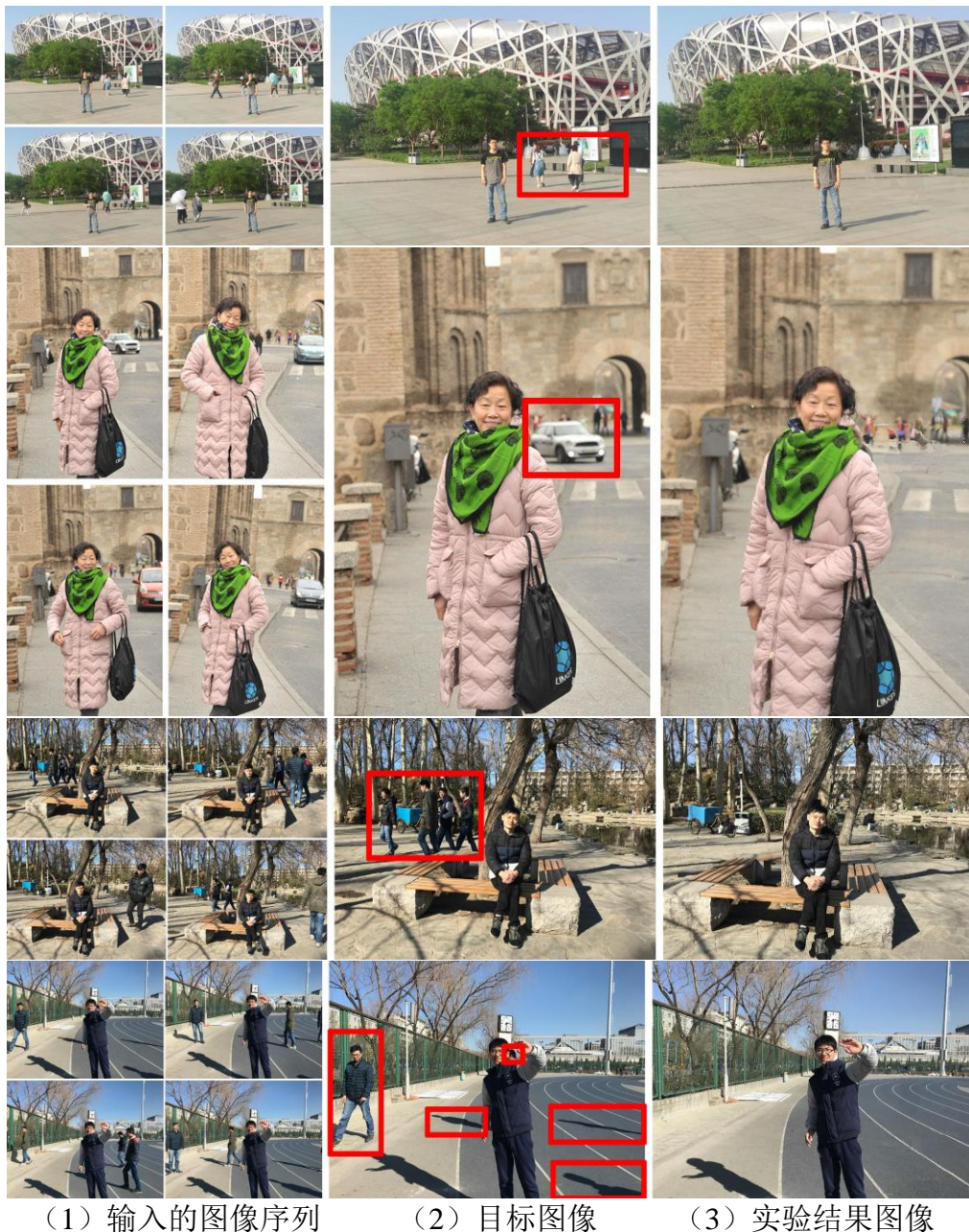


图3 我们的方法所产生的实验结果

本文将实验结果与Criminisi的图像补全算法^[2]以及传统的基于像素的投票算法进行了对比。如图4所示，左边是用户选择的目标图像，之后分别为本文的方法，Criminisi的方法以及基于像素的投票方法的实验结果。可以看出，Criminisi的方法产生的结果图像中会出现断层现象（图中用黄色矩形框标出），这是由于该方法只是在目标图像中找到了相似的背景信息，而这些图像信息并不是真实属于该区域的，一旦图像中存在大面积的区域需要恢复，该方法很容易产生不真实的结果。与Criminisi相比，本文提出的方法可以自动检测图像序列中的运动物体，并利用候选图像中相应的背景信息去恢复目标图像的背景信息，所以本文的方法能更有效地产生真实的结果。

而基于像素的投票方法是对单个像素进行投票，容易产生局部误差。从图中可以看出，基于像素的投票方法会在结果中残留下点状物体（图中用黄色矩形框标出），如鸟巢场景中扔残留游客的一部分身体，托雷多城中汽车的一部分，其它场景中也可能存在一些黑点残留物。与基于像素的投票方法相比，本文所提出的方法在检测和去除运动物体时都是基于超像素块的，因此我们的方法具有更好的鲁棒性，在不同的场景下都会产生更为真实的结果。



(1) 目标图像 (2) 我们的方法 (3) Criminisi的方法 (4) 基于像素的投票方法

图4 实验结果对比图

5.2 参数选择

本文提出的方法包含一个参数 α ， α 的取值会影响该方法产生的效果。 α 超像素块的初始边长，如图5所示，第一行为 α 的三种不同取值所产生的超像素分割图从左到右 α 的取值依次为：10，20，30（单位：像素），第二行为 α 不同的取值所产生的结果图像。可以看出，当 α 取值为10时，结果图像中仍存在小的残留物（图中用黄色矩形框标出），这是因为超像素块的面

积太小，与基于像素的投票方法类似，在计算时会产生一些局部误差，另外，超像素面积越小，计算复杂度就越高，这还会增加算法的计算量以及降低算法效率。然而，当 α 取值为30时，可以看出，结果图中残留有人的头发（图中用黄色矩形框标出），这是由于超像素块的面积过大，一个超像素可能会覆盖多个小的运动物体，如果这个超像素得票数最多，那么就会替换掉原来的超像素，导致结果图中仍残留些许运动物体信息。经过试验发现当 α 取值为20时，所产生的结果图像效果最好。

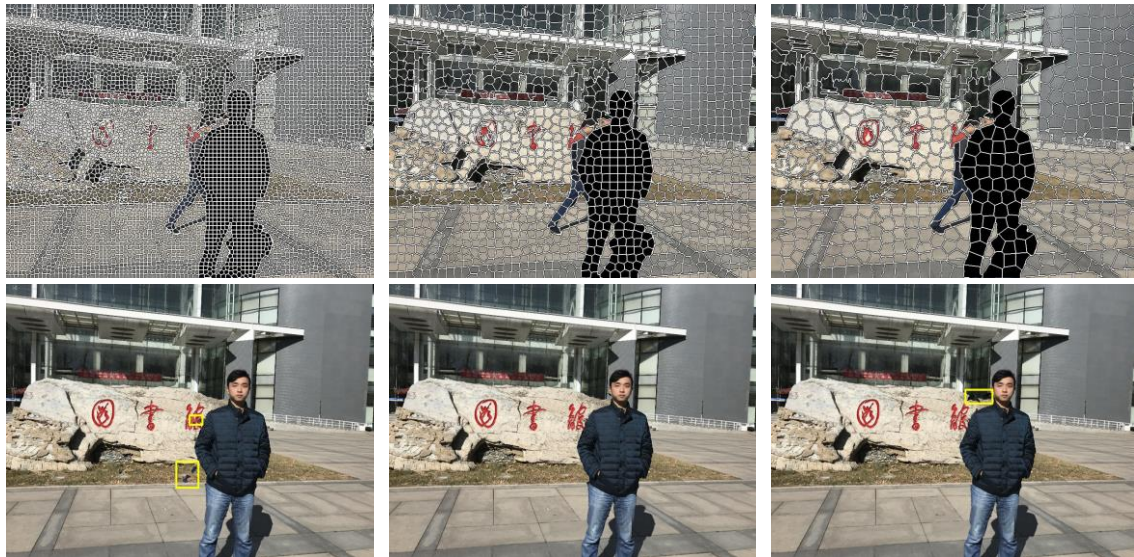


图5 不同的参数 α 产生的结果对比

5.3 局限性

由于本文所提出的基于超像素的投票算法是有假设前提的，也就是拍摄的照片序列几乎在同一位置，因此，本文的方法可能不适用于照片中存在静止的行人或车辆的场景，这是因为该方法会将这些静止的行人和车辆当作背景处理。所以对于上述提到的场景，本文提出的方法可能会产生不明显的效果。

另外，在对目标图像进行前景人物与背景分离的过程中，本文使用了交互式图割方法，为了减少用户的操作，应该对该算法进行改进，以达到全自动图像分割的效果。

6. 结论和未来工作

本文提出了一种基于超像素的投票方法，用以自动检测并去除照片序列中运动物体。此方法的输入为用相机或手机拍摄的一组照片，用户选择出需要修复的照片作为目标图像。该方法首先对图像序列进行配准，前、背景分割的预处理，然后，利用基于超像素的投票算法实现运动物体的自动检测和去除，用得票数最多的超像素去替换目标图像中相应的超像素，最后将前景人物与经过处理后的目标背景图像融合。本文还为用户提供辅助操作，帮助用户完善照片。实验结果表明，相比Criminisi的方法以及基于像素的投票方法，我们的方法具有更好的鲁棒性以及更真实的效果。

为了进一步提升图像效果，接下来我们将结合深度学习方法去做进一步探索。本文的方法在进行图像前、背景分割时还需要一些人为交互操作，在之后的实验中，我们会向全自动图像分割方向做进一步研究。

References

- [1] Bertalmio, M. and Vese, L. and Sapiro, G. and Osher, S., Simultaneous Structure and Texture Image Inpainting, *IEEE*, pp. 882-889, 2003.
- [2] Criminisi, A. and Prez, P. and Toyama, K., Object Removal by Exemplar-Based Inpainting, vol.2, pp. 292-311, 2003.
- [3] Bertalmio, M., Bertozzi, A. L., & Sapiro, G. Navier-stokes, fluid dynamics, and image and video inpainting. *Computer Vision and Pattern Recognition*, Vol.1, pp.I-355-I-362, 2001.
- [4] Ružić, T., & Pižurica, A. Context-aware patch-based image inpainting using markov random field modeling. *IEEE Transactions on Image Processing*, vol.24, pp. 444-456, 2015.
- [5] Wexler, Y., Shechtman, E., & Irani, M. Space-time video completion. *Computer Vision and Pattern Recognition*, Vol.1, pp.I-120- I-127, 2004.
- [6] Zhu, Z., Huang, H. Z., Tan, Z. P., Xu, K., & Hu, S. M. Faithful completion of images of scenic landmarks using internet images. *IEEE Transactions on Visualization & Computer Graphics*, 22(8), 1945-1958, 2016.
- [7] Granados, M, et al. How Not to Be Seen — Object Removal from Videos of Crowded Scenes. *Computer Graphics Forum*, pp: 219-228, 2012.
- [8] Kwatra, Vivek, et al. Graphcut textures:image and video synthesis using graph cuts. *ACM Transactions on Graphics (TOG)*, pp: 277-286. 2003.
- [9] Stauffer, Chris, and W. E. L. Grimson. Adaptive Background Mixture Models for Real-Time Tracking. *IEEE Computer Society*, pp: 2246, 1999.
- [10] Barron, J. L., and D. J. Fleet. Beauchemin Performance of optical flow techniques In *International Journal of Computer Vision*. 1996.
- [11] Huang, Jia Bin, et al. Image completion using planar structure guidance. *Acm Transactions on Graphics*, pp: 1-10. 2014.
- [12] Sun, Jian, et al. Image completion with structure propagation. *Acm Siggraph*, pp:861-868,2005.
- [13] Whyte, Oliver, J. Sivic, and A. Zisserman. Get Out of my Picture! Internet-based Inpainting. *British Machine Vision Conference*, pp: 7-10, 2009.
- [14] Lowe, David G. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision* vol.60, pp: 91-110, 2004.
- [15] Berg, Alexander C., and J. Malik. Geometric Blur for Template Matching. *IEEE Computer Society Conference on Computer Vision & Pattern Recognition*, pp: 607, 2001.
- [16] Malik, J. Learning a classification model for segmentation. *IEEE International Conference on Computer Vision*, vol.1, pp:10-17, 2003.
- [17] Huang, T., G. Yang, and G. Tang. A fast two-dimensional median filtering algorithm. *IEEE Transactions on Acoustics Speech & Signal Processing*, vol.27.1, pp: 13-18, 1979.
- [18] Achanta, Radhakrishna, et al. SLIC superpixels. *Epl*, 2010.
- [19] Alexandru Telea. An Image Inpainting Technique Based on the Fast Marching Method. *Journal of Graphics Tools*, vol.9.1, pp: 23-34, 2004.