Cultural Heritage Reconstruction System Based On Structure from Motion

Fei Deng, Zilong Zhao, Duanshun Li, Yulei Hu
School of Geodesy and Geomatics, Wuhan University,
430079, Wuhan, China fdeng@sgg.whu.edu.cn

ABSTRACT: We present a pipeline to reconstruct the 3D structure and color from underground image sets and to present the heritage with an interactive system. Our approach employs a Structure from motion (SfM) algorithm to recover the pose of the cameras which will be used to reconstruct the 3D structure of the scene and to render the images in a visual reality system based on state-of-the-art image based rendering techniques. Then the polished pose of cameras and the coordinates of 3D points from triangulation through bundle adjustment are used for the representation of the relics. The contribution of our work is that we never treat the reconstructed scene and the images separately as approaches to record those heritages, on the contrary, we combine them together to take the great advantage of the 3D information provided by the reconstructed point clouds and the textured 3D models and the more accurate color and texture information provided by the images simultaneously.

KEY WORDS: cultural heritage, reconstruction, image based rendering, combined

1 Introduction

Large amounts of cultural heritages have been facing damages due to aging, pollution (acid rain, haze, dust and etc.), unpredictable natural catastrophes (earthquake, flood and etc.) and increasing human activities (tourism, war and etc.). Specifically, the magnificent Dunhuang frescoes have been threatened by the dry climate and the desert, and the Great Buddha of Bamiyan were even totally damaged in the war. Though various measures have been taken to protect and restore those treasures, the vanishing of those precious relics is inevitable. To insure the descendants of human could share the heritages with us, researchers have realized the importance to record those relics for our descendants, modern techniques such as digital image system (Ren, 2005, Ke, 2006, Gruen, 2002, Remondino, 2009) and LiDAR (zhao, 2008) are used for digitalization of vanishing relics.

The Digital Michelangelo project (Levry, 2000) and the Digital Dunhuang project (Liu, 2005) are best examples for the use of scanning. However, a big disadvantage of this approach is the lack of color and texture information. Despite from that, the use of LiDAR to digitalize heritages is highly debatable because such a technique may bring harm to precious relics due to its high energy shooting on the surface of the relics directly. Comparatively, images has becoming the most popular approach to record those precious relics such as ancient books, catastrophes, buildings, statues, sculptures and so on, though most of the usages merely use images as a recording of the relics and go no further. Though 3D reconstruction and modeling from images has been deeply researched, few works has been done for the application of those techniques to protect cultural heritages. (Gruen, 2002) take the first step in reconstruction the Great Buddha of Bamiyan through an image collection from internet. Still there is an unignorable drawback of this approach, because the dense of the point clouds is limited due to various kinds of reasons such as mismatch which will be very common in the application of smooth statues with similar photometric responses.

In our work, we adopt the SfM method proposed in (Snavely, 2006) which choose the first pair of images with wide baseline and largest number of matches and add the other images successively to recover the pose of the cameras and reconstruct the scene sparsely. Following that, to provide much denser point clouds, we employ the patch based dense reconstruction method taking epipolar constraint into consideration presented in (Furukawa, 2007). Our pipeline then adopts a state-of-the-art image based rendering system which is similar to the rendering mechanism of Photo Tourism introduced by Snavely to visualize the images combined with the reconstructed point clouds. The system provides the users several interactions to view the images both from a global view and at the place where they were captured. We apply this technique several data sets including a relief, a Chinese folk style building, a cave and the exterior of a grotto.

2 Pose estimation and 3d reconstruction

The first stage of our Pipline is to recover the pose of the images. This stage follows the basic mage based reconstruction strategy. To get the correspondences, we choose to detect SIFT (Lowe, 2004) features from each image captured about the scene due to its scale invariant ability and its excellent performance in a circumstances with varying illumination or view point which are quite common while capturing images of different kinds of heritages located in different locations such as big palace, little dark caves. The 128-dimensional descriptors extracted along with the key points will be used for the following pairwise matching of those features. The result is refined by taking epipolar geometry into consideration in a RANSAC (Fischler, 1981) frame work. Figure 1 provide a pipeline of the feature detection and matching of 3 images.

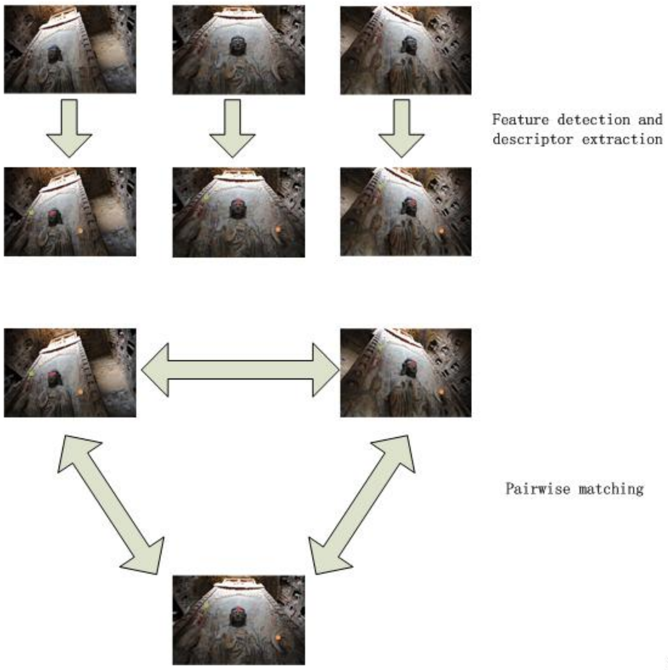


Figure 1. Feature detection and matching.

Once we get the correspondences among the image sets, we recover the pose of the images with the SfM algorithm which is illustrated in Figure 2. Firstly, we choose a pair of images to establish a local coordinate system and initialize the scene by estimating the relative orientation of these two images and triangulate the correspondence of the image pair to get initial 3D points. To establish the robustness of the algorithm, we choose the first pair of wide baseline and largest number of correspondences which ensures that we have enough points with high accuracy to start successive algorithms and the initial point cloud covers a large proportion of the scene. By compute the mean error of the estimated homography referred as Homographic Index (HI) through correspondence, we can evaluate the quality of the initial image pair since a minor mean error indicate that either the images are captured at nearly the same place which will bring large error in triangulation or the initial scene is a quasi-plane which is harmful for the estimation of successive images.

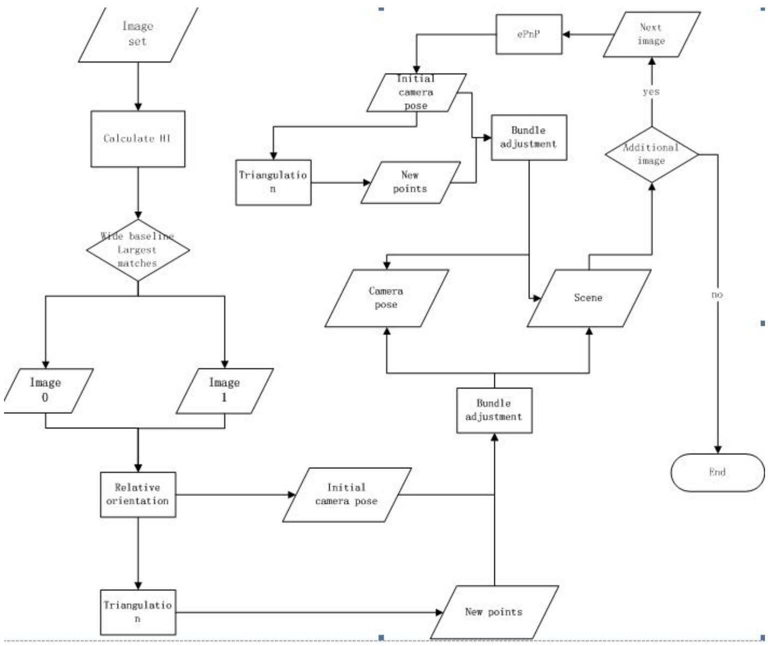
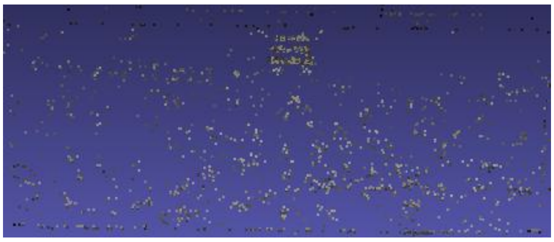


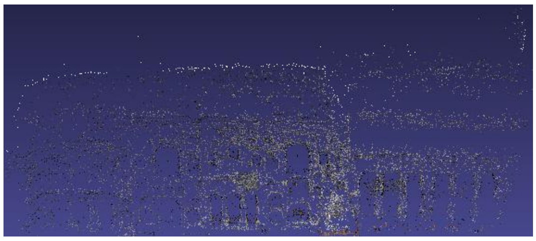
Figure 2 Pipeline of the SfM algorithm

While the first pair of image and the scene is initialized successfully, we add the other images successively into the scene according to the number of reconstructed points visible in the image. The direct linear transformation (DLT, Hartley, 2003) algorithm is usually used for the initialization of the pose of images, however, due to over parameterization, this approach may encounter singularity while processing a quasi-plane scene. Since prospective n point (PnP) algorithms with less correspondences are much more stable than DLT, we employ the efficient perspective n point (ePnP) algorithm which proposed by (Noguer, 2007) to estimate the initial pose of those images. After the pose of the camera is refined through the tie points with a least square optimization, additional 3D points are triangulated and followed by bundle adjustment which has been the golden global optimization technique in SfM.

Bundle adjustment was first introduced to refine a visual reconstruction to produce joint optimal 3D structure and viewing parameters including camera pose and calibration parameters. The bundle adjustment parameter space is a high-dimensional nonlinear manifold, which is made up by a large Cartesian product of projective 3D feature, 3D rotation and camera calibration manifolds. The cost function is expressed as the sum of squares of a large number of nonlinear, real-valued functions. To accomplish such a time consuming optimization task, various kinds of techniques have been introduced, such as preconditioned conjugate gradient () and parallel newton method with GPU acceleration. One of the most popular techniques is to take great advantage of the sparse characteristic of this problem. (Lourakis, 2009) develops a sparse bundle adjustment (SBA) algorithm which utilizes the sparse geometry of this problem to solve the camera pose and 3D structure separately. SBA solve the camera pose first with the Hessian first, to reduce calculation cost, 3D coordinates of the triangulated points are then addressed with the camera pose deduced. Additionally, the use of Levenberg–Marquardt algorithm as the nonlinear optimization framework ensures the algorithm can converge quickly from a wide range of initial guesses due to its effective damping strategy.

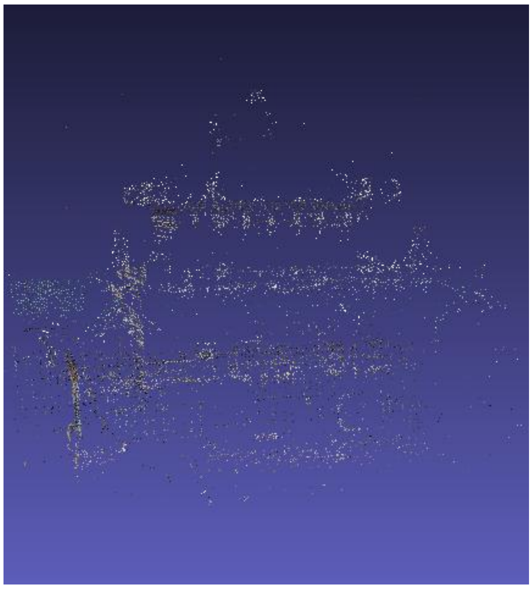


(a) Relief



(b) Exterior of Yungang Grotto

(c) A cave of Yungang Grotto

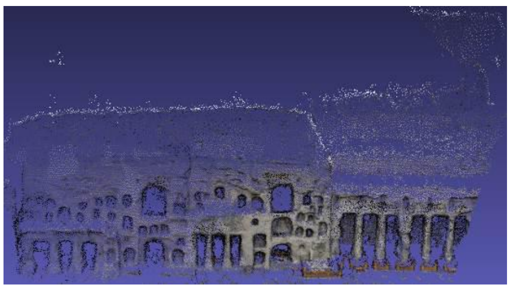


(d) Old library of Wuhan University

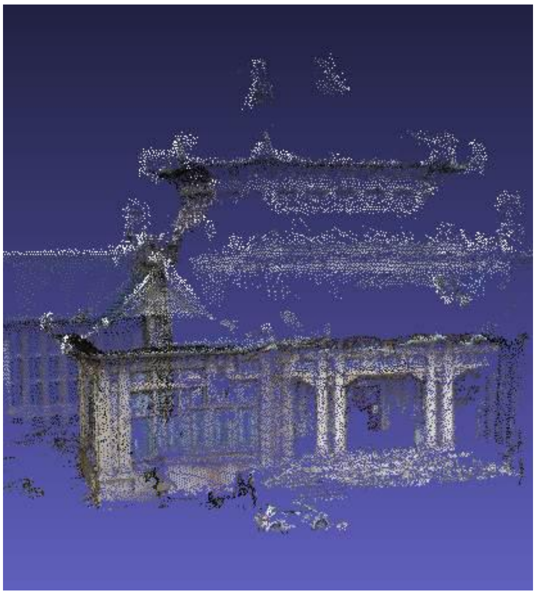
Figure 3 sparse reconstruction

Though we can get a sparse reconstruction of the scene along with the recovering of the pose of the cameras, to insure the accuracy of the estimation of the pose, only points with very high accuracy are left, see Figure 3. To get much denser point clouds, we adopt a patch based accurate dense reconstruction method introduced in (Furukawa, 2007). In this approach, we take the coarse geometry which is indicated as a series of patches as local tangent planes of a surface, each patch is defined by its center with its normal and a reference image in which it is visible. We divide each image into small cells and expand the patches into the cells without any invisible patch according to its neighbors and search its match along the epipolar line. The result is filtered using visibility constraints. Some results are presented in Figure 4.

(a) Relief



(b) Exterior of Yungang Grotto

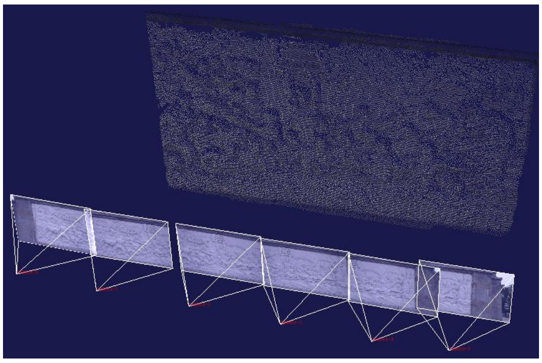
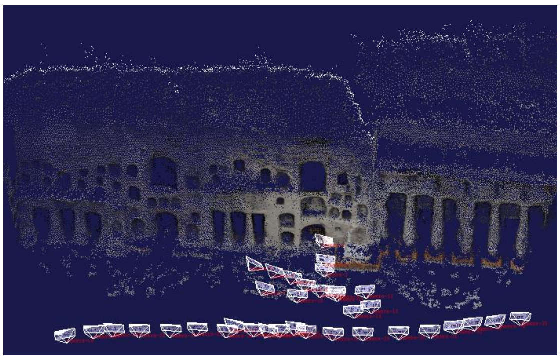


(c) A cave of Yungang Grotto (d) Old library of Wuhan University

Figure 4 sparse reconstruction

3 Rendering and Presentation

Usually, images and point clouds from LiDAR are presented separately while presenting cultural heritages. A major drawback of such an approach is that the audience can never get a high-fidelity experience because point clouds can only provide the geometric information and the images can only offer a view from a fixed position and direction. By combine them together we can get an outline of the whole scene through 3D point clouds and high resolution color and texture information from different views through shifting between images in this visualization system.

	
(a) Relief	(b) Exterior of Yungang Grotto

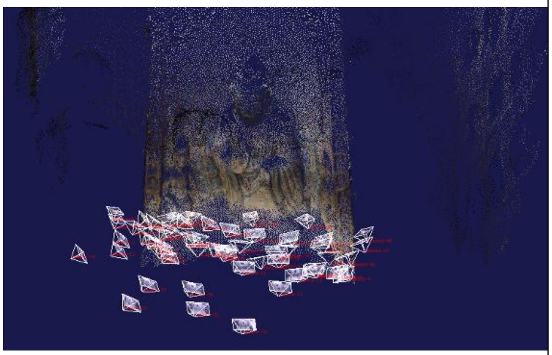
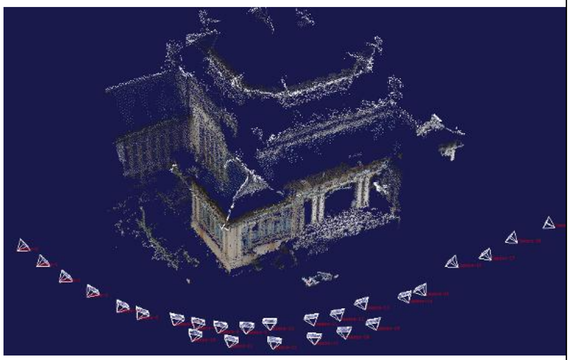
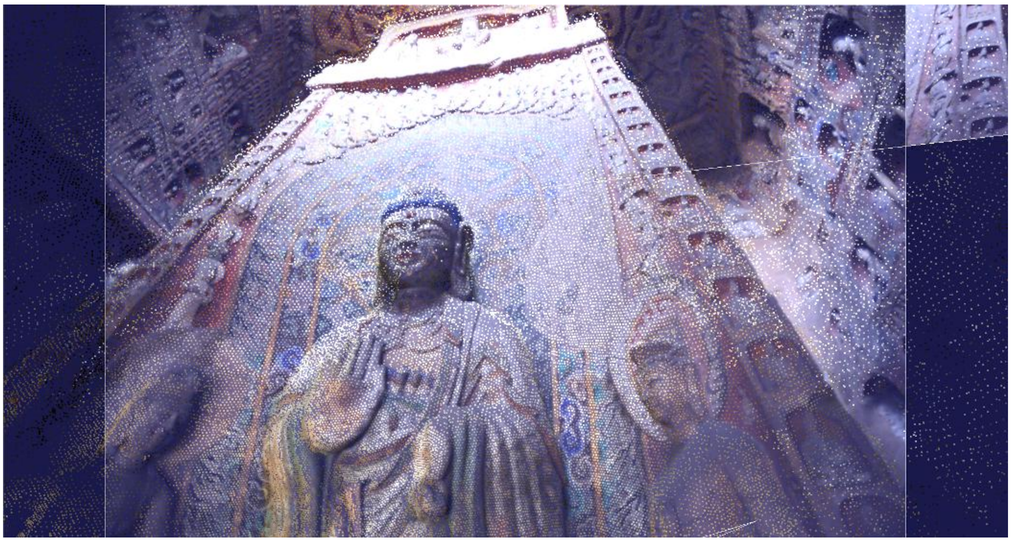
	
(c) A cave of Yungang Grotto	(d) Old library of Wuhan University

Figure 5 Global view

Our visualization system provides the users two kinds of view of the scene. The global view mode provides users a global view of both the geometry of the scene and the position of the cameras while capturing those images, see Figure 5. By double clicking one of the images in the global view, users can enter a perspective view of the scene from the position where the image was captured, see Figure 6. Users can shift from images to images both in the global view and the perspective view mode to appreciate the scene from different views.

(a) Perspective view 1 of old library of Wuhan university


(b) Perspective view 2 of old library of Wuhan university


(c) Perspective view 1 of a cave of Yungang Grotto

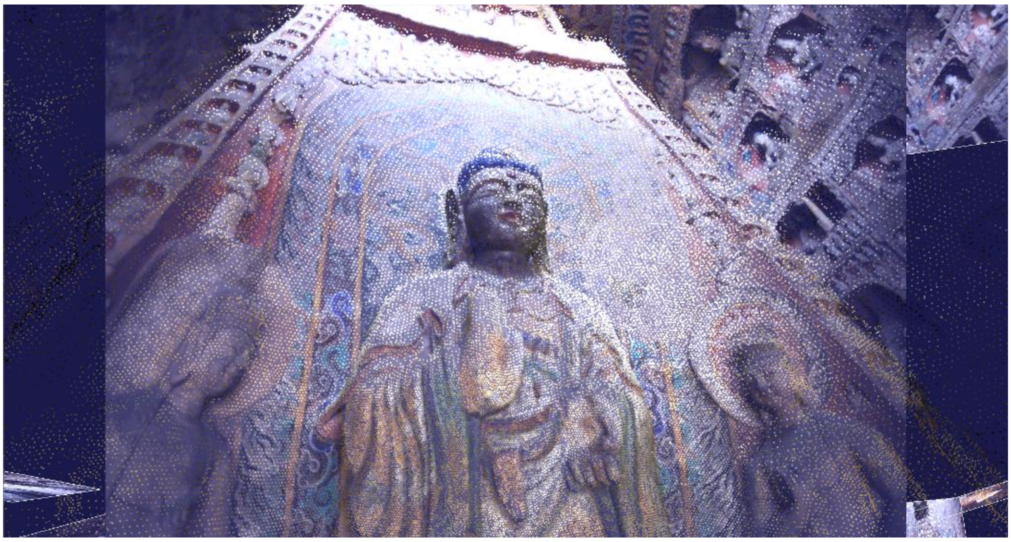

(d) Perspective view 2 of a cave of Yungang Grotto

Figure 6 Perspective view from different view points

4 Conclusions

We developed a pipeline to reconstruct the 3D structure of our vanishing precious heritages for the use of record or restoration and a visualization system which combines the reconstructed point cloud and the image set together to ensure an informative and enjoyable platform. We apply this approach to several different kinds of circumstance, the result is stable and impressive on image sets with large overlapping.

Acknowledge: This Research was sponsored by NSF 41001305.

References

[1]Ren Qing, Diao Chang yu, Lu Dong-ming, Liu Gang, Adopting Structure Light and Triangulation Technology to Implements A 3D Reconstruction Prototype System of Historic Heritage, Dunhuang Research, 5: 102-106, 2005.

[2]Ke Changqing Feng Xuezhi Gu Guoqin, Three Dimensional Information Restoration of the Digital Images of the Dunhuang Mural Paintings, JOURNAL OF NANJING UNIVERSITY (NATURAL SCIENCES), 42(6):628-634, 2006.

[3]A.Gruen, F. Remondino, L. Zhang, Reconstruction of the Great Buddha of Bamiyan, Afghanistan, Int. Arch. Photogrammetry Rem. Sens. 34 (5):363–368, 2002.

[4]Abio Remondino, Stefano Girardi, Alessandro Rizzi, and Lorenzo Gonzo, 3D modelling of complex and detailed cultural heritage using multi-resolution data. ACM Journal on Computing and Cultural Heritage2(1):Article 2, 2009.

[5]Zhao Xu, Zhou Keqin, Yan Li, Deng Fei, 3D Reconstruction Method for Large Scale Relic Landscape from Laser Point Cloud, Geomatics and Information Science of Wuhan University, 33(7):684-687, 2008.

[6]Liu gang, Zhang jun, Diao Changyu, Study 3D Digitization Techniques Applying to Dunhuang Caves, Dunhuang Research, 4: 104-109, 2005.

[7]Levoy, M., Pulli, K., Curless, B., Rusinkiewicz, S., Koller, D.,Pereira, L., Ginzton, M., Anderson, S., Davis, J., Ginsberg, J., Shade, J.,and Fulk, D. “The Digital Michelangelo Project: 3D Scanning of LargeStatues,” Proc. SIGGRAPH, 2000.

[8]Noah Snavely, Steven M. Seitz, Richard Szeliski, "Photo tourism: Exploring photo collections in 3D," ACM Transactions on Graphics (SIGGRAPH Proceedings), 25(3), 2006, 835-846.

[9]Richard Hartley and Andrew Zisserman (2003). Multiple View Geometry in computer vision. Cambridge University Press. ISBN 978-0-521-54051-3.

[10]M. I. A. Lourakis and A. A. Argyros. Sba: A software package for generic sparse bundle adjustment. ACM TOMS,36(1):1–30, 2009.

[11]B. Triggs, P. McLauchlan, H. R.I, and A. Fitzgibbon. Bundle adjustment - a modern synthesis. In Vision Algorithms’99,pages 298–372, 1999.

[12]F.Moreno-Noguer, V.Lepetit and P.Fua , Accurate Non-Iterative O(n) Solution to the PnP Problem, IEEE International Conference on Computer Vision, Rio de Janeiro, Brazil, October 2007.

[13]Y. Furukawa and J. Ponce. Accurate, dense, and robust multiview stereopsis. In IEEE Conference on Computer Visionand Pattern Recognition, 2007.

[14]Martin A.Fischler and Robert C.Bolles. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography, Comm. of the ACM 24 (6): 381–395.