Finding Intrinsic and Extrinsic Viewing Parameters from a Single Realist Painting

Tadeusz Jordan[1], David G. Stork[2, 3], Wai L. Khoo[1], Zhigang Zhu[1]

[1] CUNY City College, Department of Computer Science, Convent Avenue and 138th Street, New York NY 10031
tedjj123@gmail.com, wlkhoo@gmail.com, zhu@cs.ccny.cuny.edu
[2] Ricoh Innovations, 2882 Sand Hill Road, Suite 115, Menlo Park CA 94025-7054
[3] Stanford University, Department of Statistics, Stanford CA 94305
artanalyst@gmail.com

Abstract. In this paper we studied the geometry of a three-dimensional tableau from a single realist painting – Scott Fraser's *Three way vanitas* (2006). The tableau contains a carefully chosen complex arrangement of objects including a moth, egg, cup, and strand of string, glass of water, bone, and hand mirror. Each of the three plane mirrors presents a different view of the tableau from a virtual camera behind each mirror and symmetric to the artist's viewing point. Our new contribution was to incorporate single-view geometric information extracted from the direct image of the wooden mirror frames in order to obtain the camera models of both the real camera and the three virtual cameras. Both the intrinsic and extrinsic parameters are estimated for the direct image and the images in three plane mirrors depicted within the painting.

Keywords: camera calibration, perspective geometry, art analysis

1 Introduction

The problem of reconstructing a three-dimensional scene from multiple views is well explored, and a number of general methods, such as those based on correlation, relaxation, dynamic programming, have been developed and fully characterized [5, 6]. Three-dimensional reconstruction and metrology can be based on single views as well [3, 1]. Criminisi and his colleagues [2] have recently applied such techniques to the analysis of paintings, for instance reconstructing the virtual spaces in Masacio's *Holy Trinity* (c. 1425), Piero della Francesca's *Flagellation of Christ* (c. 1453), Hendrick V. Steenwick's *St. Jerome in his study* (1630), Jan Vermeer's *A lady at the Virginals with a gentleman* (1662-1665), and others. These methods reveal both the high geometric accuracies in some passages, and the geometric inconsistencies in others, properties that are nearly impossible to determine by eye. Such analyses shed new light on these works and the artists' working methods, for instance revealing whether an artist likely used geometrical aids during the execution of their work.

Recently Smith, Stork and Zhang reconstructed the three-dimensional space depicted in a highly realistic modern painting, Scott Fraser's *Three way vanitas* (Fig.

1) using traditional multiple-view reconstruction methods applied to the direct view and a view visible in a depicted mirror [11]. Even though using reflected images by mirrors is a very popular approach for stereo vision in computer vision [7, 9, 10, 12], it was the first time to analyze a painting with such a setup. However, there were some limitations in that previous work as well as unexplored opportunities. For instance, the images of the frames of the mirrors provide geometric constraints about the centers of projection of the images depicted within each mirror, and the earlier scholarship did not incorporate that information when reconstructing the three-dimensional space.

Section 2 describes the painting, previous scholarship and an overview of our new approach to camera parameter estimation. Section 3 introduces some notations and constraints used in the paper. Sections 4 to 7 describe the details of estimating both intrinsic and extrinsic parameters of the main and virtual cameras: finding image center, estimating focal length, representing mirror planes and locating virtual cameras. Section 8 gives a brief summary and discussion.
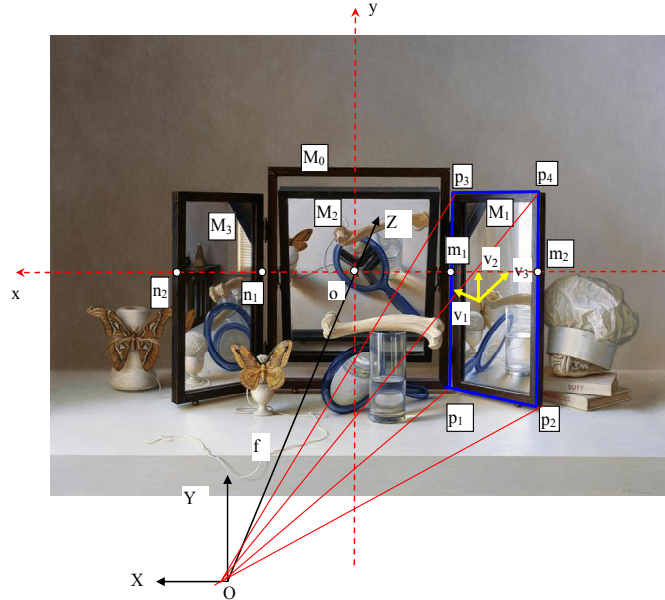

Fig. 1. The work, the notations and the constraints.

2   The Work and Problem Addressed

Fig. 1 shows the work we consider, Scott Fraser's *Three way vanitas* (2006). This painting was commissioned as part of The Object Project, in which fifteen artists were commissioned to create works, each containing five specified objects: hand mirror, bone, moth, ball of string and drinking glass [4]. Our method for estimating the camera models for the virtual cameras is based on the single-image information of the

mirror frames in the primary image of the painting; the image seen from the artist's viewing point (the main camera). We construct geometric invariants such as horizontal/vertical lines and vanishing points from the locations of the vertices of its frame, to estimate the focal length, the center of projection of the main "camera", and the location and orientation of each mirror. Then, we use the pose information of each mirror to compute the location and orientation of the virtual camera by finding its coordinate system that is symmetric to the artist's viewing point in the plane of the mirror. Using the knowledge of the mirror frames, both the intrinsic and extrinsic parameters between a pair of stereo cameras can be found. This is difficult if basic fundamental matrix method is used, as in [11], which used only limited number of points on the table in both the real image and the mirror images.

3   Notations and Constraints

We label the mirrors, reading right to left, M_i, their associated reflected images I_i, and corresponding centers of projection C_i, for i = 1, 2, 3. The middle frame is labeled as M_0. Let's use the mirror M_1 as an example. Each of the four vertices of the frame rectangle, P_i has a corresponding image point $p_i = (a_i, b_i, f)$, i = 1, 2, 3, 4. These image points can also be viewed as vectors from the optical center O to those points. Since $P_1P_2 \parallel P_3P_4$, the direction of the parallel lines can be estimated as

$$v_1 = (p_3 \times p_4) \times (p_1 \times p_2) . \tag{1}$$

Since $P_1P_3 \parallel P_2P_4$, the direction of the parallel lines can be estimated as

$$v_2 = (p_4 \times p_2) \times (p_3 \times p_1) . \tag{2}$$

The normal of the rectangular mirror surface is

$$v_3 = v_1 \times v_2 . \tag{3}$$

Here are a few notes:
1. Point (a_i, b_i) should be measured in the xoy coordinate system that is aligned with the main camera coordinate system O-XYZ. Therefore the image center o should be estimated first for using the above equations.
2. The focal length f is unknown and should be estimated first to use Eqs. (1) – (3).
3. If the projections of a pair of parallel lines are not parallel in the image, such as p_1p_2 and p_3p_4, the direction of the pair can be calculated through their vanishing point in the image plane, (v_{1x}, v_{1y}, f), therefore

$$v_1 \cong (v_{1x}, v_{1y}, f) . \tag{4}$$

Otherwise, the vanishing point is in infinite, and the third dimension of the direction vector such as v_2 should be zero. This is the advantage to use Eqs. (1) to (3) to find directions of parallel lines instead of using vanishing point estimation.

In the following, we will describe methods in estimating the following parameters of the real camera and the three virtual cameras created by the three mirrors: 1). the image center $o(c_x,c_y)$; 2). the focal length f; 3). the plane equation of each mirror; and 4). the pose of the three virtual cameras related to the real camera (the artist's eye). The four cameras share the same focal length and the same image center, but the images of the three mirrored cameras are reversed in the x direction. The intrinsic geometric constraints (*assumptions*) we use are the following:

1. All three mirrors are rectangular.

2.. The two flanking plane mirrors have the same inherent shape and are arranged vertically, each rotated by an unknown angle, and that the central plane mirror is viewed frontally, tipped forward by an unknown angle.

3. The back edges of the two flanking mirrors are at the same distance.

4. The aspect ratio of the image is 1:1.

In addition to the above constraints, by analyzing the images of the frames and mirrors, we have also observed that the left and right flanking mirrors (M_3 and M_1) and the central frame (M_0) are vertical, and the middle inset mirror (M_1) is only tilted in the y direction.
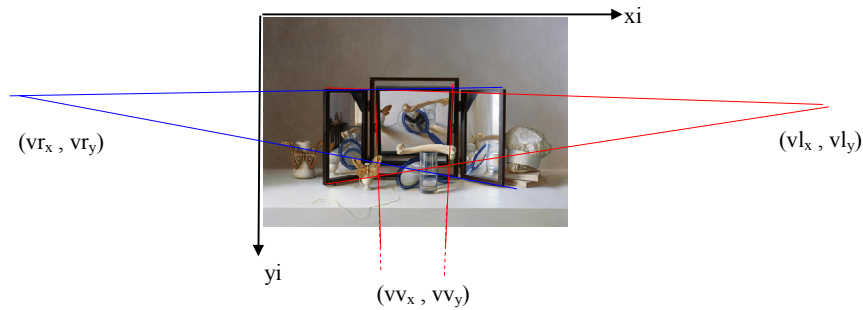

Fig. 2. Estimating image center using vanishing points.

4   Finding Image Center

To find the image center, we will need to have both a set of 3D horizontal lines and a set of vertical lines whose projections are not parallel in the image. By finding the vanishing point of the two pairs of horizontal edges of the left and right flanking mirrors, (vl_x, vl_y) and (vr_x, vr_y), the y coordinate of the image center can be determined as $c_y = (vl_y + vr_y)/2$. By finding the vanishing point of the two vertical edges of the inset mirror, (vv_x, vv_y), the y coordinate of the image center can be determined as $c_x = vv_x$. In Fig. 2, using the digital image coordinate system (x_i, y_i) as shown in Fig. 2, we obtain the results in Table 1:

Table 1. Vanishing points portrayed in Fig. 2.

Using left mirror	Using right mirror	Using middle mirror
(3219.8, 572.1)	(-1360.0, 519.7)	(868.6, 6524.3)

5 Estimating Focal Length

The focal length estimation is the most critical step. Once we find f, the rest of steps will be rather straightforward. We can find the focal length f by using the fact that the left and right flanking mirrors have exactly the same width, i.e. $W_1 = W_2$ in Fig. 3.
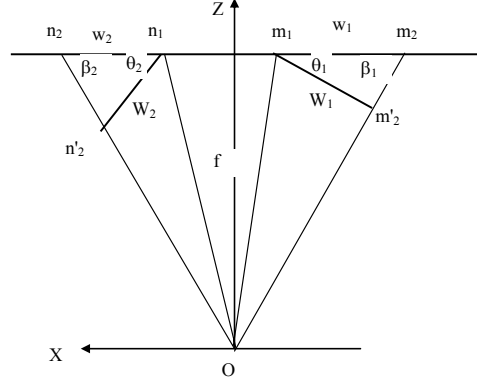

Fig. 3. Focal length estimation using the equal-width constraint

We scaled the mirrors into pixel representation so the distance of the two back edges (m_1 and n_1) are f (refer to Fig. 1). The directions of $n_1n'_2$ and $m_1m'_2$ can be represented by the vanishing points, $v_m = (v_{mx}, v_{my}, f) = (2228.6, 7.2, f)$, generated from the two sets of horizontal edges. The dot product of normalized v_m and the X-axis (1, 0, 0) should be $v_m / |v_m| \cdot (1, 0, 0) = \cos\theta_1$, assuming that v_m is in the horizontal plane as the X-axis (from assumption No 2). We can derive a similar relation for v_n. Therefore, the angles θ_1 and θ_2 can be represented as functions of f:

$$\theta_1 = \tan^{-1}(f / |v_{mx}|), \;\; \theta_2 = \tan^{-1}(f / |v_{nx}|) \tag{5}$$

From the image, location m_1, m_2 and n_1, n_2 can be measured (in the x direction, refer to Fig. 1). Therefore, the two distances w_1 and w_2 can be calculated as $w_1 = |m_2-m_1|$, $w_2 = |n_2-n_1|$. The two angles β_1 and β_2 can be also represented as functions of f:

$$\beta_1 = \tan^{-1}(f / |m_2|), \;\; \beta_2 = \tan^{-1}(f / |n_2|) \tag{6}$$

Using the sine law with triangles $n_1n_2n'_2$ and $m_1m_2m'_2$, and using the fact that $W_1 = W_2$, we can derive that:

$$f = \sqrt{\frac{a_n^2 v_{nx}^2 - a_m^2 v_{mx}^2}{a_m^2 - a_n^2}} \tag{7}$$

where $a_m = \dfrac{w_1}{|m_2| + |v_{mx}|}, a_n = \dfrac{w_2}{|n_2| + |v_{nx}|}$.

Using the measurements from the image, we obtain f = 2050.7 (pixels) Note that this method does not work if the two mirrors are symmetric. We calculated the two angles using Eq. (5): $\theta_1 = \tan^{-1}(f / |v_{mx}|) = 42.62°$, $\theta_2 = \tan^{-1}(f / |v_{nx}|) = 39.36°$. The two mirrors have different flanking angles, so they are not symmetric, therefore we can obtain the focal length estimation.

6   Representing Mirror Planes

Once we obtain the value of the focal length f, we can use the general equations (1) to (3) to calculate the directions of the two edges of each mirror M_i, and its normal n_i. These equations work for all the cases no matter if we can find vanishing points or not. The results of the rotation matrices for M_i (i=0, 1, 2, 3) are listed in Table 2.

Table 2. The rotation matrices (2nd row) and angles (3rd row, in degrees) of M_i (i = 0, 1, 2, 3).

M_0	M_1	M_2	M_3
0.9999 -0.0054 0.0129 0.0053 0.9999 0.0112 -0.0130 -0.0114 0.9999	0.7363 0.0157 -0.66765 0.0201 0.9988 0.0451 0.6764 -0.0468 0.7350	1.0000 -0.0028 0.0054 0.0009 0.9455 0.3257 -0.0060 -0.3257 0.9454	0.7743 -0.0046 0.6328 0.0028 1.0000 -0.0038 -0.6328 -0.0012 0.7743
(0.66, -0.76, 0.31)	(3.59, 43.16, 1.26)	(19.38, -0.32, 0.16)	(0.29, -39.84, 0.35)

From the rotation matrices it can be seen that the three angles of the middle frame M_0 are almost zero degrees, the middle inset mirror M_2 is mainly tipped forward, the right mirror M_1 and the left mirror M_2 mainly have flanking angles consistent with the results obtained by using Eq. (5). The results highly agree with our assumptions of the intrinsic geometry constraints in Section 2.

The plane equation of the mirror can be represented as

$$\mathbf{n}_i \bullet \mathbf{P}_c + d_i = 0 \ . \tag{8}$$

where $P_c = (X_c, Y_c, Z_c)^t$ is represented in the camera coordinate system O-XYZ. In order to find d_i, and further find the virtual camera parameters mirrored by each mirror, we will build a world coordinate system on each mirror. For example, for mirror M_1, after we find the three vectors and then normalized them into column unit vectors, still represented in v_1, v_2 and v_3, we can define a world coordinate system using the middle of the mirror plane as its origin, and the three vectors as its three coordinate axes. The transformation between the camera coordinate system and the world coordinate system of the ith mirror M_i (i = 0, 1, 2, 3) can be represented as

$$P_c = \mathbf{R}_{1i}P_w + \mathbf{T}_{1i} \ . \tag{9}$$

where $P_c = (X_c, Y_c, Z_c)^t$, $P_w = (X_w, Y_w, Z_w)^t$, $R_{1i} = (r_{pq})3x3$, which is (v_1, v_2, v_3) for M_1, and T_{1i} to be determined. The projection of P_c into the image of the main camera is $(x, y) = (f X_c/Z_c, f Y_c/Z_c)$.

To find the translational vectors for all the three mirrors, we use the dimension of the middle frame M_0 as a reference. To calculate the translations for mirrors M_1 and

M_3, we use their heights h= 566.464 pixels, which can be measured in image with respect to the height H of the middle frame, since they are at about the same distance. Similarly, to calculate the translation for the middle inset mirror, we use its width w = 461.5911 pixels, which can be measured against the width W.

Table 3. The translations and the distances of M_0 to M_3

Mirror	T_{1i} (i = 0, 1, 2, 3)	d_i (i = 0, 1, 2, 3)
M_0	(-51.3, -176.1, 2048.5)	2045.6
M_1	(-430.1, -215.7, 1929.7)	1802.8
M_2	(-50.9, -182.6,2051.1)	1879.5
M_3	(340.5, -214.6, 1968.0)	1801.1

7 Locating Virtual Cameras

After we obtain the plane equation for each mirror (Eq. (8), the mirrored coordinate system, i.e., the virtual camera C_i, can be easily obtained, as in [9]. Here we use a coordinate transformation method to find the relation between each virtual camera C_i and the real camera C, by finding the rotation matrix R_i and translational vector T_i (i = 1, 2, 3). In our implementation, we use Eq. (9) to represent the origin and the three axes of the camera coordinate system in the world coordinate system $X_wY_wZ_w$ of each mirror M_i. Since X_wOY_w is the mirror plane, the mirrored origin and axes can be simply obtained by changing the signs of their Z_w components. Then we do a similar procedure as in Eq. (9) to find the transformation (characterized by R_{2i} and T_{2i}, i = 1, 2, 3) between the world coordinate system M_i and the virtual camera C_i:

$$P_w = \mathbf{R}_{2i}P_i + \mathbf{T}_{2i} . \tag{10}$$

Combining Eqs (9) and (10), we can find the transformation between the real camera and the virtual camera:

$$P_c = \mathbf{R}_i P_i + \mathbf{T}_i . \tag{11}$$

$$\mathbf{R}_i = \mathbf{R}_{1i}\mathbf{R}_{2i}, \mathbf{T}i = \mathbf{R}_{1i}\mathbf{T}_{2i} + \mathbf{T}_{1i} . \tag{12}$$

Table 4 shows the estimated results. Using the full 6 degree-of-freedom (DOF) relation between each virtual camera and the main camera, stereo reconstruction can be performed, and the accuracy of the painting can be analyzed.

Table 4. The transformations between virtual camera C_i (i = 1, 2, 3) and the main camera

M_i	M_1			M_2			M_3		
R_i	-0.0846	0.0610	0.9945	-0.9999	-0.0035	-0.0103	-0.1992	-0.0049	-0.9800
	-0.0610	0.9959	-0.0663	0.0035	0.7878	-0.6159	0.0049	1.0000	-0.0059
	-0.9945	-0.0663	-0.0806	0.0103	-0.6159	-0.7877	0.9800	-0.0059	-0.1991
T_i	(-2299.7, 153.4,2498.7)			(20.4, 1224.4, 3553.8)			(2200.2, 13.3, 2692.3)		

8   Conclusions and Discussions

Our method for estimating the camera models for the virtual cameras was based on single-view analysis. The relative 3D structures of the rectangles mirrors and frames are estimated by using their perspective analysis with a few assumptions. The camera calibration for the cameras is then successfully performed. Its results can then be used for 3D reconstruction and painting analysis, which is described in an accompanying paper [8]. The 3D estimates of both the frames and regular objects are consistent among the single-view analysis, and results from multiple stereo triangulations. However, there are some stereo and perspective inconsistencies across four views. These could either be the accuracies of the perspective distortions, orientations and sizes of the mirrors, or of the locations of objects inside the mirrors, which cannot be easily observed by eye. As such, our work extends the new discipline of computer vision applied to the study of fine art.

References
[1] A. Criminisi. Accurate visual metrology from single and multiple uncalibrated images. ACM Distinguished Dissertation Series. Springer-Verlag, London, 2001.
[2] A. Criminisi, M. Kemp, and A. Zisserman. Bringing pictorial space to life: computer techniques for the analysis of paintings. In A. Bentkowska-Kafel, T. Cashen, and H. Gardner(eds), Digital Art History: A Subject in Transition, 2005: 77–100.
[3] A. Criminisi, I. Reid, and A. Zisserman. Single view metrology. IJCV 40(2), 1999:123–148.
[4] M. S. Doherty. Object Project: Five objects, fifteen artists. Evansville Museum of Arts, History and Science, Evansville, IN, 2007.
[5] O. Faugeras. Three-Dimensional Computer Vision: A Geometric Viewpoint. MIT Press, Cambridge, MA, 1993.
[6] R. Hartley and A. Zisserman. Multiple View Geometry in Computer Vision. Cambridge University Press, Cambridge, UK, 2nd edition, 2003.
[7] P.-H. Huang, S.-H. Lai. Contour-based structure from reflection. CVPR'06, v 1: 379-386.
[8] W. Khoo, T. Jordan, D. Stork, Z. Zhu, Reconstruction of a three-dimensional tableau from a single realist painting, *15th Int. Conf. Virtual Systems & Multimedia*, Sept. 9-12, 2009
[9] R. K. Kumar , A. Ilie, J.-M. Frahm, M. Pollefeys. Simple calibration of non-overlapping cameras with a mirror. CVPR'08
[10] S.A. Nene, and S. K. Nayar. Stereo with mirrors, ICCV'98: 1087-1094
[11] B. Smith, D. G. Stork, and L. Zhang. Three-dimensional reconstruction from multiple reflected views within a realist painting: an application to Scott Fraser's Three way vanitas. SPIE/IS&T, Bellingham, WA, 2009.
[12] Z. Zhu, X. Lin, D. Shi, G. Xu, A Single camera stereo system for obstacle detection. 4th Int. Conf. Information Systems Analysis and Synthesis, July 12-16, 1998, v 3:230-237