INSERTION OF THREE-DIMENSIONAL OBJECTS IN ARCHITECTURAL PHOTOS

Beatriz Silva Villa Alvarez
Paulo Cezar Pinto Carvalho
Marcelo Gattass

Instituto de Matemática Pura e Aplicada - IMPA
Estrada Dona Castorina 110
22.460-320 Rio de Janeiro/RJ
Brasil

Departmento de Informática
Pontifícia Universidade Católica do Rio de janeiro - PUC-Rio
Marquês de São Vicente 225
22.453-900 Rio de Janeiro/RJ
Brasil

bia@visgraf.impa.br, pcezar@impa.br , gattass@tecgraf.puc-rio.br

Abstract

This paper proposes a simple and interactive system that allows modifying a photographic picture of a three-dimensional scene involving architectural elements, so that the user can evaluate the aesthetic effects and the impact such modifications in the real environment would cause in other people. The method is based on the existence, in architectural pictures, of three main directions of interest, which are mutually orthogonal. The identification of vanishing points of such directions allows calibrating the camera used to take the pictures and also inserting new elements into the scene.

Keywords: three-dimensional edition, camera calibration, projection, pictures, vanishing points, orthogonal directions.

1 INTRODUCTION

This paper presents a simple and interactive system capable of modifying a photographic picture of a three-dimensional scene involving architectural elements, by inserting new objects into certain planes in the scene. The system is targeted to Architecture, Photography, Publicity and Visual Arts professionals whose needs involve modifying real settings. For instance, architects could insert new three-dimensional objects in building façades to evaluate the result of refurbishment or restoration before spending time and money in the actual execution. Publicity professionals might desire to have an early visualization of their outdoor campaigns and test their visual impact (see Figure 1). Graphics artists could also enhance their wall paintings, nowadays so common in big cities, by testing the aesthetic result virtually and studying the color and size of their works before they become real.

The system also allows extracting texture with adequate perspective correction, which could be used for rendering synthetic 3D models, and obtaining a new view of the scene in which the projections of vertical lines become parallel. This last tool is useful in those cases in which the camera is too close to the object of interest and the picture cannot be taken by a camera parallel to the vertical direction (hence, vertical lines seem to converge to their vanishing point).

The techniques proposed here can be easily embedded into existing 2D image editing software (for instance, as plug-ins), in order to provide the user with new, easy-to-use, geometrically correct tools for modifying 3D scenes.

In Section 2, we will describe the problem of editing three-dimensional scenes and its possible approaches. Then, in Section 3, we will discuss in detail the implementation of the proposed method. Finally, in Section 5, we will draw some conclusions and suggest possible extensions of this work.

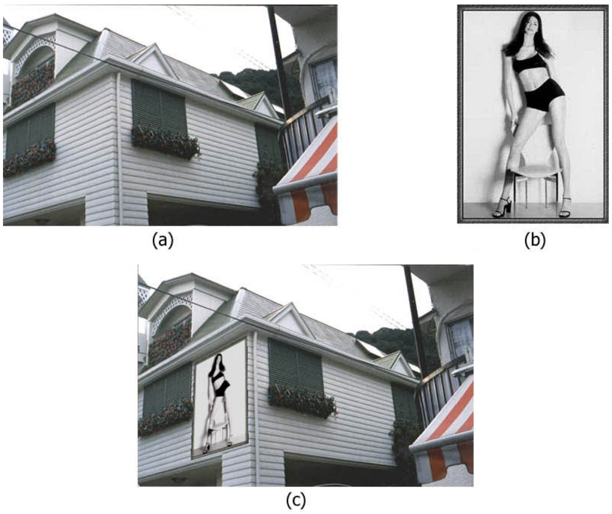

Figure 1: (a,b) Original images; (c) outdoor applied to the house's façade

2 PROBLEM DESCRIPTION

When editing photographic pictures, one commonly wishes to insert new images in certain planes of the original picture or to obtain a new view of the scene using a projection that preserves the proportions parallel to a given plane. Such tasks are often executed *by feeling* - the user deforms the image, by means of trial and error, until the desired effect is obtained. However, such process is time-consuming, tedious, and it not always provides good results, even for experienced and capable users.

The problem of obtaining three-dimensional information from pictures has been receiving a good deal of attention in the literature, but usually with purposes different from the ones related to the present work. Several works [Chen95, MB95, HAA97] approach the visualization problem using image-based rendering techniques whose goal is to obtain new views of a scene from a set of images. Other works deal with the three-dimensional reconstruction of a scene using stereo vision [FRL98] or adopting mathematical models for the objects present in the scene [DTM96].

Differently from such approaches, the method presented here does not seek to obtain a complete reconstruction of the scene. On the other hand, it uses as input data only one picture, for which no metric information is assumed to be known, neither about the scene (such as coordinates of reference points or measures of elements) nor about the camera position.

The present work has several points in common with [P3D99], which provides an environment for extracting three-dimensional information from a single picture. However, the emphasis of that work is the partial modeling of three-dimensional objects in the scene, whereas our goal is to provide tools for editing, in a geometrically correct manner, planes of the original picture, by inserting new objects into such planes with the proper perspective. This method also aims at allowing the user to extract a texture from the input picture with automatic perspective correction. It also provides a way to simulate a photographic lens capable of correcting the angular distortion caused by the perspective projection. We must note that this projection can not always be obtained with a real photographic lens.

3 METHOD DESCRIPTION

3.1 Camera Calibration

One of the preconditions to construct geometric representations from photographic pictures is to know or identify the parameters of the camera that captured the picture of interest; among them, the camera orientation and virtual position in the 3D space. With this information, a transformation is obtained such that, given a point in space, its corresponding point in the image is located. This problem is called *camera calibration*, and a more detailed study can be seen in [Fau93, TGTL91]. We are considering a simple pin-hole camera model, which does not take into account the existence of lenses: the image of a 3D object is obtained by its perspective projection on a plane 2D surface through the camera optical center (pin-hole).

Several camera-calibration methods are based on the existence of reference points in the picture, whose coordinates in the real world are known [CSG98, Tsai86]. Others [DTM96] employ known metric information, such as the dimensions of architectural elements like windows or doors. In the present paper, we assume that no such information is available. This is important, for instance, in situations involving old pictures, for which it is impossible to recover these data.

The calibration method used in this work, which is the same as in [P3D99], resorts exclusively to information on the directions - more precisely, to the vanishing points relative to three mutually orthogonal directions. As will be shown next, these vanishing points determine the position of the camera optical center and angle of view. However, without complementary metric information, it is impossible to retrieve the focal distance. As a consequence, objects can be inserted in the scene with correct proportions, but without an absolute control of their size. For example, in Figure 1, the size of the outdoor can be specified only in relation to elements in the scene, such as the floor-to-ceiling height.

The need to identify vanishing points of three orthogonal directions makes this method especially suitable for architectural pictures, in which there are usually three easily identifiable directions. We also admit the elements to be inserted on the image, or from which one wishes to obtain information from the image, to be aligned with such directions.

3.2 Obtaining the Camera's Position

In this subsection we will show how to obtain the camera position in relation to the picture from vanishing points F_x, F_y and F_z relative to the three orthogonal directions X, Y and Z. In architectural pictures, the choice of these directions is natural, as they correspond to the main height, width and depth directions of buildings. The vanishing points are the points where the lines parallel to these three directions passing through the optical center intersect the projection plane. Therefore, together with the optical center C, they determine three mutually orthogonal directions. This condition allows retrieving the position of the optical center C in relation to the image, by solving a system of equations expressing the orthogonality of $C\vec{F}_x$, $C\vec{F}_y$ e $C\vec{F}_z$ or by using the fact that C is projected on the orthocenter of the triangle having vertices F_x, F_y and F_z, as illustrated in Figure 2. The position (u_c, v_c, w_c) of C is expressed in the image coordinate system S_i, with origin on the left bottom corner, two axes u and v aligned with the borders of the image, and a third axis w orthogonal to these borders. The coordinates in this system are expressed in pixels. As already mentioned, it is not possible to obtain the real focal distance unless one knows the dimensions of a pixel. It is also important to note that the projection of the optical center does not necessarily coincide with the center of the image (Figure 3), since the image might represent only a portion of the original picture.

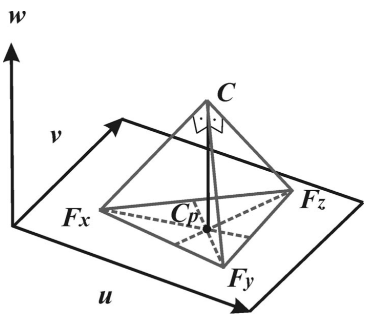

Figure 2: Location of the optical center

In fact, the user does not directly specify vanishing points F_x, F_y and F_z : he/she indicates, for each of the main directions, two or more straight lines in the image which are projections of straight lines in the scene having these directions. The corresponding vanishing point is obtained by intersecting these straight lines. To allow the user to indicate the straight lines corresponding to each direction, the system offers an interface with three pairs of guidelines to be positioned over the straight lines in the image, as illustrated in Figure 3.

3.3 Projection Equations

The camera projection equations associate each point in space to its perspective projection on the image. The position of a point P in space is described by means of its coordinates in the world coordinate system S_m, with origin at the camera optical center C and axes aligned

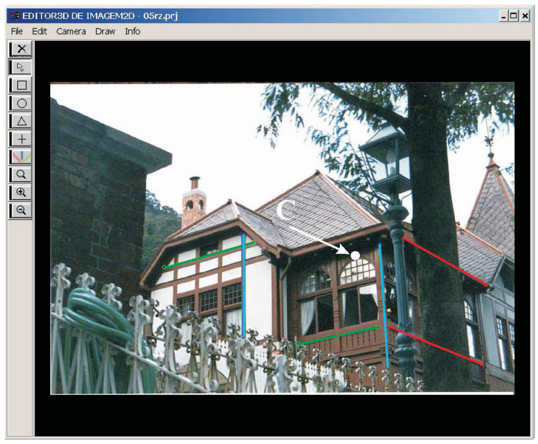

Figure 3: Guidelines for the directions, and position of the optical center **C**

with the main directions X, Y and Z. The axes in this system pass through points F_x, F_y and F_z, respectively. Since the real focal distance is not known, the coordinates in this system are also expressed in pixels. To express them in conventional distance measures one would have to know, once again, the dimensions of a pixel.

Therefore, the projection equations include a point (X, Y, Z) of the world in their projection (u, v) on the image. To write these equations, it is convenient to consider an intermediate coordinate system: that of the camera, S_c, with origin on the optical center C and axes U, V and W parallel to the axes of the coordinate system of the image, S_i. Figure 4 illustrates these systems. Given a point (X, Y, Z) of the world, its coordinates on the camera's system are given by:

$$\begin{bmatrix} U \\ V \\ W \end{bmatrix} = \begin{bmatrix} X_u & Y_u & Z_u \\ X_v & Y_v & Z_v \\ X_w & Y_w & Z_w \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \quad (1)$$

where (X_u, X_v, X_w), (Y_u, Y_v, Y_w) e (Z_u, Z_v, Z_w) are unit vectors corresponding to $C\vec{F}_x$, $C\vec{F}_y$ e $C\vec{F}_z$, respectively.

Once such camera coordinates are obtained, obtaining their projection (u, v) is immediate. In homogeneous coordinates, it is given by:

$$\begin{bmatrix} tu \\ tv \\ t \end{bmatrix} = \begin{bmatrix} w_c & 0 & u_c \\ 0 & w_c & v_c \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} U \\ V \\ W \end{bmatrix} \quad (2)$$

Thus, the camera projection equation is expressed by:

$$\begin{bmatrix} tu \\ tv \\ t \end{bmatrix} = \begin{bmatrix} w_c & 0 & u_c \\ 0 & w_c & v_c \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X_u & Y_u & Z_u \\ X_v & Y_v & Z_v \\ X_w & Y_w & Z_w \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \quad (3)$$

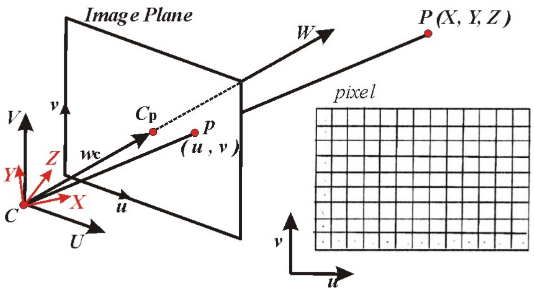

Figure 4: Coordinate systems

3.4 Editing in a Main Plane

Once the camera position and the projection equations have been obtained, it is possible to edit portions of the image corresponding to projections of images contained in planes parallel to two of the main axes. We will consider planes parallel to axes X and Y, but the other cases can be handled similarly.

We assume that the portion of the image to be edited is the projection of a rectangle with sides parallel to X and Y, specified by selecting two diagonally opposed points p_1 and p_2 corresponding to the projections of two vertices P_1 and P_2 of the rectangle. There are infinitely many points in space that project on p_1 and p_2, corresponding to the several depths of Z in which the rectangle can be located. Since our purpose is to obtain the ratio between the rectangle dimensions, and not to retrieve their absolute values, this depth can be arbitrarily set. Thus, let us set $Z = 1$. We can now retrieve the position of P_1 and P_2, and consequently that of the two other vertices Q_1 and Q_2 of the rectangle and their projections q_1 and q_2.

We have now established a correspondence between a rectangle in space $P_1Q_1P_2Q_2$ and its projection - quadrilateral $p_1q_1p_2q_2$ - by means of a two-dimensional projective transformation $T^{'}$, which is the perspective projection executed by the camera restricted to plane $Z = 1$. This correspondence allows retrieving a non-deformed image of the quadrilateral $p_1q_1p_2q_2$, obtained by a warping transformation given by the inverse of $T^{'}$. Over this non-deformed image, new elements can be applied, such as signs or pictures. These elements can be inserted back into the original image by means of a warping transformation inverse to the one applied in the previous step. All the process described is illustrated in Figure 5.

Figure 6 shows another example, where an outdoor is applied to a building façade seen from two diferent positions.

In many cases, it can be interesting to execute only part of the process above - for instance, when one wishes to extract information from the scene instead of inserting new elements. The non-deformed image correctly displays the proportions among the elements present on the plane being edited and preserves their angles, so it can be used to retrieve the relations among element dimensions (such as doors or windows) or to extract textures from it, to be used in the construction of virtual models of the scene. Figure 7 illustrates this process: the texture on the wall was extracted in a geometrically correct manner.

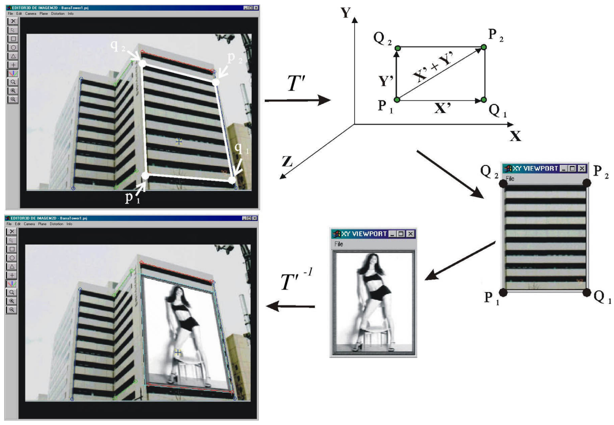

Figure 5: Stages of the editing process on a main plane

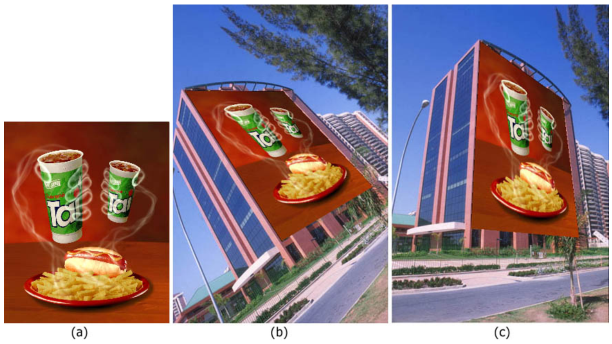

Figure 6: (a) original outdoor image , (b)(c) buildings with inserted outdoors

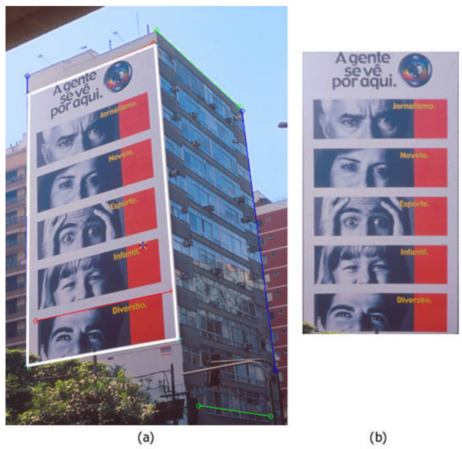

Figure 7: (a) original picture, (b) extracted texture

3.5 Perspective Correction

Though the images produced by means of perspective correction are geometrically correct, they are not always the ideal representation of the object of interest. Particularly, angles are not preserved unless they have sides parallel to the projection plane. For instance, when producing the picture of a building whose height is large with respect to its distance to the camera optical center (i.e., situations in which there is little depth of field), the camera has to be inclined in order to include the greatest possible portion of the object. This causes the vertical direction to be non-parallel to the plane of the picture, introducing a vanishing point corresponding to this direction. This means that the observer has the impression that the vertical edges of the building will meet. Even though this is perfectly normal, architects often prefer images in which the vertical direction is parallel to the plane of the picture, in order to avoid this effect. A possible solution consists in using special lenses capable of producing pictures with the desired characteristics [Nik01].

Our system offers an alternative solution, allowing the user to eliminate the vanishing point corresponding to the vertical direction. The original picture is reprojected on a new plane, parallel to vertical direction Y and positioned at the same distance from the camera as the original projection plane. We select coordinate axes in this new plane so that Y is projected according to the vertical axis in this new picture. This is equivalent to introducing a new camera with the same optical center and focal distance, and new coordinate axes α, β and γ. These new axes are selected so that β has the same direction as Y and γ has a direction as similar to W as possible. This is done by taking γ as the projection of W on the plane orthogonal to Y, as illustrated in Figure 8. Thus, vectors $\vec{\alpha}$, $\vec{\beta}$ e $\vec{\gamma}$ corresponding to these new axes can be obtained from unit vectors $\vec{Y}$ e $\vec{W}$ (respectively vertical and orthogonal to the original projection plane) using the equations below:

$$\begin{aligned} \vec{\beta} &= \vec{Y} \\ \vec{\gamma} &= \vec{W} - (\vec{W} \cdot \vec{Y})\vec{Y} \\ \vec{\alpha} &= \vec{\beta} \times \vec{\gamma} \end{aligned} \tag{4}$$

Finally, to obtain the reprojected picture, one only needs to consider each point (u, v) in the image to have the form (u, v, w_c), then transform it to the coordinate system of the new camera and apply the new perspective projection. Thus, a point (u, v) in the original picture is transformed into a point whose homogeneous coordinates (tu', tv', t) are given by:

$$\begin{bmatrix} tu' \\ tv' \\ t' \end{bmatrix} = \begin{bmatrix} w_c & 0 & u_c \\ 0 & w_c & v_c \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \alpha_u & \alpha_v & \alpha_w \\ \beta_u & \beta_v & \beta_w \\ \gamma_u & \gamma_v & \gamma_w \end{bmatrix} \begin{bmatrix} u \\ v \\ w_c \end{bmatrix} \tag{5}$$

where $(\alpha_u, \alpha_v, \alpha_w)$, $(\beta_u, \beta_v, \beta_w)$ and $(\gamma_u, \gamma_v, \gamma_w)$ are the unit vectors corresponding to vectors α, β and γ obtained in Equation (4).

Equation 5 determines a warping transformation that allows generating the new image.

Figure 9 shows the original images and the results of the reprojected pictures generated by the system. One can see the desired effect: the lines corresponding to the vertical direction of buildings have become parallel and vertical in the reprojected picture.

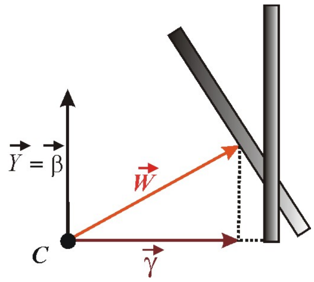

Figure 8: New camera's computation process

Figure 9: Original images and reprojected pictures with vertical inclination

4 APPLICATION

The method described in the previous section was implemented in an application for Windows 98, NT4 or above, available at [VCG00]. The user starts by loading a picture and performing the camera calibration procedure, through the positioning of pairs of guidelines for each one of the three main directions X, Y or Z (Figure 10 and Figure 11). Calibration results can be saved for use in future sessions.

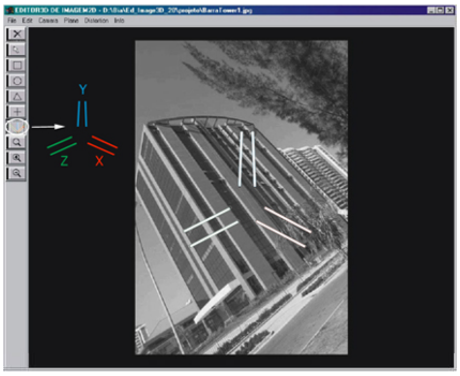

Figure 10: Guideline initialization.

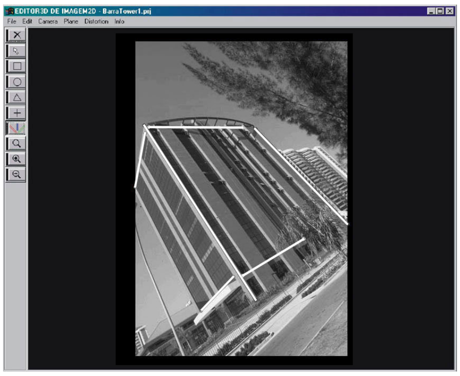

Figure 11: Guideline positioned along lines parallel to the main directions.

After calibration, the user can retrieve, with proper perspective correction, a portion of the image corresponding to the projection of a rectangle parallel to two of the main directions. First, he/she specifies whether the rectangle is parallel to plane XY, XZ or YZ. Then, two diagonally opposed points of the projected rectangle must be indicated (Figure 12). The projected rectangle is drawn over the picture and the corrected (unprojected) rectangle is shown in a separate window (Figure 13).

The unprojected picture can be saved to a file and used for texture extraction or proportion measurement. It can also be replaced by a new image and placed back into the original image, as shown in Figure 14.

There is also a function for eliminating the vanishing point for the vertical (Y) direction, in such a way that vertical lines project vertically in the new image. Figure 15 shows the result obtained when applying that function.

Figure 12: Opposing corners of projected rectangle.

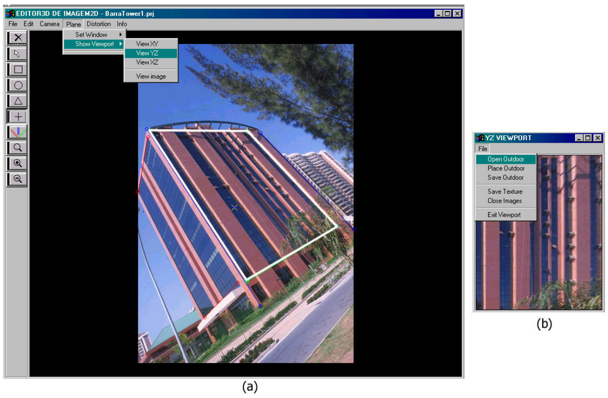

Figure 13: (a) Projected, (b) unprojected rectangle.

5 CONCLUSION AND FUTURE WORK

We have presented a system for three-dimensional editing of architectural photographic pictures. The method is simple enough to be used by the professionals it is aimed at, requiring intuitive information which can be easily identified in the scenes of interest. The method does not require previous data about the scenes, so it can be used with any picture (even historical ones) for extracting information or inserting new virtual elements.

A limitation of this method is that it assumes the edited region to be on a single plane, parallel to two main axes. When this hypothesis is not satisfied, the extracted image

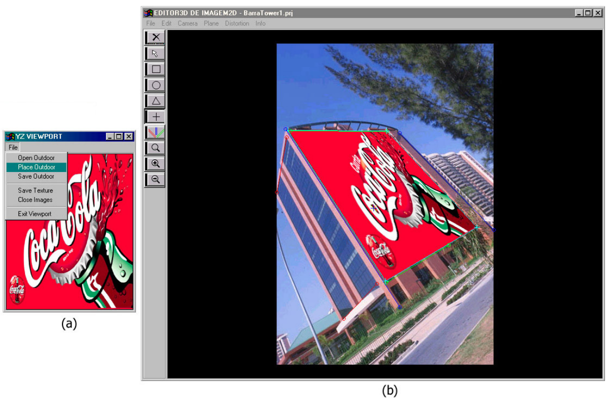

Figure 14: (a) Unprojected outdoor, (b) outdoor inserted into the original image.

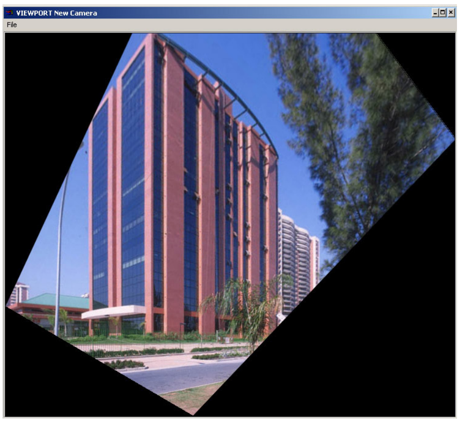

Figure 15: Image with vertically projected vertical lines.

will contain elements with proportions not corresponding to the real ones. Figure 16 illustrates this. When extracting the marked region from the picture, the elements in the balcony are also extracted and are displayed with deformations (actually, the balcony is treated as if it were drawn on the wall, instead of being a three-dimensional element). To address this problem, it would be necessary to extend the editing surface, considering surfaces formed by rectangles parallel to the main axes.

Another natural extension would be to consider planes parallel to only one of the axes. This would be the case of ceilings, ramps or walls not aligned with the main directions, for instance.

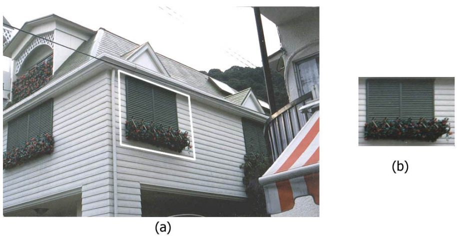

Figure 16: (a) Original image; (b) transformed image, considering all elements on the same plane.

It should also be considered the introduction of automatic or semi-automatic methods for extracting guidelines, through the recognition of straight lines in the image ([SCG00]). For better results, these lines should be extracted at sub-pixel level.

References

[HAA97] Anjyo, K.; Horry, Y. and Arai, K. Tour into the picture: Using a spidery mesh interface to make animation from a single image. In *Conference Proceedings SIGGRAPH'97*, 1997, pp 225–232.

[FRL98] Laveau, S.; Csurka,G.; Zeller,C.; Gauclin,C.; Faugeras,O.; Robert,L. and Zoghlami,I. 3D reconstruction of urban scenes from image sequences. *Computer Vision and Image Understanding*, Vol. 69 No.3(Article No. IV970665.):292–309, March 1998.

[Nik01] Lens Nikon 28mm f/3.5 PC-nikkor perspective control technology. `http://www.nikon.com`, 2001.

[MB95] McMillan, L. and Bishop, G. Plenoptic modeling: An image-based rendering system. In Robert Cook, editor, *Conference Proceedings (SIGGRAPH '95)*, ACM SIGGRAPH, Addison Wesley, Annual Conference Series, Los Angeles, California, 06-11 August 1995, pp 39–46.

[Fau93] Olivier Faugeras. *Three-Dimensional Computer Vision: A Geometric ViewPoint*. MIT Press, Cambridge, Massachusetts, 1993.

[P3D99] Photo 3D. `http://www.photo3D.com`, 1999.

[Tsai86] Roger Y. Tsai. An efficient and accurate camera calibration technique for 3d machine vision. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Miami Beach, FL, 22-26 June 1986.

[Chen95] Shenchang Eric Chen. QuickTime VR - an image-based approach to virtual environment navigation. In Robert Cook, editor, *Conference Proceedings (SIGGRAPH '95)*, ACM SIGGRAPH, Addison Wesley, Annual Conference Series, Los Angeles, California, 06-11 August 95, pp 29–38.

[SCG00] Szenberg, F.; Carvalho,P.C. e Gattass,M. Automatic Camera Calibration for Image Sequences of a Football Match. *Interactive Conference on Pattern Recognition*. Rio de Janeiro, RJ, Brasil, Março de 2000.

[CSG98] Szenberg,F.; Carvalho,P.C. and Gattass,M. Image-based modeling using a two-step camera calibration method. In *Proceedings of International Symposium on Computer Graphics, Image Processing and Vision*, 1998, pp 388–395.

[DTM96] Taylor,C.J.; Debevec,P. and Malik,J. Modeling and rendering architecture from photographs: A hybrid geometry-and image-based approach. In Holly Rushmeier, editor, *Conference Proceedings (SIGGRAPH ' 96)*, ACM SIGGRAPH, Addison Wesley, Annual Conference Series, New Orleans, Louisiana, 04-09 August 1996, pp 11–20.

[TGTL91] Tommaselli,A. and Tozzi, C. Tutorial:calibração de câmeras usando feições geométricas. In *IV Simpósio Brasileiro de Computação Gráfica e Processamento de Imagens*, SIBGRAPI'91, Centro de Convenções Rebouças, São Paulo, 14-17 July 1991.

[VCG00] Villa Alvarez,B.; Carvalho,P.C. and Gattass,M. `http:www.tecgraf.puc-rio.br/~bia/paper/3dedition.html`