Single-View Metrology: Algorithms and Applications

Antonio Criminisi

invited paper
Microsoft Research, One Microsoft Way, Redmond, WA
`antcrim@microsoft.com`

Abstract. This paper addresses the problem of extracting three-dimensional geometric information from a single, uncalibrated image of a scene.
This work, building upon [7], is divided into two parts. The first part describes, in simple steps, the basic algorithms to obtain partial or complete geometric reconstruction from single perspective images of a scene. The second part presents a panorama of applications of single-view metrology and discusses its relationship with different disciplines such as architecture, history of art and forensic science. Furthermore, techniques for increasing the level of automation of the reconstruction process are herein described.
Several examples on photographs and historical paintings demonstrate the power and flexibility of the proposed techniques.

1 Introduction

This paper aims at describing, in a coherent framework, simple and effective algorithms for extracting geometric information and constructing compelling three-dimensional models from single, uncalibrated perspective images of a scene. Furthermore, this paper discusses the relationship of single-view metrology tecnhiques with other disciplines such as architecture, forensic science and history of art.

When only one view of a scene (either real or imaginary) is available, multi-view geometry algorithms [11, 16, 19] cannot be applied to construct three-dimensional models of the oberved scene. Recently, novel techniques for the partial or complete reconstruction of scenes from single images have been developed [7, 21, 24, 28, 30, 31, 35]. The main challenge of such algorithms lies in correctly modeling the perspective distortions introduced by the imaging process from a single input image (fig. 1).

We are mainly concerned with two canonical types of measurement: (i) lenghts of segments on planar surfaces and (ii) distances of points from planes. In many cases, these two kinds of measurements are proved to be sufficient for a partial or complete three-dimensional reconstruction of the observed scene.

The proposed algorithms have been designed to work in an uncalibrated framework (i.e. no need for the camera pose or internal parameters to be known or computed). On the other hand, scene constraints such as orthogonality and parallelism of structures are exploited, thus making our algorithms especially suitable for scenes containing man-made structures such as architectural elements and geometric patterns.

The ideas in this paper can be seen as reversing the rules for drawing perspective images laid out for the first time by Leon Battista Alberti in his treatise on linear perspective [1].
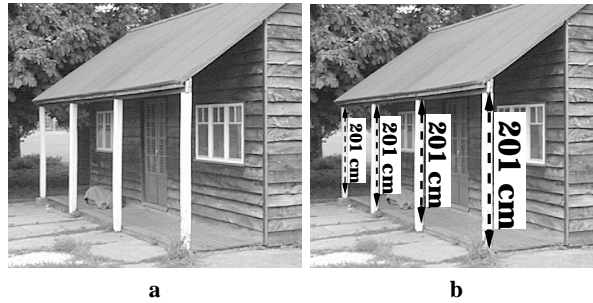
Fig. 1. Modeling perspective distortions in single images: **(a)** the four pillars have the same height in the world, although their images clearly are not of the same length because of perspective effects; **(b)** as shown, however, all pillars are correctly measured to have the same height. The perspective distortion has been removed.
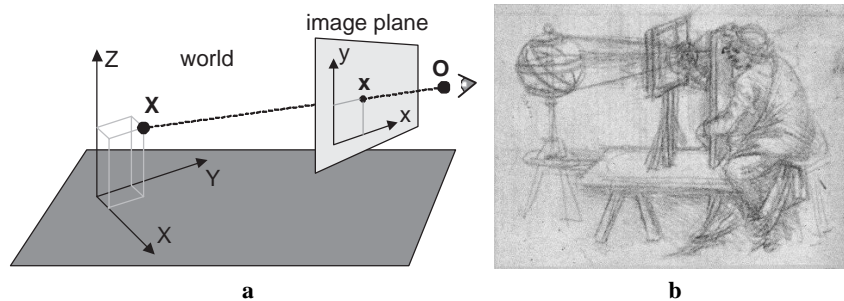

Fig. 2. The pinhole camera model. **(a)** A point $\mathbf{X}$ in the three-dimensional space is imaged as $\mathbf{x}$. Euclidean coordinates X, Y, Z and x, y are used for the world and image reference systems, respectively. $\mathbf{O}$ is the centre of projection, i.e. the camera optical centre. **(b)** Leonardo's *Perspectograph* (detail), by Leonardo da Vinci (1452–1519), Codex Atlanticus c.5r. Property of the Ambrosian Library, Milano. The similarity between figure (a) and (b) is striking.

2 The basic algorithm

This section describes, in simple terms, the basic techniques for: (i) measuring lengths of segments on planar surfaces and (ii) distances of points from planar surfaces. Clear examples and step-by-step algorithms make the implementation of these techniques straightforward[1].

The camera model employed here is central projection (see fig. 2). Effects such as radial distortion, which corrupt the central projection model, can generally be removed [14] and are therefore not detrimental to these methods.

---

[1] The reader is referred to [7] for a more general and comprehensive description of the single-view techniques.
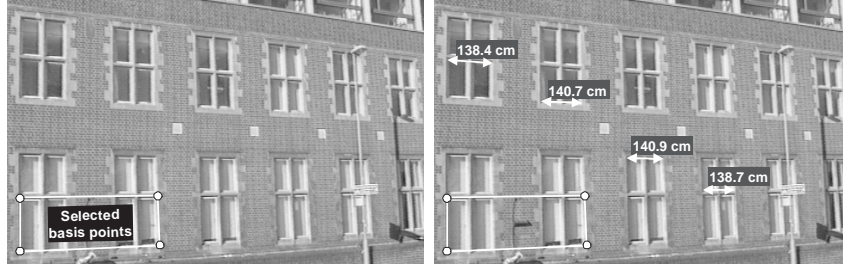
Fig. 3. Measuring distances on planar surfaces: **(Left)** A photograph of a wall of a building in Oxford. The position of four points have been manually measured in the world and the corresponding image points selected and marked in white on the image plane. **(Right)** Once the image-to-world homography has been computed (see appendix), measurements can be taken on the building wall as described in **Algorithm 1**.

2.1 Planar measurements

Given an image of a planar surface π, points on the image plane can be mapped into corresponding points in the world plane by means of a projective transformation called *homography* [19].

Points in one plane are mapped into the corresponding points in the other plane as follows:

$$\mathbf{X} = \mathtt{H}\mathbf{x} \tag{1}$$

where $\mathbf{x}$ is an image point, $\mathbf{X}$ is the corresponding point on the world plane (both expressed in homogeneous coordinates) and $\mathtt{H}$ is the 3×3 matrix representing the homography transformation.

Therefore, once the homography matrix $\mathtt{H}$ is known (or has been computed), any image point can be mapped into the corresponding location on the world surface and distances between world points can be extracted as illustrated below.

Algorithm 1: planar measurements.

1. Given an image of a planar surface estimate the image-to-world homography matrix $\mathtt{H}$;
2. Repeat
    (a) Select two points $\mathbf{x}_1$ and $\mathbf{x}_2$ on the image plane;
    (b) Back-project each image point into the world plane via (1) to obtain the two world points $\mathbf{X}_1$ and $\mathbf{X}_2$;
    (c) Compute the Euclidean distance $d(\mathbf{X}_1, \mathbf{X}_2)$.

Thus, the only remaining problem is that of estimating the homography matrix $\mathtt{H}$ (point 1 in **Alg.1**). The homography may be computed directly from a set of at least four corresponding points as described in the appendix. A statistical analysis of the measurements accuracy may be found in [7]. Figure 3 shows an example where windows of a building wall are measured directly on the image. Furthermore, the computed homography may also be used to rectify images of slanted planar surfaces into front-on views as demonstrated in fig. 11d [9].

2.2 Measuring distances from planes

This section addresses the problem of measuring distances of points *from* planes in the usual uncalibrated framework.

Figure 4a,b describes the problem in a schematic way. The aim is to compute the height of an object (the man in the figure) relative to a reference (the height of the column). Here, we assume that the vanishing line of the ground plane has been computed [2].

If $\mathbf{v}$ is the vanishing point for the vertical direction, $\mathbf{l}$ is the vanishing line of the ground plane, $\mathbf{t}_r$ and $\mathbf{b}_r$ are the top and base points of the reference, respectively and $\mathbf{t}_x$ and $\mathbf{b}_x$ are the top and base points of the object to be measured, then the following equation holds:

$$\alpha Z_i = -\frac{||\mathbf{b}_i \times \mathbf{t}_i||}{(\mathbf{l} \cdot \mathbf{b}_i)||\mathbf{v} \times \mathbf{t}_i||} \quad \forall i = r, x \tag{2}$$

where Z_x the height of the object we wish to measure, Z_r is the reference height and α a scalar quantity herein referred to as *metric factor*. Since αZ_i scales linearly we have obtained affine structure. If α is known, then a metric value for the height Z is obtained. Conversely, if the height Z is known then equation (2) provides a way of computing α and hence removing the affine ambiguity. Proof for (2) may be found in [7].

The complete algorithm for height computation from single images is described below, and examples of the computations are shown in fig. 4d.

Algorithm 2: computing heights of objects in single views.

1. Estimate the vanishing point $\mathbf{v}$ for the vertical direction;
2. Estimate the vanishing line $\mathbf{l}$ of the reference plane;
3. Select top and base points of the reference segment (points $\mathbf{t}_r$ and $\mathbf{b}_r$, respectively);
4. Compute the metric factor α by applying: $\alpha = -\frac{||\mathbf{b}_r \times \mathbf{t}_r||}{Z_r(\mathbf{l} \cdot \mathbf{b}_r)||\mathbf{v} \times \mathbf{t}_r||}$;
5. Repeat
    (a) Select top and base of the object to measure (points $\mathbf{t}_x$ and $\mathbf{b}_x$, respectively);
    (b) Compute the height Z_x by applying: $Z_x = -\frac{||\mathbf{b}_x \times \mathbf{t}_x||}{\alpha(\mathbf{l} \cdot \mathbf{b}_x)||\mathbf{v} \times \mathbf{t}_x||}$;

The key to the success of this algorithm is an accurate estimation of the vertical vanishing point $\mathbf{v}$ and the vanishing line $\mathbf{l}$ of the reference plane. The following section describes a simple technique for the automatic computation of vanishing points and lines as well as providing useful links to other techniques in the literature.

Estimating vanishing points and lines. Given an uncalibrated input image, vanishing points and vanishing lines may be computed either from the image-to-world homography H (if known) or by applying automatic and semiautomatic techniques which work directly on the image plane [6, 10, 13, 25–27, 32–34].

Here a simple RANSAC-based algorithm is employed to automatically estimate dominant vanishing points and lines [3].

---

[2] For the purpose of this section the vanishing line of the ground plane suffices; i.e. a full metric calibration of the ground plane is not necessary.

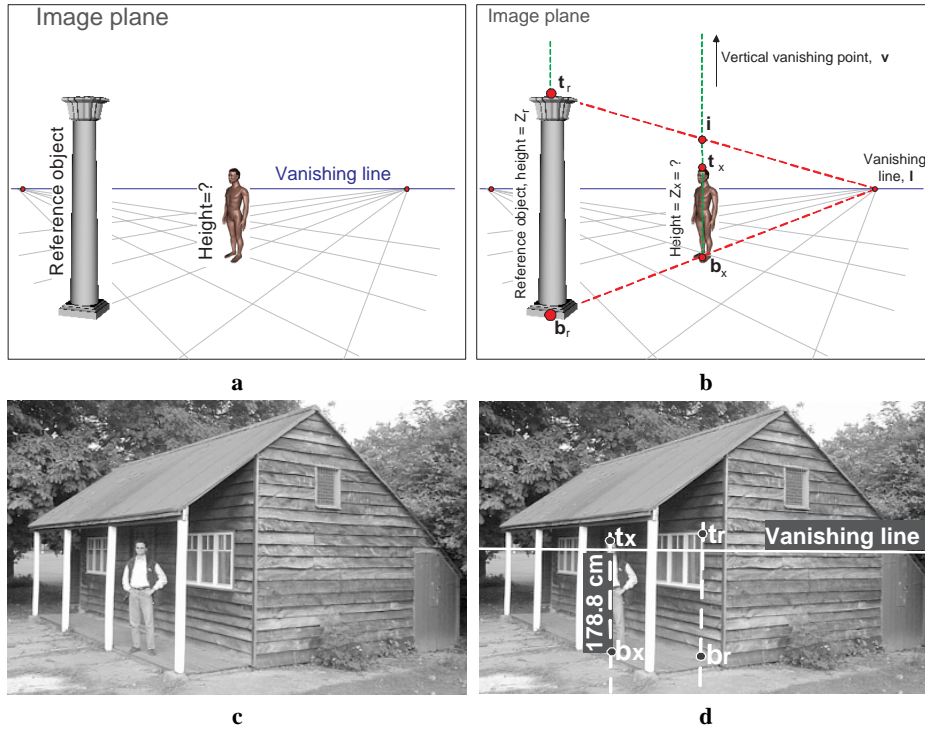[3] A more detailed and comprehensive description of this algorithm may be found in [32]

Fig. 4. Measuring heights in single images. **(a)** The aim is to compute the height of the human figure relative to the height of the column (reference). The vanishing line of the ground plane has been computed and is shown in white. **(b)** The unknown height ratio $\frac{Z_x}{Z_r}$ can be computed from image quantities only. See **Alg.2** for details. **(c)** A photograph of a garden shed in Oxford. **(d)** Once the height of the window top edge from the floor has been measured (reference), the height of the man is computed to be $178.8cm$; about $1cm$ off the ground truth.

The algorithm can be outlined as follows (*cf.* fig. 5):

1. Automatic Canny edge detection and straight line fitting to obtain the set of straight edge segments $\mathcal{E}$ (fig. 5b) [4];
2. Repeat
   (a) Randomly select two segments $\mathbf{s}_1, \mathbf{s}_2 \in \mathcal{E}$ and intersect them to give the point $\mathbf{p}$;
   (b) The support set $\mathcal{S}_p$ is the set of straight edges in $\mathcal{E}$ going through the point $\mathbf{p}$;
3. Set the dominant vanishing point as the point $\mathbf{p}$ with the largest support $\mathcal{S}_p$;
4. Remove all edges in $\mathcal{S}_p$ from $\mathcal{E}$ and goto 2 for the computation of the next vanishing point.

Different metrics may be used to decide when a straight line $\mathbf{s}$ goes through a given point $\mathbf{p}$. The Euclidean distance of a point from a line has been used herein; thus, $\mathbf{s} \in \mathcal{S}_p$ iff $d(\mathbf{p}, \mathbf{s}) < \sigma$, where σ is a fixed distance threshold ($\sigma = 2pix$ is used here)[4].

---

[4] Notice that this algorithm groups together lines which are parallel to each other in the scene (same vanishing point in the image), regardless of their coplanarity.
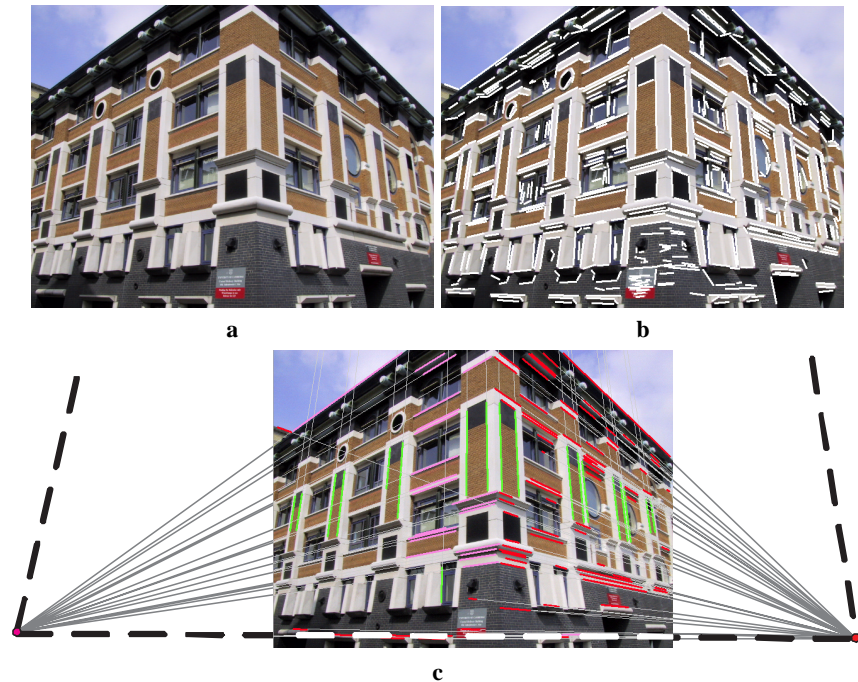
Fig. 5. Automatic computation of vanishing points and lines. (a) A photo of a building in Cambridge. (b) The automatically computed straight edges are superimposed (in white). (c) Automatic computation of the three dominant vanishing points. Edges of the same colour intersect in the same vanishing point. The three vanishing lines (thick dashed lines) are defined by joining the three pairs of automatically estimated vanishing points.

In fig. 5c vanishing lines are defined by joining pairs of vanishing points. Algorithms for the maximum likelihood estimation of vanishing points and lines are described in [32].

2.3 Constructing 3D models

In the previous sections the basic algorithms to extract planar and off-plane distance measurements from single images have been described. In order to construct complete three-dimensional models two more ingredients are necessary: (i) segmentation of scene objects and (ii) filling of occluded areas.

Given an input image, meaningful objects, such as planar walls and human figures, need to be segmented, measured and placed in the output model consistently with the three-dimensional scene geometry.

Object segmentation is achieved, here, by interactive silhouette cut-out.

Several techniques have been investigated in the past [2, 5, 29]. Amongst those, the dynamic programming-type algorithms cast the problem of estimating the contour be-
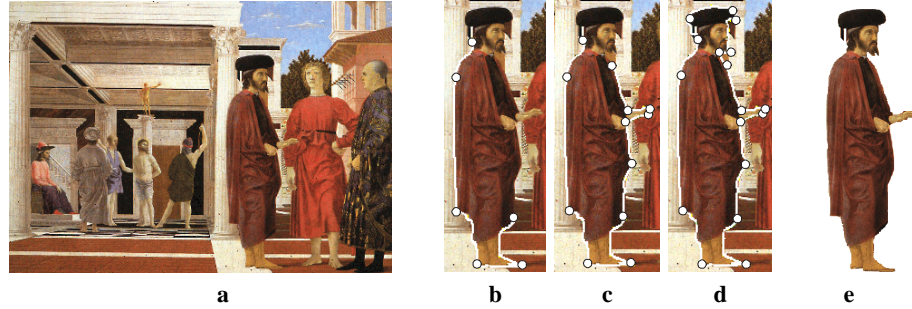
Fig. 6. Interactive silhouette cut-out. **(a)** The original painting. **(b,c,d)** Three stages of the interactive silhouette cut-out process. The dots indicate the user clicks, the white curve indicates the automatically estimated best contour. **(e)** The extracted silhouette.

tween two user-specified points as one of finding the optimal path between them. The technique employed here, based on a Viterbi algorithm [5] [17], can be thought of as a simple variant on the dynamic-programming methods. The costs of the edges of the Viterbi diagram are defined in a typical minimum cumulative way where the incremental cost associated to each pair of points in consecutive columns is given by:

$$Cost = \frac{1 - Ncc}{2} * w; \quad \text{with} \quad Ncc = \frac{\sum_{\Omega}(I(x,y) - \mu) * (I(x+u, y+v) - \mu')}{\sqrt{\sum_{\Omega}(I(x,y) - \mu)^2 \sum_{\Omega}(I(x+u, y+v) - \mu')^2}}$$

where Ncc is the normalized cross-correlation between two patches centred at (x, y) and $(x + u, y + v)$, respectively, μ and μ' are the average intensities for the two patches and w is a smoothing weight which tends to discourage very sharp changes in the contour curvature. The normalized cross-correlation is computed over patches Ω of fixed size (generally 3×3).

The use of normalized cross-correlation is justified by the observation that a contour can be thought of as a one dimensional curve that separates two dissimilar regions and such that points along the curve are "locally" similar in terms of the texture in their neighbourhood. The normalized cross-correlation measure in the costs of the Viterbi edges tends to constrain the extracted contour to follow peaks in the gradient map of an image, without the explicit computation of the image gradient.

This algorithm has the added benefit of being simple and easily implemented. Often, this technique is sufficient for a quick interactive silhouette cut-out (fig. 6), but does carry some drawbacks, namely: (i) as in most dynamic-programming approaches, incorporating dynamics and smoothness priors is not an easy task and (ii) the extracted silhouettes are restricted to pixel precision. These difficulties are overcome by the powerful particle-filtering technique described in [29].

Occlusion filling. In order to achieve visually compelling 3D models it is also necessary to fill-in occluded areas in an "undetectable" way. Two main techniques exist: (i)
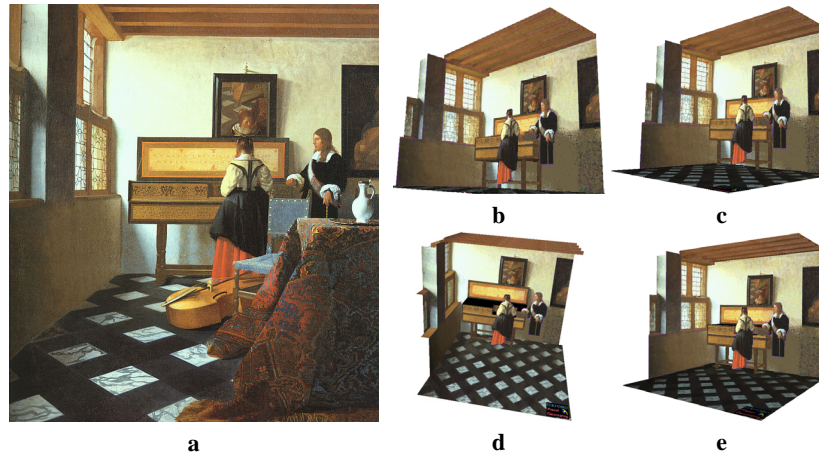
---

[5] http://www.sonic.net/~ejr/viterbi/viterbi.htm

Fig. 7. Three-dimensional reconstruction of a painting. (**a**) *The Music Lesson* (1662-65), by J. Vermeer (1632-1675). (**b-e**) Snapshots of a virtual fly-through inside the reconstructed painting to show different views of the reconstructed room.

Fill-in by exploiting symmetries and pattern regularities [8, 32] and (ii) Non-parametric texture synthesis [15, 20].

The first set of algorithms applies to regular geometric patterns and is used, in fig. 7d, to recover areas of the floor which where hidden in the original view.

Instead, non-parametric texture synthesis algorithms prove more useful for synthesizing stochastic (or generally less regular) textures. However, it is important to notice that these techniques cannot be applied directly to images showing strong perspective distortions, and a preliminary rectification of slanted planar surfaces is necessary.

The algorithm for the construction of complete three-dimensional models is outlined below and an example of complete reconstruction is shown in fig. 7.

Algorithm 3: complete 3D reconstruction

1. *Reference Plane Calibration:* select a reference plane and estimate the homography H (**Alg.1**);
2. *Height Calibration:* select a reference height and compute the metric factor α (**Alg.2**);
3. Repeat
     – Segment an object and measure its height and position on the reference plane;
     – Fill-in areas occluded by the selected object;
     – Insert the selected object in the output three-dimensional model.

---

3   Applications and interdisciplinarity

The first part of this paper has described the basic algorithms for a partial or complete geometric reconstruction from single images. This second part discusses possible applications of single-view techniques and their relationships with other disciplines such as architecture, forensic science and history of art.
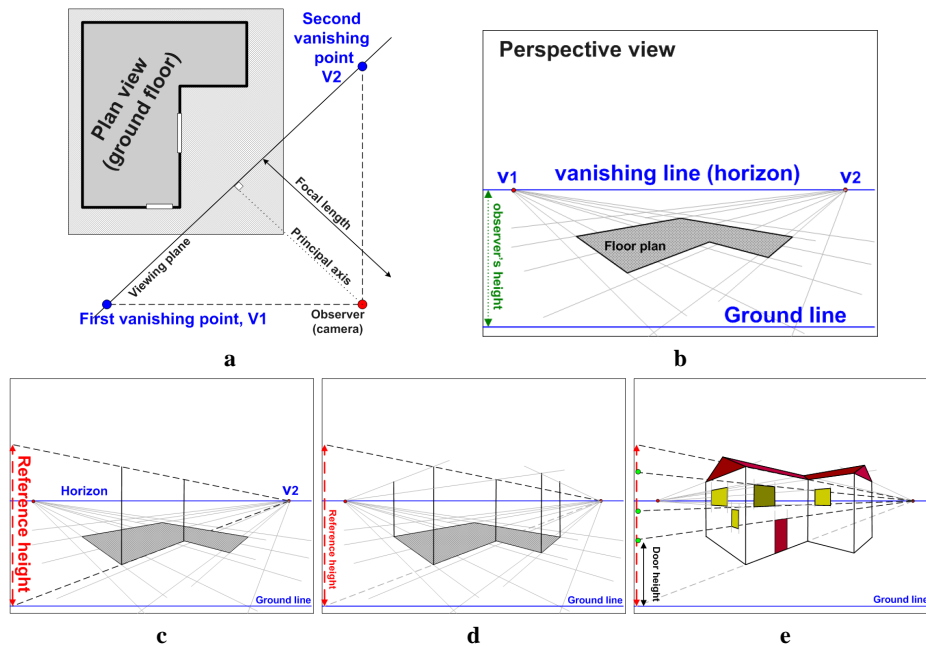
Fig. 8. Constructing a perspective image of a house. (a) Drawing the floor plan and defining the viewing conditions (observer position and image plane). (b) Constructing a perspective view of the floor. (c) A reference height (in this case the height of an external wall) is drawn from the ground line and the first wall is constructed in perspective by joining the reference end points to the horizontal vanishing point $\mathbf{v}_2$. (d) All four external walls are constructed. (e) The elevations of all other objects (the door, windows and roofs) are first defined on the reference segment and then constructed in the rendered perspective view.

3.1 Architectural drawing

Often architects need to create perspective views of three dimensional objects such as buildings or indoor environments on paper. This section describes the basic procedure for constructing perspective images (drawings) of three-dimensional objects [6] and its relationship to single-view metrology.

Figure 8 shows the process of constructing a perspective view of a house starting from its three-dimensional measurements. The basic steps of such procedure may be summarized as:

– **AD 1.** Draw a plan view of the ground floor (fig. 8a);
– **AD 2.** Set the viewing conditions: observer position and orientation, focal length, and viewing plane (fig. 8a);
– **AD 3.** Construct a perspective view of the ground plane (the reference plane, fig. 8b);
– **AD 4.** Draw a reference height and construct all elevations in perspective (fig. 8c,e).

[6] The interested reader may find useful reading material in any text book on technical drawing and descriptive geometry [3, 18].
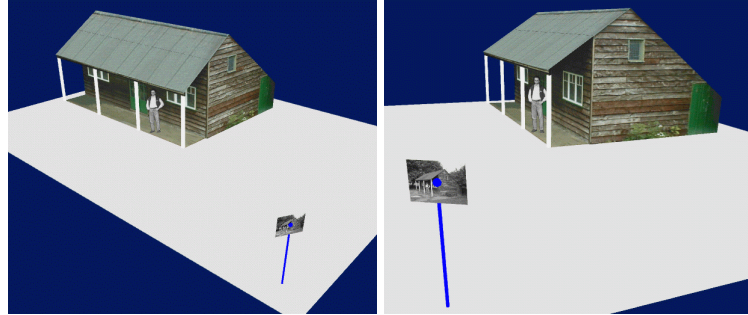
Fig. 9. Complete 3D reconstruction from single views. Two snapshots of the reconstructed three-dimensional model reconstructed from fig. 4c. The camera pose has, also, been estimated.

Instead, single-view metrology algorithms compute three-dimensional measurements from flat images. The procedure may be summarised as follows:

– **SVM 1.** Select a planar surface directly on the input perspective image (this can be viewed as the inverse of **AD 1**);
– **SVM 2.** Rectify the reference plane and take distance measurements (inverse of **AD 3**);
– **SVM 3.** Select a reference height on the image plane and estimate the height of any other object directly in the input image (inverse of **AD 4**);
– **SVM 4.** Estimate the camera pose and intrinsic parameters [35] (inverse of **AD 2**).

A comparison of the above procedures shows that single-view metrology can be seen as the technique inverting the long estabilished rules of linear perspective and descriptive geometry through the powerful algebraic modeling provided by projective geometry. Notice that in general, architectural drawing assumes infinite vertical vanishing point and horizontal vanishing line of the reference plane. Such limitations do not exist in the single-view metrology framework. An example of three-dimensional reconstruction from a single photograph of an architectural structure is shown in fig. 9.

Nowadays, architectural rendering is no longer done manually. Sofisticated CAD programs have replaced the architect's drafting, but the underlying construction steps remain as before.

3.2 Forensic investigation

Single-view metrology may also be used to analyse forensic imagery. A common requirement in surveillance images is to obtain measurements from the scene, such as the height of a suspect. Even when the suspect is no longer present in the scene, reference lengths can be measured from fixtures such as tables and windows.

An example of heights measurements is shown in fig. 10. Figure 10a is the input image, taken from a poor-quality security camera. In fig. 10b the input image has been corrected for radial distortion and the floor taken as the reference plane. After estimating the vertical vanishing point and the vanishing line of the ground plane, the height of the man has been computed from three known references. Details on the optimal use of multiple references may be found in [7].

Fig. 10. Measuring heights of people from single views. **(a)** Original photograph. **(b)** Image corrected for radial distortion and measurements superimposed. Three reference heights (marked with white segments) have been used and the man height has been measured to be $Z = 190.4 \pm 3.27$ cm. The uncertainty on the measurements has been estimated according to [7].

Single-view metrology techniques are currently used by forensic agencies in crime investigation applications.

3.3 Art History

Finally, in this section single-view metrology is used for analysing the geometry of paintings. Further details may be found in [8].

Comparing heights of people in paintings. *Flagellation* (in fig. 11a) by Piero della Francesca is one of the most studied paintings from the Italian Renaissance period. The "obsessive" accuracy of its geometry makes it one of the most mathematically rewarding paintings for detailed analysis purposes.

In fig. 11b the metrology algorithms described in the first part of this paper have been applied to compute the heights of the people in the painting. Due to the lack of an absolute reference the heights have been computed relative to a chosen unit reference, which in this case is the height of Christ. Therefore, height measurements are expressed as percentage variations from the height of Christ. Despite little variations, the measurements are all satisfactory consistent with each other, thus confirming the extreme accuracy and care in details for which Piero della Francesca has become famed [12].

Analysing shapes and patterns. This section demonstrates generation of new views of portions of paintings to better investigate the shape of patterns of interest.

The painting in figure 11a shows, an interesting black and white floor pattern viewed at a grazing angle (fig. 11c). Kemp in [22, 23] has manually analysed the shape of the pattern and demonstrated that it follows the "square root of two" rule. Figure 11d shows the rectification achieved by applying our homography-based technique (section 2.1) to obtain a front-on view of the floor. The result of automatic rectification is strikingly similar to the manual rectification in [22] but has the added advantage of being much faster and allowing retention of the original colour and shading. Furthermore, our computer rectification reveals a second instance of the same geometric pattern (on the top part of fig. 11d). A complete 3D reconstruction of this painting may be found in [7].
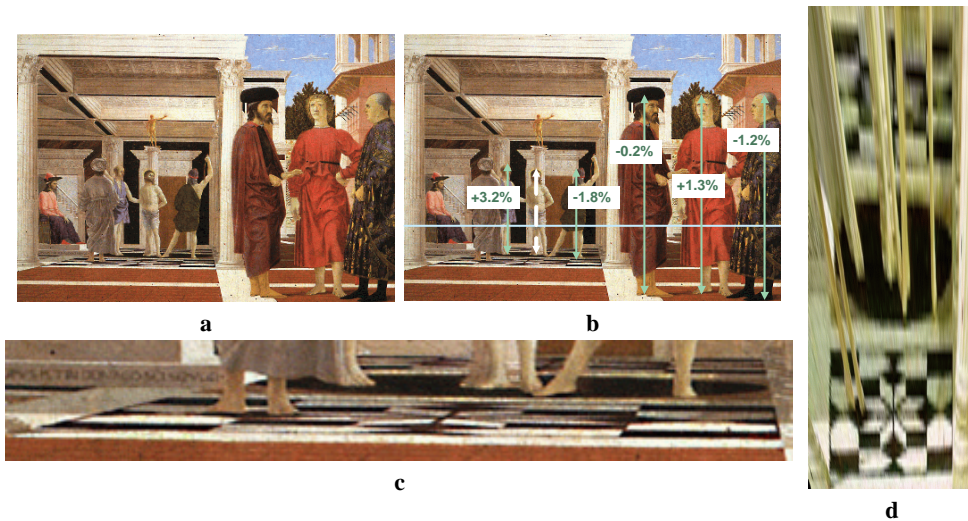
Fig. 11. Comparing heights of people in a Renaissance painting. (**a**) The original painting: *Flagellation* (approx. 1453), by Piero della Francesca (1416–92), Galleria Nazionale delle Marche, Urbino, Italia. (**b**) Heights of people have been measured relative to the height of Christ. They are expressed in percentage difference. (**c**) Enlarged image of the black and white floor pattern. (**d**) Automatic rectification of the floor into a front-to-parallel view. The rectified image has been obtained by applying our planar warping algorithm to the image of the painting directly. Note the striking similarity between the rectified pattern and that obtained by manual drafting by Kemp [8, 22]. A complete 3D reconstruction of this painting may be found in [7].

Analysing geometric ambiguities. The church of Santa Maria Novella, in Florence boasts one of Masaccio's best known frescoes, *The Trinity* (fig. 12a). The fresco is the first fully-developed perspectival painting from the Renaissance to use geometry to set up an illusion in relation to the spectator's viewpoint.

Single-view reconstruction algorithms were applied to an electronic image of the fresco to achieve a three-dimensional model of the chapel (fig. 12b,e) and to help art historians reach a consensus over debated disputes such as the relationship between the shape of the floor plan and the entabulatures of the chapel's vault.

In fact, since only one image is used and no scene metric information is available (the chapel is not real), an ambiguity arises in the reconstruction: it is not possible to uniquely recover the depth of the chapel without making some assumptions about the geometry of the scene.

Two plausible assumptions may be made: either the coffers on the vault of the chapel are square or the floor is square. The application of our single-view techniques has demonstrated that the two assumptions cannot coexist [8], i.e. square coffers imply a rectangular ground plan and vice-versa. Here the two models stemming from the two assumptions have been generated. Once the first model was constructed, the second one was obtained by applying a simple "affine transformation", a scaling in the direction orthogonal to the plane of the fresco.

Fig. 12. Three-dimensional reconstruction of Masaccio's *Trinity*. (**a**) The original fresco: *The Trinity* (approx. 1426), by Masaccio (1401–1428), Santa Maria Novella, Florence. (**b-e**) Different views of the reconstructed three-dimensional model of the chapel in the Florentine fresco.

The images of the chapel floor and that of the vault pattern shown in fig. 13 for both cases demonstrate that the square-ground-plan assumption yields rectangular angular coffers and the square-coffers assumption yields a rectangular ground plan. The advantage in terms of speed and accuracy over manual techniques is blatant.

Whatever the reason for Masaccio's ambiguity, the computer analysis performed has allowed both assumptions to be investigated in a rigorous and efficient manner.

The work in [8] shows further examples of analysing paintings using single-view techniques and presents an interactive virtual museum where the observer can not only move freely within the museum and look at the paintings, but even "dive" into the three-dimensional scenes reconstructed behind the plane of the canvas in a smooth and seemless way. This demonstrates the viability of our techniques in achieving compelling visual experiences which may be used to teach art students and art lovers about the power of linear perspective and its use in the Renaissance period.

4   Conclusion

This paper has presented easy-to-implement techniques to turn flat images into three-dimensional models and take distance measurements directly on the image plane.

The second part of this document has shown applications of our techniques to solve real problems such as meauring the height of a suspect in forensic images, or help resolve disputes over historical paintings. Furthermore, single-view metrology has been shown to be the inverse of the process of creating perspective architectural drawings.

Currently, we are planning to augment the flexibility and ease of use of single-view techniques by increasing the level of automation of the basic algorithms such as scene calibration and object segmentation.
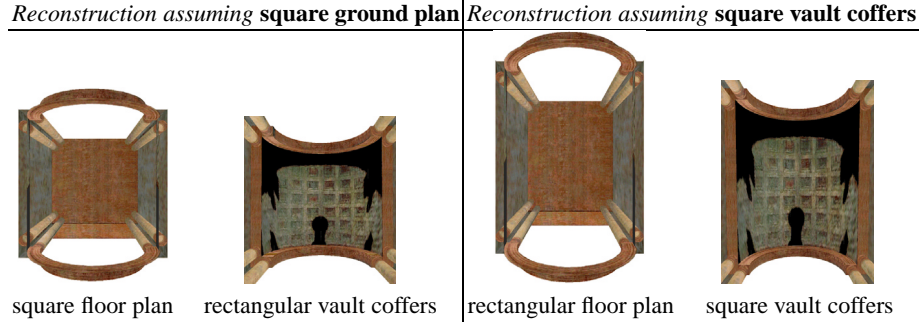
Reconstruction assuming **square ground plan**	*Reconstruction assuming* **square vault coffers**

| square floor plan | rectangular vault coffers | rectangular floor plan | square vault coffers |

Fig. 13. Ambiguity in reconstructing the depth of the chapel in Masaccio's *Trinity*. Comparing two possible reconstructions from an infinite set of plausible ones. (**Left**) Assuming a square ground plan leads to rectangular vault coffers and (**Right**) Assuming square vault coffers leads to a rectangular ground plan, thus demonstrating that ground plan and coffers cannot be both square.

Appendix: estimating the image-to-world homography

In the case of uncalibrated cameras, accurate estimation of the homography between the image and the world planes can be achieved directly from a set of known image-world correspondences (points or lines).

There are three standard methods for estimating the homography matrix H: (i) non-homogeneous linear solution; (ii) homogeneous solution; (iii) non-linear geometric solution. Only the second case is described herein. Details about the other cases may be found in [9].

Homogeneous solution. From (1) each image-to-world point correspondence provides two equations which are linear in the elements of the matrix H. They are:

$$h_{11}x + h_{12}y + h_{13} = h_{31}xX + h_{32}yX + h_{33}X$$
$$h_{21}x + h_{22}y + h_{23} = h_{31}xY + h_{32}yY + h_{33}Y$$

For n correspondences we obtain a system of $2n$ equations in eight unknowns. If $n = 4$ (as in fig. 14) then an exact solution is obtained. Otherwise, if $n > 4$, the matrix is over-determined, and H is estimated by a suitable minimization scheme.

The solution is obtained using Singular Value Decomposition (SVD). This method minimizes an algebraic error which does not have a geometric meaning. It is good practice to employ this method to obtain a reliable initial solution and, then run a non-linear minimization step to refine the solution by minimizing a more meaningful geometric error.
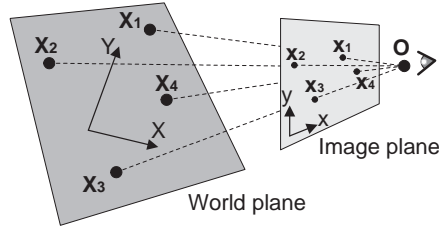
Fig. 14. Computing the plane-to-plane homography: at least four corresponding points (or lines) are necessary to determine the homography between two planes.

Writing the H matrix as a 9-vector $\mathbf{h} = (h_{11}, h_{12}, h_{13}, h_{21}, h_{22}, h_{23}, h_{31}, h_{32}, h_{33})^\top$ the homogeneous equation (1) for n points become $\mathtt{A}\mathbf{h} = \mathbf{0}$, with $\mathtt{A}$ the $2n \times 9$ matrix:

$$\mathtt{A} = \begin{pmatrix} x_1 & y_1 & 1 & 0 & 0 & 0 & -x_1 X_1 & -y_1 X_1 & -X_1 \\ 0 & 0 & 0 & x_1 & y_1 & 1 & -x_1 Y_1 & -y_1 Y_1 & -Y_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_n & y_n & 1 & 0 & 0 & 0 & -x_n X_n & -y_n X_n & -X_n \\ 0 & 0 & 0 & x_n & y_n & 1 & -x_n Y_n & -y_n Y_n & -Y_n \end{pmatrix}$$

The problem of computing the $\mathbf{h}$ vector is now reduced to the constrained minimization of the cost function $C = \mathbf{h}^\top \mathtt{A}^\top \mathtt{A} \mathbf{h}$ subject to the constraint that $\|\mathbf{h}\| = 1$. The corresponding Lagrange function is: $\mathcal{L} = \mathbf{h}^\top \mathtt{A}^\top \mathtt{A} \mathbf{h} - \lambda(\mathbf{h}^\top \mathbf{h} - 1)$. Differentiating this with respect to $\mathbf{h}$ and setting these derivatives equal to zero we obtain $\mathtt{A}^\top \mathtt{A} \mathbf{h} = \lambda \mathbf{h}$. Therefore the solution $\mathbf{h}$ is a unit eigenvector of the matrix $\mathtt{A}^\top \mathtt{A}$ and $\lambda = \mathbf{h}^\top \mathtt{A}^\top \mathtt{A} \mathbf{h}$ is the corresponding eigenvalue. In order to minimize the C function, only the eigenvector $\tilde{\mathbf{h}}$ corresponding to the minimum eigenvalue $\tilde{\lambda}$ should be considered. This eigenvector can be obtained directly from the Singular Value Decomposition of $\mathtt{A}$. In the case of $n = 4$, $\mathbf{h}$ is the null-vector of $\mathtt{A}$ and the residuals are zero.

References

1. Leon Battista Alberti. *De Pictura*. 1435. Reproduced by Laterza (1980).
2. W. Barrett and Mortensen. Interactive live-wire boundary extraction. *Medical Image Analysis*, 1(4):331–341, 1997.
3. H. C. Browning. *The Principles of Architectural Drafting : A Sourcebook of Techniques and Graphic Standards*. Whitney Library of Design, Oct 1996.
4. J. F. Canny. A computational approach to edge detection. *IEEE T-PAMI*, 8(6), 1986.
5. L. Cohen and R. Kimmel. Global minimum for active contour models: A minimum path approach. *IJCV*, 24(1):57–78, 1997.
6. R. T. Collins and R. S. Weiss. Vanishing point calculation as a statistical inference on the unit sphere. In *Proc. ICCV*, pages 400–403, Dec 1990.
7. A. Criminisi. *Accurate Visual Metrology from Single and Multiple Uncalibrated Images*. Distinguished Dissertation Series. Springer-Verlag London Ltd., Sep 2001. ISBN: 1852334681.
8. A. Criminisi, M. Kemp, and A. Zisserman. Bringing pictorial space to life: computer techniques for the analysis of paintings. Technical report, Microsoft Research, 2002.

9. A. Criminisi, I. Reid, and A. Zisserman. A plane measuring device. *Image and Vision Computing*, 17(8):625–634, 1999.

10. A. Criminisi and A. Zisserman. Shape from texture: homogeneity revisited. In *Proc. BMVC*, pages 82–91, UK, Sep 2000.

11. P. E. Debevec, C. J. Taylor, and J. Malik. Modeling and rendering architecture from photographs: A hybrid geometry- and image- based approach. In *Proceedings, ACM SIGGRAPH*, pages 11–20, 1996.

12. Piero della Francesca. *De Prospectiva Pingendi*. Firenze, Italy, 1474. Reproduced by ed. Sansoni (1942), Edizione Critica.

13. J. Deutscher, M. Isard, and J. MacCormick. Automatic camera calibration from a single manhattan image. In *Proc. ECCV*, page 175ff, Copenhagen, 2002.

14. F. Devernay and O. D. Faugeras. Automatic calibration and removal of distortion from scenes of structured environments. In *Proc. of SPIE*, volume 2567, San Diego, CA, Jul 1995.

15. A. Efros and W.T. Freeman. Image quilting for texture synthesis and transfer. In *Proc. ACM SIGGRAPH*, pages 341–346, Eugene Fiume, August 2001.

16. O. D. Faugeras. *Three-Dimensional Computer Vision: a Geometric Viewpoint*. MIT Press, 1993.

17. G. D. Forney. The viterbi algorithm. *Proc. IEEE*, 61:268–278, Mar 1973.

18. R. W. Gill. *Basic Perspective*. W.W. Norton and Company, Feb 1980.

19. R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN: 0521623049, 2000.

20. A. Hertzman, C. E. Jacobs, N. Oliver, B. Curless, and D. H. Salesin. Image analogies. In *Proc. ACM SIGGRAPH*, pages 341–346, Eugene Fiume, August 2001.

21. Y. Horry, K. Anjyo, and K. Arai. Tour into the picture: Using a spidery mesh interface to make animation from a single image. In *Proc. ACM SIGGRAPH*, pages 225–232, 1997.

22. M. Kemp. *The Science of Art*. Yale University Press, New Haven and London, 1989. ISBN: 0-300-05241-3.

23. M. Kemp. *Visualizations: the nature book of art and science*. The University of California Press, Berkeley and Los Angeles, California, USA, 2000. ISBN: 0-520-22352-7.

24. T. Kim, Y. Seo, and K. Hong. Physics-based 3D position analysis of a soccer ball from monocular image sequences. *Proc. ICCV*, pages 721 – 726, 1998.

25. J. Koseka and W. Zhang. Video compass. In *Proc. ECCV*, page 476ff, Copenhagen, 2002.

26. D. Liebowitz and A. Zisserman. Metric rectification for perspective images of planes. In *Proc. CVPR*, pages 482–488, Jun 1998.

27. G. F. McLean and D. Kotturi. Vanishing point detection by line clustering. *IEEE T-PAMI*, 17(11):1090–1095, 1995.

28. MetaCreations. http://www.metacreations.com/products/canoma/.

29. P. Perez, A. Blake, and M. Gangnet. JetStream: Probabilistic contour extraction with particles. *Proc. Int. Conf. on Computer Vision (ICCV)*, II:524–531, Mar 2001.

30. M. Proesmans, T. Tuytelaars, and L. J. van Gool. Monocular image measurements. Technical Report Improofs-M12T21/1/P, K.U.Leuven, 1998.

31. I. Reid and A. Zisserman. Goal-directed video metrology. In R. Cipolla and B. Buxton, editors, *Proc. ECCV*, volume II, pages 647–658. Springer, Apr 1996.

32. F. Schaffalitzky and A. Zisserman. Planar grouping for automatic detection of vanishing lines and points. *Image and Vision Computing*, 18:647–658, 2000.

33. J. A. Shufelt. Performance and analysis of vanishing point detection techniques. *IEEE T-PAMI*, 21(3):282–288, Mar 1999.

34. T. Tuytelaars, L. van Gool, M. Proesmans, and T. Moons. The cascaded Hough transform as an aid in aerial image interpretation. In *Proc. ICCV*, pages 67–72, Jan 1998.

35. M. Wilczkowiak, E. Boyer, and P. Sturm. Camera calibration and 3D reconstruction from single images using parallelepipeds. In *Proc. ICCV*, Vancouver, 2001.