

Low cost digital reconstruction procedure of 3D objects

G. Arcidiacono and F. Portuese
IR&T engineering, Catania, Italy

Computer graphics has made a fast progress in visualizing 3D objects, because the increasing in the storage and computational power of today computer. The range of applications of the 3D digital visualization spans from architecture and heritage conservation to any virtual environment. For cultural heritage, one useful application is the visualization of artistic objects through digital virtual museums, that give the ability to anybody to look monuments and other artistic assets, without necessarily go to the place in which they lie. All these digital products cause important demand for more complex and realistic models; however, a limitation to the widespread use of these techniques is the high costs of proposed standard solutions, which include currently the use of expansive devices as stereo rigs or laser scanners. Thanks to the growing computational power of modern computer, it's possible to reduce the hardware complexity of those systems, moving some important stages from hardware to software implementation, with a significant simplification of the 3D reconstruction procedure.

The system presented in this work is a passive approach that can be realized with low cost consumer image acquisition hardware. Initially a sequence of images can carry out from a commercial digital video device (resolution 1Mpixel), taking care to get pictures with enough overlap, typically a content change below 10-15%, between subsequent images. The image overlap is necessary to get the correspondence search for many interest points in two subsequent views. Knowing the camera position and orientation, conventional stereo algorithms can be used on the image in order to perform the reconstruction. Afterwards, between small surface portions, a stereo correspondence analysis is used to calculate dense depth images for each input view, including for every pixel the distance object-camera. Moreover to improve the image quality, multi-view-stereo matching algorithms are included, together specific interpolation procedures in order to get rid of small regions with either missing or incorrect depth data. After all mentioned steps, the full set of depth images are fused together to build a 3D-surface-mesh, where the original images are used as textures on the triangular mesh. The 3D-model is then exported and showed with any standard viewer or can be imported into 3D-CAD systems to get a realistic impression of the virtual reconstructed object.

APPROACH DESCRIPTION

Step 1: Image acquisition

Initially, the images are acquired via a digital video or photo camera:



A sufficient number of image must be taken in order to perform a full reconstruction of the scene.

The images must be acquired with a content change below 10-15% between two consecutive images in order to perform a reasonable robust feature matching (see next steps). The resolution of the images should be high enough (typically from 1 to 2 Mpixels).

Step 2: Feature extractions

Starting from a collection of images or a video sequence the next step consists in relating the different images to each other.

A restricted number of corresponding points is sufficient to determine the geometric relationship or multi-view constraints between the images. Since not all points are equally suited for matching or tracking (e.g. a pixel in a homogeneous region), the first step consist of selecting a number of interesting points or *feature points*.



One of the most important requirements for a feature point is that it can be differentiated from its neighboring image points.

To ensure that, the following second order *dissimilarity function* is used:

$$D(x, y) = \frac{1}{W} \int \int \frac{I(x, y)}{I(x, y)} w(x, y) dx dy$$

W is an image window and $w(x, y)$ is a weighting function defined over W .

Step 3: Feature matching

In the next step, the feature points detected in the first image are matched in the second image:



The search of the matching feature, in the second image, start from the feature's position found in the first picture. The range of search is restricted to a small area in order to improve performance and to avoid wrong matches (due to similar regions).

The following *similarity* measurement is used to compare two regions:

$$S = \frac{\int \int \frac{I(x, y)}{I(x, y)} w(x, y) dx dy}{\sqrt{\int \int \frac{I(x, y)}{I(x, y)} w(x, y) dx dy} \sqrt{\int \int \frac{I(x, y)}{I(x, y)} w(x, y) dx dy}}$$

With $\bar{I} = \frac{1}{W} \int \int I(x, y) dx dy$ and $\bar{I} = \frac{1}{W} \int \int I(x, y) dx dy$ the mean image intensity in the considered region. Note that this measure is invariant to global intensity and contrast changes over the considered regions.

Step 4: Fundamental Matrix estimation

Once the features are robustly matched, the projective camera setup of the views can be obtained through the *Fundamental Matrix*. The fundamental matrix (introduced by Faugeras and Hartley) describes the relationship between two matching points. If m is a point on the first image and m' is the matching point on the second image, we have the following relation:

$$m'^T F m = 0$$

where F is the fundamental matrix.

Every pair of corresponding points gives one constraint on F . Since F is a 3×3 matrix which is only determined up to scale, it has $3 \times 3 - 1$ unknowns. Therefore 8 pairs of corresponding points are sufficient to compute F with a linear algorithm. The precision of the estimation of the fundamental matrix is really important, so it's necessary a robust method to compute it. An important problem to consider is the presence of *outliers* (wrong matches). A solution to this problem was proposed by Fischler and Bolles. Their algorithm is called RANSAC (RANDOM SAMPLING CONSENSUS) and it can be applied to all kinds of problems.

From the fundamental matrix the projective camera setup can be obtained:

$$\begin{matrix} P_1 & I_{3 \times 3} & | & 0_3 \\ P_2 & e_{12} & F & e_{12} & | & e_{12} \end{matrix}$$

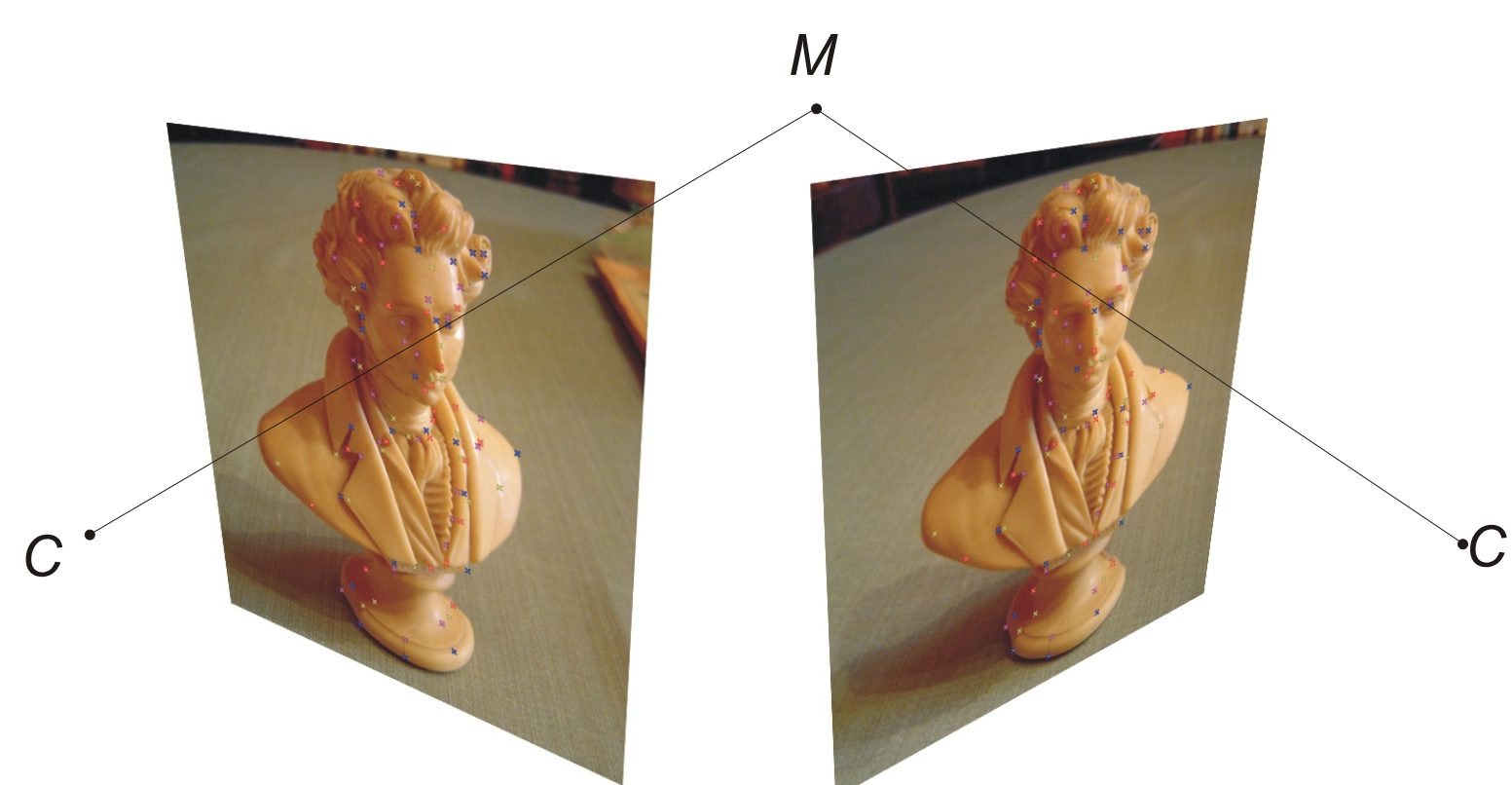
Where e_{12} is the projection of the projection center of the first camera in the second image (also called *epipole*).

Step 5: Triangulation and projective reconstruction

Once two projection matrices have been fully determined the matches can be reconstructed through triangulation. Due to noise the lines of sight will not intersect perfectly. In the uncalibrated case the minimizations should be carried out in the images and not in projective 3D space. Therefore, the distance between the reprojected 3D point and the image points should be minimized:

$$D(m_1, P_1 M)^2 + D(m_2, P_2 M)^2$$

where M is the reconstructed 3D point.



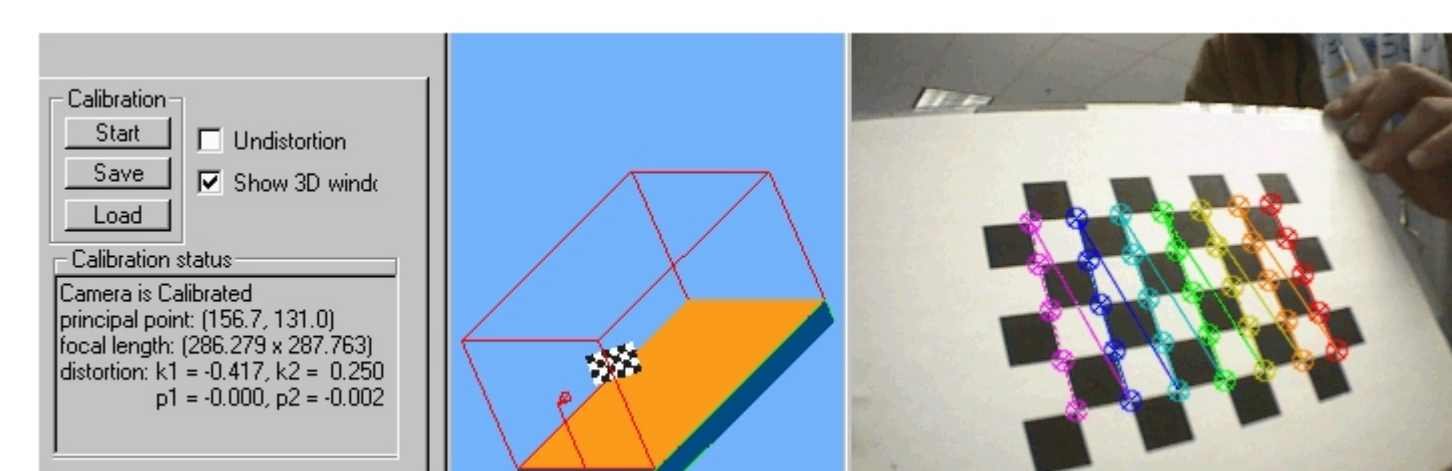
The reconstructed point M are then reprojected onto the images to detect further outliers.

Step 6: Self-Calibration

The reconstruction obtained in the previous step is only determined up to an arbitrary projective transformation. This might be sufficient for some robotics or inspection applications, but certainly not for visualization. Therefore we need a method to upgrade the reconstruction to a metric one (i.e. determined up to an arbitrary Euclidean transformation and a scale factor).

To achieve this we can apply constraints on the *intrinsic camera parameters*. These kind of parameters are constant for the specific acquiring device (like the aspect ratio, the principal point and the focal length). Reducing the ambiguity on the reconstruction by imposing restrictions on the intrinsic camera parameters is termed *self-calibration*.

The intrinsic camera parameters can be obtained through off-line calibration with a calibration object that have a known geometry (usually a chessboard pattern).



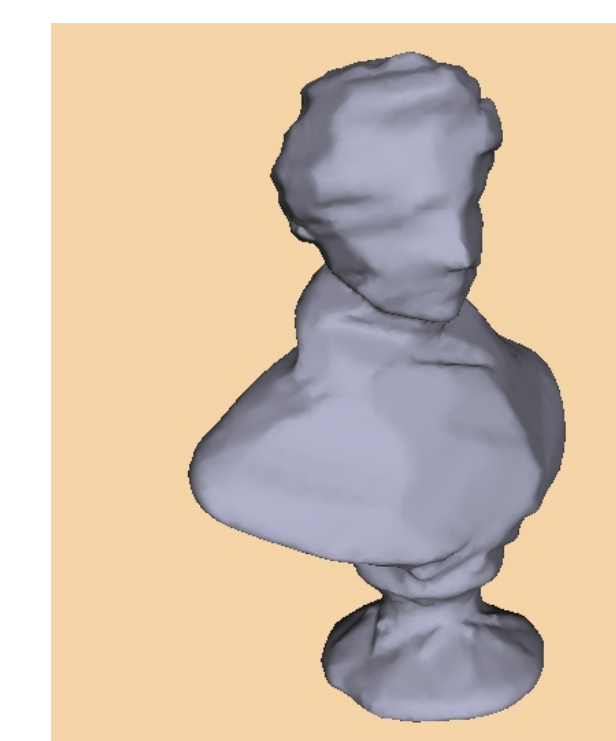
(Image taken from the Camera Calibration Tool of the Open Source Computer Vision Library)

A major drawback of this method is that some parameters (like the focal distance) may change due to operations like zoom and autofocus, so a more flexible way of doing self-calibration is under development (where most intrinsic parameters can vary).

Step 7: Surface modeling

With the camera calibration given for all viewpoints of the sequence, we can proceed with methods developed for calibrated structure from motion algorithms. The feature tracking algorithm already delivers a sparse surface model based on distinct feature points. This however is not sufficient to reconstruct geometrically correct and visually pleasing surface models. This task is accomplished by a dense disparity matching that estimates correspondences from the grey level images directly by exploiting additional geometrical constraints. The stereo matching problem can be solved much more efficiently if images are rectified.

The 3D surface is approximated by a triangular mesh to reduce geometric complexity and to tailor the model to the requirements of computer graphics visualization systems. The image itself can be used as texture map (the texture coordinates are trivially obtained as the 2D coordinates of the vertices).



Results

The following results are obtained through a digital photo camera (10 photos) with a resolution of 1 Megapixels.



The flexibility of the proposed systems allows applications in many domains, especially in cultural fruition (virtual museums).

Further enhancements to improve the precision of the reconstructed models are under development, including an automatic system to estimate the intrinsic parameters of the camera.

The hardware requirements are quite low, enabling the system to be transported easily in every place with really contained costs respect to the existing 3D scanners. In addition, the proposed system is not invasive making it ideal to handle antique objects.