Methodology for 3D scene reconstruction from digital camera images

V. Patrangenaru, M. A. Crane, X. Liu, X. Descombes, G. Derado, W. Liu, V. Balan, V. P. Patrangenaru, H. W. Thompson

Abstract. Digital images provide today an important source of data that deserves a careful statistical analysis. This paper concerns methods for retrieval of 3D information, including shape and texture, from cheap digital camera imaging outputs. It includes a three step reconstruction of a 3D scene with texture, from arbitrary partial views, in absence of occlusions. In Patrangenaru and Patrangenaru [23], Mardia et. al. [20] and Patrangenaru and Mardia [24] a planar scene was reconstructed using image fusion, around representatives of sample mean projective shapes or sample mean affine shapes of landmark configurations shared by a number of partial views of the scene. In this paper we first analyze the advantages and limitations of such a reconstruction of a close to planar remote scene from its partial aerial views, by specializing this algorithm to affine transformations. Furthermore, we combine a projective shape reconstruction of a finite 3D configuration from its uncalibrated camera views, as developed in Patrangenaru, Liu and Sughatadasa [22], with a VRML technique, to reconstruct projectively a 3D scene with texture from a pair of digital camera images, thus allowing a more detailed statistical analysis of the scene pictured. We give three such examples of 3D reconstructions.

M.S.C. 2010: 62H35, 62H10, 68U10, 62H11.
Key words: projective shape reconstruction; statistical analysis; texture; affine transformations.

1 Introduction

Gray level images are generated by matrices, with entries (corresponding to grey levels at pixel location) ranging on a scale from 0 to 1, that measures the brightness of the scene pictured. The combined effect of the brightness is perceived as a continuous digital image. Therefore, it is the reflectance of objects, and not the content, that dictates the appearance of a scene in a digital image. If the objective is analyzing the

BSG Proceedings, Vol. 19, 2012, pp. 110-124.

contents of a scene from digital images, its features would be localized by the high frequencies of the signal. Thus, studying the contents of a scene can be regarded as a *high level image analysis.* In this context we consider landmarks, labeled points, that have a significance on the scene, and their configuration yields a basic descriptor of the contents. Depending on the image acquisition principle, examples of descriptors are the projective shape, the affine shape or the similarity shape of the landmark configuration.

Often, the effect of the signal acquisition principle to the output image of the scene, may be described in terms pseudogroups of transformations. A family G of pairs (U, f_U), where U is an open subset of $\mathbb{R}^m$ and f_U is a diffeomorphism from U onto the subset $f_U(U)$ of $\mathbb{R}^m$, such that (i). for any element of G, U, f_U, its inverse $f_U(U), f_U^{-1})$ is in G, and (ii). for any pairs $(U, f_U), (V, f_V) \in \mathcal{G}$, the pairs $((f_U)^{-1}(V), f_U|(f_U)^{-1}(V))$ and $((f_U)^{-1}(V), f_V \circ f_U)$ are in G is said to be a *pseudogroup* of transformations of $\mathbb{R}^m$. A typical example of pseudogroup of transformations on $\mathbb{R}^m$, is given by the pairs (U, f_U), where f_U is the restriction of a map f defined by $y = f(x)$, with

$$y^j = \frac{\sum_i a_i^j x^i}{\sum_i a_0^j x^i}, \forall j = 1, ..., m \tag{1.1}$$

is such that $\det((a_i^j)_{i,j=0,...,m}) \neq 0$.

The general study of configurations of a system of k points in $\mathbb{R}^m$, where the effects of a pseudogroup $\mathcal{G}$ of transformations have been removed, was motivated by applications in machine vision involving the group of affine transformations of $\mathbb{R}^m$ (Sparr [27]), or the pseudogroup of projective transformations of $\mathbb{R}^m$ (Hartley and Zisserman [13, p. 91]). The $\mathcal{G}$ - shape of a subset K of $\mathbb{R}^m$, is the set of all orbits of K under transformations in $\mathcal{G}$. Such a type of set may consist in k labeled points, called *landmarks.* The resulting orbifold, quotient of $(\mathbb{R}^m)^k/\mathcal{G}$ was called the *$\mathcal{G}$-shape space of k-ads* (Patrangenaru and Patrangenaru [23]), and was labeled $\mathcal{G}\Sigma_m^k$. If the pseudo-group action is generated by a locally free group action of a Lie group G of dimension g, the dimension of the $\mathcal{G}$-shape space of k-ads is $mk - g$.

If $\mathcal{G}$ is the pseudogroup of transformations generated by affine transformations of $\mathbb{R}^m$, since the affine group has dimension m^2+m, the resulting $\mathcal{G}$-shape space of k-ads, also called the *affine shape space* is $m(k-m-1)$ dimensional. Sparr [27] showed that if one considers only k-ads in general position, the corresponding affine shape space is diffeomorphic to the real Grassmann manifold $G_m(k-1)$ of m-dimensional linear subspaces of $\mathbb{R}^{k-1}$ (see Heyden [14]). A landmark based investigation of projective invariants was first initiated by Maybank and Beardsley in [17] and by Goodall and Mardia [11]. Later on, the projective shape of configurations with the first $m + 2$ landmarks yielding a *projective frame*, was identified by Patrangenaru [25] with $(\mathbb{R}P^m)^{k-m-2}$. Note that the landmark coordinates are usually recorded manually or automatically from images, registration procedures that are subject to errors. Subsequently, one assumes that the observed $\mathcal{G}$-shapes yield a random sample $(X_1, ..., X_n)$ from a given probability distribution Q, and that the sample mean $\mathcal{G}$-shape is a consistent estimator of the true mean $\mathcal{G}$-shape of Q (see Bhattacharya and Patrangenaru [4]). This note should be considered in the context of statistics on manifolds, since the mean $\mathcal{G}$-shape is defined in terms of a distance considered on the $\mathcal{G}$-shape space, which is not a linear space. Also note that $\mathcal{G}\Sigma_m^k$ has always a singular point, $\mathcal{G}$-orbit of the

k-ad of identical points, and this singularity is automatically removed. In case the resulting $\mathcal{G}$-shape space has a manifold structure, a distance ρ can be obtained via any embedding j of this manifold into an Euclidean space, by pulling back the restriction of the Euclidean distance. The resulting *extrinsic (sample) mean* $\hat{\mu}_{E,j}$ will be used in the reconstruction. In the case of affine shape spaces, one may use an equivariant embedding. Dimitric [9] showed that an equivariant embedding j of $G_m(k-1)$ in the space $Sym(k-1)$ of $(k-1)\times(k-1)$ symmetric matrices endowed with the Euclidean square norm $\|A\|^2 = Tr(AA^t)$ is obtained by identifying each m-dimensional vector subspace L of $\mathbb{R}^{k-1}$ with the matrix p_L associated to the orthogonal projection into L. This embedding was considered in Bhattacharya and Patrangenaru [4] in the particular case of the real projective space. Assume a probability distribution Q of affine shapes of configurations in general position is *nonfocal* w.r.t. j. Patrangenaru and Mardia [24] showed that the extrinsic mean $\mu_j(Q)$ exists if the mean of $j(Q)$ of $(k-1)\times(k-1)$ has the eigenvalues $\lambda_1 \geq \cdots \geq \lambda_{k-1}$ such that $\lambda_m > \lambda_{m+1}$; in this case $\mu_j(Q)$ is the vector subspace spanned by eigenvectors corresponding to the first m eigenvalues of $\mu_{j(Q)}$. In particular, if $\pi_1, \ldots, \pi_n$ is a sample of size n of m-dimensional vector subspaces $\pi_1, \ldots, \pi_n$ of $\mathbb{R}^{k-1}$, and the subspace π_r is spanned by the orthonormal unit vectors $\{x_{r,a}\}_{a=1,\ldots,m}$ and if we set $x_r = (x_{r,a})_{a=1,\ldots,m}$, the j-extrinsic sample mean, when it exists, is the m dimensional vector subspace $\overline{\pi}_j$ generated by the unit eigenvectors corresponding to the first largest m eigenvalues of $\sum_{r=1}^n x_r x_r^t$.
On the other hand, in the case of projective shape spaces of k-ads, one may use equivariant embeddings of products of $q = k-m-2$ copies of $\mathbb{R}P^m$, and the resulting expression of the extrinsic sample mean (see Patrangenaru et. al. [22]).
While *fusing images around mean configurations* is a successful methodology for reconstruction of planar flat scenes (Patrangenaru and Mardia [24], Patrangenaru and Patrangenaru [23]), in this paper besides applying this technique to the general case when the scene pictured is almost flat (eg. aerial images of a city), we consider a more complex case, of reconstructing the surface a polyhedral object, or even of a curved scene in 3D, including texture from two of its arbitrary images, taken in absence of occlusions. This operation is completed in three steps. We first reconstruct a finite configuration of points, using the eight point algorithm in Ma et. al. [18, p. 121], then we build an wireframe from this configuration, and finally we add the texture from the two images to this wireframe. We give three such examples or reconstructions, one for a polyhedral object, one for an architectural scene, and another one for a medical imaging scene from a stereo camera pair of the eye fundus.

2 Reconstruction of a planar scene

Two ideal camera images of a remote flat scene approximately differ by an affine transformation,

$$\begin{cases} x' = a_1 x + a_2 y + a_3, \\ y' = b_1 x + b_2 y + b_3. \end{cases}$$

and images of a configuration pictured two such images have almost the same affine shape. Such an example is given in figure 1

Figure 1: Photographs of an airfield taken by a squadron.

For scene reconstruction we first represent the partial views of a remote scene, under an angle bounded as images on a larger canvas. Two views that partially overlap, are transformed via affine transformations by aligning a given configuration of landmarks, which coincide in their region of overlap, to yield a convenient representative of their extrinsic mean affine shape. This is method also known as the DLT method in computer vision (Hartley and Zisserman [13, p. 88]). The gray level at each pixel on the canvas is the mean gray level from the contributing images containing that pixel. If more than two images are considered, this extended image is built around the extrinsic mean affine shape of a configuration shared by those images. The practical algorithm Patrangenaru and Mardia [24] consists in starting with a view that is overlapping with the largest number of other images, and selecting convenient shared configurations, and resulting transformations to estimate the reconstructed scene, see Figure 2.
For close-up images of a flat scene, the outputs differ approximately by a projective transformation,

$$x' = \frac{a_1x + a_2y + a_3}{c_1x + c_2y + c_3}, \qquad y' = \frac{b_1x + b_2y + b_3}{c_1x + c_2y + c_3},$$

therefore configurations of the same group of landmarks, pictured in two such images have approximately the same projective shape. Two or more partial images of a larger 2D scene, from arbitrary view points, may be used to reconstruct the scene. An example of using a DLT based patching method was presented in Patrangenaru and Patrangenaru [23], where seventeen landmarks were used to estimate the reconstructed scene, via projective transformations between images in Figure 3. The output scene in Figure 4, despite some rough points (least squares were not used here), is a reasonable reconstruction. We applied this methodology for the reconstruction of an almost planar ground scene ($2\frac{1}{2}$D scene) by fusing three aerial images given in Figure 5. The reconstruction, including texture is displayed in Figure 6. It clearly shows some partial blurring around the shared portions of the original images, essentially

Figure 2: Fusion of images in figure 1, around an extrinsic mean affine shape.

due to the camera movement. Note that given the pixel size (about 10' in ground units), the blurring effect could not be removed by any smoothing.

3 Projective reconstruction of a 3D configuration from a pair of digital camera images

Image acquisition from the 3D world to the 2D camera film is based on a *central projection principle*, therefore *projective geometry* governs the physics of the ideal pinhole cameras. A point in the outer space and its central projection via the camera pinhole, determine a unique line in space, leading to the notion of projective point. In general, consider a real vector space V, and let 0_V be the zero of this vector space. Two vectors $x, y \in V\backslash\{0_V\}$ represent the same projective point if they differ by a scalar multiple. The equivalence class of $x \in V\backslash\{0_V\}$ is a *projective point* $[x]$, and the set of all such projective points is the *projective space* $P(V)$ associated with V, $P(V) = \{[x], x \in V\backslash\{0_V\}\}$. The real projective space in m dimensions is $\mathbb{R}P^m = P(\mathbb{R}^{m+1})$. Another notation for a *projective point* $p = [x] \in \mathbb{R}P^m$, equivalence class of $x = (x^1, \ldots, x^{m+1}) \in \mathbb{R}^{m+1}$, is $p = [x^1 : x^2 : \cdots : x^{m+1}]$ featuring the *homogeneous coordinates* of p. A *projective transformation* is a bijective function α of $\mathbb{R}P^m$, associated with a matrix $A \in GL(m+1, \mathbb{R})$, defined by $\alpha([u]) = [Au]$.

Consider an *ideal pinhole camera* and assume for simplicity that the focal distance may be taken as unit, $f = 1$. The camera coordinate system is assumed to have the origin O at the camera pinhole and the one axis perpendicular on the film, so that the image of a world point p of coordinate X_1 is $x_1 = (\xi_1, \eta_1, 1)^T$, where $X_1 = \lambda_1 x_1$. If we move the camera to a different location the pinhole will move to a point O', new location of the origin of the new rigid Cartesian coordinate system of the camera.

Figure 3: Partial views of one side of the old ATT building in Atlanta.

The point p has the new coordinates X_2 and if the two camera positions involve a roto-translation $(R,T) \in SO(3) \times \mathbb{R}^3$, where $T = O\bar{O}'$, then $X_1 = T + RX_2$, and the new recorded image of the point p is $x_2 = (\xi_2, \eta_2, 1)^T$, where $X_2 = \lambda_1 x_2$ (see Figure (3). From these basic equation we have $\lambda_1 x_1 = T + \lambda_2 R x_2$ therefore $x_2^T(T \times Rx_1) = 0$. Note that $x \to T \times x$ is linear function and the matrix associated with this function is labeled $T_\times$. Therefore if $E = T_\times R$. It turns out that, given an ideal camera, we have the pairing

$$x_2^T E x_1 = 0. \tag{3.1}$$

If in addition we take into account the internal camera parameter matrix K, in extended pixel coordinates u_1, u_2 ($x_a = Ku_a, a = 1, 2$) we obtain

$$u_2^T F u_1 = 0, F = K^T E K. \tag{3.2}$$

F is the so called *fundamental matrix*. If the camera internal parameters are unknown , or the camera is unavailable, we say that the camera is *noncalibrated*.

The problem of the reconstruction of a configuration of points in 3D from two ideal noncalibrated camera images with unknown camera parameters, is equivalent to the following: given matched configurations of points $[u_i], i = 1, 2$, in two pictures, seek a configuration of points $[u] \in \mathbb{R}P^3$ such that $[u_i] = \tilde{\beta}_i[u], i = 1, 2$, where β_i is the camera projection for the i-th picture. Using the fundamental matrix, Faugeras [10] and Hartley-Chen-Gupta [12] showed that for an ideal noncalibrated camera, the reconstruction problem has a solution that is unique up to a projective transformation in $\mathbb{R}P^3$. The projective ambiguity of the reconstruction was given the following new interpretation in Patrangenaru et. al. [22] : In absence of occlusions, the reconstructed

Figure 4: Estimate of the reconstructed scene from views in Fig. 3.

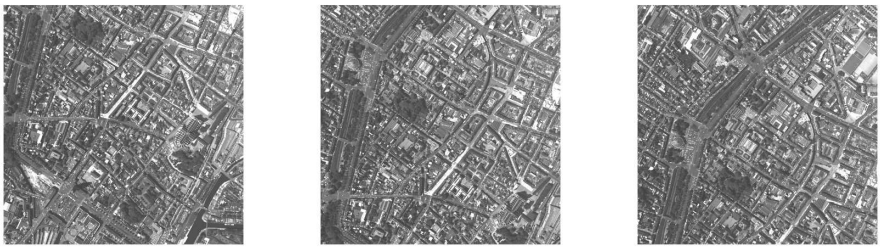

Figure 5: Aerial partially overlapping images of a city.

3D configuration $\mathcal{C}'$ from a pair of matched configurations in uncalibrated cameras views of a 3D configuration $\mathcal{C}$, and $\mathcal{C}'$ have the same projective shape, meaning that given a number of uncalibrated digital camera images of a 3D scene, the best one can do, is to reconstruct the projective shape of that scene. The concept of *similarity shape*, as originally developed by D.G.Kendall cannot be applied to pinhole camera imaging when the camera is not available, since the lack of internal camera parameters and the nature of the scene imaged most often can not allow camera calibration (Ma et. al., [18]).

4 Projective shape reconstruction of a 3D scene

The problem of the reconstruction of a 3D scene assumes an existing matched configuration of points in two camera images. The initial goal is finding points in space $X_1, \ldots, X_k \in \mathbb{R}^3$ given matched pixel configurations of labeled points $\{u_{a,1}, \ldots, u_{a,k}\} \subset \mathbb{R}^2_a, a = 1, 2$, extracted from two noncalibrated cameras $c_a, a = 1, 2$ of certain in-

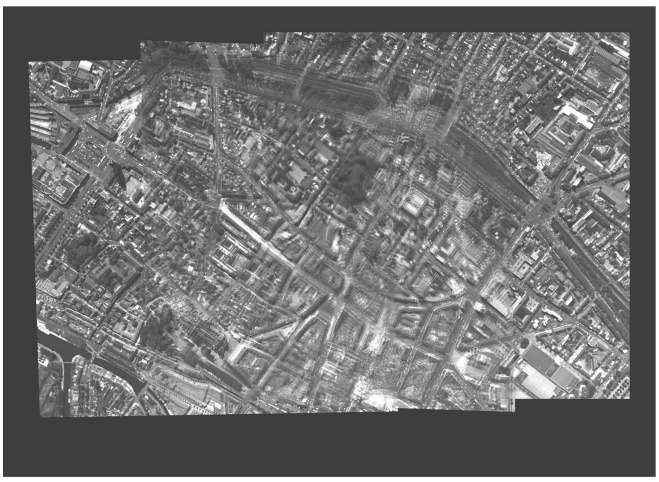

Figure 6: Reconstruction of a $2\frac{1}{2}$D scene from images in Figure 5 based on 2D extrinsic mean affine shapes.

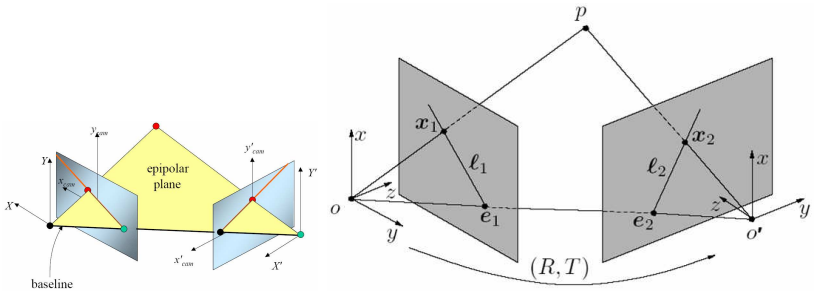

Figure 7: $T = O\bar{O}'$ and R is the camera rotation.

ternal camera parameters, such that there exist two positions of the camera films (planes) π_1, π_2 with the property that pixel coordinates of the c_a-image of X_j are $x_{a,j}, \forall a = 1, 2, j = 1, \ldots, k$.

Since, equation (3.2) is homogeneous as a linear equation in F and F has rank two, in principle, F can be recovered from corresponding configurations of eight points. Due to the inherently discrete nature of digital imaging data and other registration errors, F can be estimated by a matrix $\hat{F}$ using configurations of eight or more recorded matched points $p_{a,i}, a = 1, 2, i = 1, \ldots k, k \geq 8$, whose stacked homogeneous coordinates are the $k \times 3$ matrices $u_a, a = 1, 2$. The resulting linear system for $\hat{F}$ is

$$u_2^T \hat{F} u_1 = 0. \tag{4.1}$$

can be written as

$$\hat{f}^T Y = 0, \tag{4.2}$$

where $\hat{f}$ is a vectorized form of $\hat{F}$. If k is large, the linear homogeneous system is over-determined and the optimal estimated solution $\hat{f}$ can be obtained using simple least squares algorithm by minimizing $\|U^T\hat{f}\|^2$ subject to $\|\hat{f}\| = 1$ (see Hartley and Zisserman [13, p. 593]). Such a reconstruction algorithm that we used here can be found in Ma. et. al. [18, p. 121].

Once a cloud of points in the $3D$ is reconstructed, some of these points are joined in a triangulation that approximates the underlying surface of the scene. We then identify which points vertices the triangulation, enabling us to build a polyhedral model for the scene, where each polygonal face is selected from one image.

This triangulation is used to make a more realistic model of the scene, by adding textures to each triangle from from the corresponding triangles in the original images on the surfaces. This last step can be accomplished "easily" in a Virtual Reality Modeling Language (VRML) file: for each triangle, specify an image using "Texture ImageTexture", and specify the texture coordinates using "texCooord TextureCoordinate", and then specify the correspondence between the points for the vertices and the texture coordinates using texCoordIndex.

5 Examples of a 3D scene reconstruction

5.1 Example 1 of 3D virtual reconstruction of a polyhedral scene

Here we give a synthetic example to illustrate how the spatial scene reconstruction works. In our first example, we consider a 3D polyhedral object manufactured from three cubes that sit on the top of each other, whose sides from top to bottom are four, six and ten units. We selected two from a number of random pictures. The faces are flat therefore the visible portion of its surface is determined by its visible corners, see Figure 8 with visible corners, numbered from 1 to 19. Four images of the object are

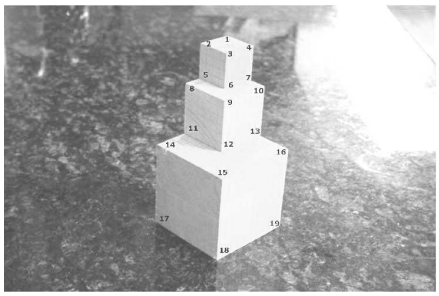

Figure 8: Nineteen landmarks displayed that are used for 3D scene reconstruction, and statistical analysis.

displayed in figure 9. The 2D coordinates of the landmarks selected are and listed in a Table in Appendix 2 of Patrangenaru et. al. [22].

Step 1. The four images in 9 were randomly paired, and for each pair the fundamental matrices were estimated. Using an algorithm from Ma. et. al. [18, p. 121] for each

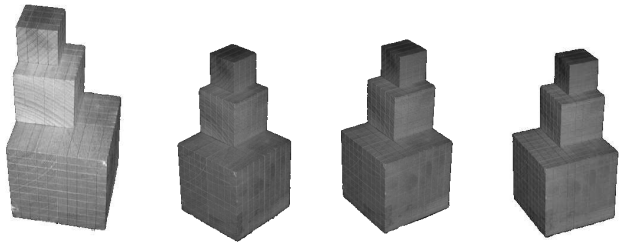

Figure 9: Four digital camera views of a 3D object resembling the blueprint.

pair of corresponding planar configurations, a $3D$ reconstructed configuration of 19 points (*point cloud*) was reconstructed. Homogeneous coordinates of the reconstructed configurations are given in Appendix 2 of Patrangenaru et. al. [22]. Recording of landmark coordinates of camera image pairs, and the first step were done in Matlab. **Step 3.** The points in the $3D$ reconstructed configurations are joint (a *wire-frame* is created) to resemble the original object, as displayed in Figure 10:

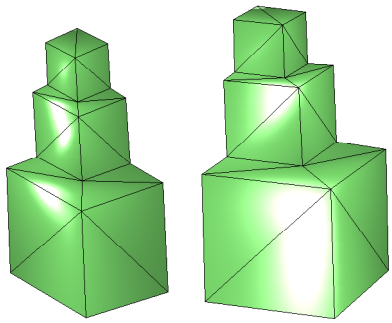

Figure 10: The two reconstructed 3D projective shapes.

Step 3. The triangles are then filled with texture obtained from the two contributing images, based on the computer vision using a VRML (see 11. VRML (Virtual Reality Modeling Language) is a text file format where, e.g. vertices and edges for a 3D polygon can be specified along with UV mapped textures (in our case the texture is borrowed from the contributed images-only two of the outputs are displayed). This step was completed in Blender.

5.2 Architectural style example

The three dimensional projective shape of a scene from two noncalibrated camera views may be obtained by using the 8 point algorithm. Texture matching provides additional information. We used a Delaunay triangulation for texture matching and the graphics software package Blender to apply texture to the reconstructed scene from

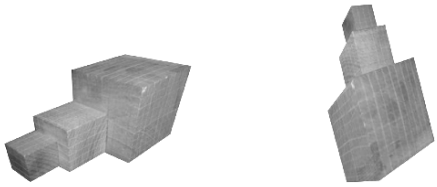

Figure 11: Reconstruction of a polyhedral scene, including texture contents using VRML.

both images. Texture matching provides additional information. We used a Delaunay

Figure 12: Noncalibrated camera views of a building in Florence, Italy.

triangulation for texture matching and the graphics software package Blender to apply texture to the reconstructed scene from both images.

5.3 Eye fundus stereo imaging

Glaucoma is a disease affecting millions of people worldwide. It is a gradual loss of vision, due to an increased Internal Ocular Pressure (IOP) that pushes back the Optic Nerve Head (ONH) Region, thus causing a "tunnel" vision effect that eventually leads to a total loss of vision in one eye or in both eyes. There are two types of devices that measure glaucomatous change: Heidelberg Retina Tomograph (HRT) yields 3D outputs based on the depth measured from the ONH ridge (see Patrangenaru et al. [21], Derado et al. [8], Bhattacharya and Patrangenaru [4]). Stereo camera are a cheaper and less invasive device, which is available to many eye care specialists. Louisiana Experimental Glaucoma Study (LEGS) provided the gold standard data for testing glaucomatous change, since both HRT and stereo data was collected from Rhesus monkeys that were imaged in both eyes. One eye of each monkey was temporarily treated to have the IOP increased, and the other eye was left as a control.

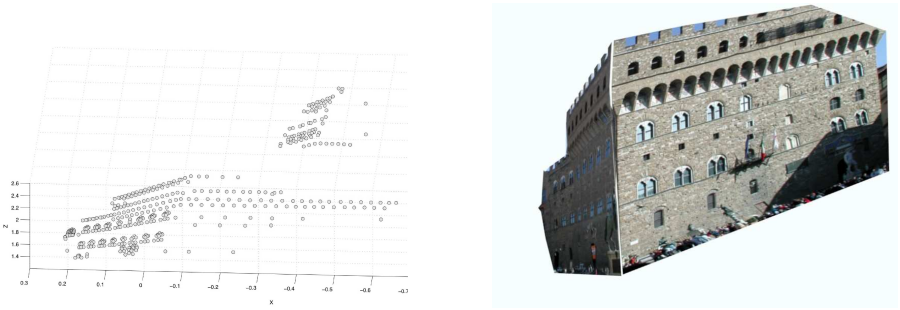

Figure 13: Partial 3D projective shape reconstruction: without and with texture.

The stereo images could be retrieved only in 2008, since the project was halted after Katrina. The coordinates of 50 landmarks on the approximate elliptic contour that determines the ridge of the ONH or are at the junctions of certain blood vessels were recorded across the sample. These include the nine landmarks used for the 3D reconstruction and for detection of glaucomatous projective shape change used in Crane and Patrangenaru [6] displayed in Figure 14, a subset of which are are S(superior), I(inferior), N(nasal), T(templar), V(vertex-the deepest point of the ONH cup) that have been used earlier in statistical analysis for glaucomatous change detection from 3D HRT outputs (Derado et. al. [8]).

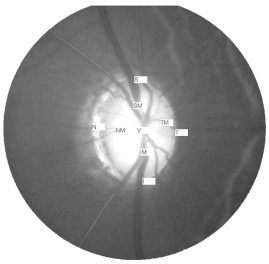

Figure 14: Nine anatomical landmarks on the O N H region.

5.4 Example of colored stereo data of one ONH region

The 3D reconstructed ONH (Optic Nerve Head) region including texture, based on a wireframe from 50 landmarks (including the 9 landmarks used for the 3D reconstruction and for a landmark based statistical detection of glaucomatous projective shape change), is displayed below.

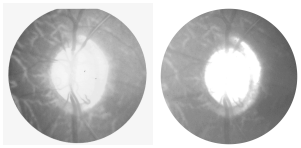

Figure 15: Colored stereo pair of the ONH of a control eye in a Rhesus monkey.

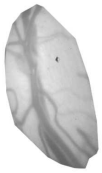

Figure 16: ONH region reconstruction from stereo input in figure 15 including texture contents using VRML.

6 Discussion

It is no accident that most of the living creatures having the ability to see, from insects and snails to mammals and humans, have a *binocular vision.* Stereopsis helps humans mentally recreating a 3D scene from the two retina 2D views of that scene. Nevertheless, this mental reconstruction is far from being similar to the observed scene, since for example, two observed parallel lines, seem two meet at a point in space in their mental reconstructions. The mentally reconstructed scene differs from the observed scene by a projective transformation of the surrounding Euclidean space. While the "built in" biological mechanism of 3D binocular vision is not yet fully understood, an important step towards its understanding is due to work of solving the problem of reconstruction of a 3D configuration of points from a pair of ideal noncalibrated camera images from uncalibrated cameras due to Faugeras [10] and to Hartley et al.[12], who proved that a finite configuration $\mathcal{C}$ of eight or more points in general position in 3D can be reconstructed, using the *fundamental matrix* from the coordinates of the images of these points in two ideal noncalibrated digital camera views, and the reconstruction $\mathcal{R}$ is unique up to a projective transformation in 3D. This projective ambiguity, was holding the key to statistical analysis of 3D configurations. Sughatadasa and Patrangenaru (2006) gave the following simple and practical interpretation the Faugeras-Hartley-Chang-Gupta result: The projective shapes of the 3D configurations of points $\mathcal{R}$ due to F-H-C-G and $\mathcal{C}$ are the identical. In a more suggestive language, this mean that *all we see are 3D projective shapes.* This is the key to a inexpensive data driven future understanding our 3D neighboring reality, that we may now partially reconstruct on its surface, from a number of digital images, as shown in Hartley and Zisserman [13], and many other Computer Vision papers, as well as in this article. This is inviting to a more research on the interface of Statistics and Computer Vision, that should enhance the recent results in Patrangenaru et. al. [22], and Crane and Patrangenaru [6].

Acknowledgements. Research supported by National Science Foundation Grants DMS-1106935, DMS-0805977, DMS-0713012, by INRIA–FSU Associated Team SHAPES and by the NSA Grant MSP-H98230-08-1-0058. The authors are grateful to the National Science Foundation, National Security Agency and INRIA and Centre National de la Recherche Scientifique - France for their support. Thanks to Sam Sughatadasa, for providing the images for Example 1.

References

[1] V. Balan, M. Crane, V. Patrangenaru and X. Liu, *Projective shape manifolds and coplanarity of landmark configurations. A nonparametric approach*, Balkan J. Geom. Appl. 14, 1 (2009), 1-10.

[2] V. Balan, V. Patrangenaru, *Geometry of shape spaces*, BSG Proc. 13, Geometry Balkan Press, Bucharest 2006, 28-33.

[3] R. Berthilsson, A. Heyden, *Recognition of planar objects using the density of affine shape*, Computer Vision and Image Understanding 76 (1999), 135-145.

[4] R. N. Bhattacharya, V. Patrangenaru, *Large sample theory of intrinsic and extrinsic sample means on manifolds-I*, Ann. Statist. 31, 1 (2003), 1-29.

[5] C. F. Burgoyne, H. W. Thompson, D. E. Mercante, R. Amin, *Basic issues in the sensitive and specific detection of optic nerve head surface change within longitudinal LDT TOPSS images*, In: H.G. Lemji, J.S. Schuman (eds), "The Shape of Glaucoma, Quantitative Neural Imaging Techniques", Kugler Publications, The Hague, The Netherlands, 2000, 1-37.

[6] M. Crane and V. Patrangenaru, *Random change on a Lie group and mean glaucomatous projective shape change detection from stereo pair image*, Journal of Multivariate Analysis, 102 (2011), 225-237.

[7] M. Crane, *Nonparametric estimation of three dimensional projective shapes with applications in medical imaging and in pattern recognition*, Ph.D. Dissertation, The Florida State University, 2010.

[8] G. Derado, K. V. Mardia, V. Patrangenaru and H.W. Thompson, *A shape based glaucoma index for tomographic images*, Journal of Applied Statistics 31 (2004), 1241-1248.

[9] I. Dimitric, *A note on equivariant embeddings of Grassmannians*, Publ. Inst. Math. (Beograd) (N.S.) 59 (1996), 131-137.

[10] O. D. Faugeras, *What can be seen in three dimensions with an uncalibrated stereo rig?*, Proc. European Conference on Computer Vision, LNCS 588, 1992, 563-578.

[11] C. Goodall, K. V. Mardia, *Projective shape analysis*, J. Graphical & Computational Statist., 8 (1999), 143-168.

[12] R. I. Hartley, R. Gupta, T. Chang, *Stereo from uncalibrated cameras*, Proc. IEEE Conference on Computer Vision and Pattern Recognition, 1992.

[13] R.I. Hartley, A. Zisserman, *Multiple view Geometry in computer vision*, 2-nd edition, Cambridge University Press, 2004.

[14] A. Heyden, *Geometry and Algebra of Multiple Projective Transformations*, Ph.D. Thesis, University of Lund, Sweden, 1995.

[15] A. Heyden, *Geometry and algebra of multiple projective transformations*, Diss. Summ. Math. 1, 1-2 (1996), 211-218.

[16] F. Lafarge, X. Descombes, J. Zerubia, M. Pierrot-Deseilligny, *3D city modeling based on hidden Markov model*, Proc. IEEE International Conference on Image Processing (ICIP), San Antonio, Etats-Unis, Septembre 2007.

[17] S. J. Maybank, P. A. Beardsley, *Classification based on the cross ratio*, In: "Applications of Invariance in Computer Vision", ed. J.L. Mundy, A. Zisserman and D. Forsyth, Springer, Berlin 1994, 433-472.

[18] Y. Ma, S. Soatto, J. Kosecka, S. S. Sastry, *An invitation to 3-D vision*, Springer, New York, 2006.

[19] K. V. Mardia, V. Patrangenaru, *Directions and projective shapes*, Ann. Statist. 33, 4 (2005), 1666-1699.

[20] K. V. Mardia, V. Patrangenaru, G. Derado, V. P. Patrangenaru, *Reconstruction of planar scenes from multiple views using affine and projective shape*, Proc. of the 2003 Workshop on Statistical Signal Processing 2003, 285-288.

[21] V. Patrangenaru, H. W. Thompson, G. Derado, *Large sample and bootstrap methods on for 3d shape change with applications to detection of glaucomatous change in images of the optic nerve head*, in "Abstracts of the Leeds Annual Statistics Research Workshop in honor of the 65th birthday of Prof. K.V.Mardia", 2000, 30-33.

[22] V. Patrangenaru, X. Liu, S. Sugathadasa, *Nonparametric 3D projective shape estimation from pairs of 2D images - I, In Memory of W.P. Dayawansa*, Journal of Multivariate Analysis, 101 (2010), 11-31.

[23] V. Patrangenaru, V. P. Patrangenaru, *Mean shapes, image fusion and scene reconstruction*, Proc. of The Conference of Applied Differential Geometry - Aristotle University of Thessaloniki, Greece (Ed: Gr.Tsagas), 2004, 230-242.

[24] V. Patrangenaru, K. V. Mardia, *Affine shape analysis and image analysis*, Proc. of the Leeds Annual Statistics Research Workshop 2003, 57-62.

[25] V. Patrangenaru, *New Large sample and bootstrap methods on shape spaces in high level analysis of natural images*, Communications in Statistics Theory and Methods, 30 (2001), 1675-1695.

[26] M. Pierrot-Deseilligny, N. Paparoditis, *A multiresolution and oprimization-based image matching approach: an application to surface reconstruction from SPOT5-HRS stereo imagery*, MATIS Laboratory, 2006.

[27] G. Sparr, *Simultaneous reconstruction of scene structure and camera locations from uncalibrated image sequences*, Proc. of the 13-th Int. Conf. on Pattern Recognition, Vienna, Austria, IEEE Comput. Soc. Press, Washington DC, 1996, 328-333.

Authors' addresses:

Vic Patrangenaru, X. Liu, W. Liu	Florida State University, USA.
M.A. Crane	EPA, Cincinnati, OH, USA.
X. Descombes	INRIA, Sophia Antipolis, France.
G. Derado	Center for Diseases Control, Atlanta, GA, USA.
Vladimir Balan	University Politehnica, Bucharest, Romania.
Vlad P. Patrangenaru	Georgia Tech Alumni Association, USA.
H.W. Thompson	Louisiana State Univ., New Orleans, LA, USA.