Structured light on dynamic scenes using standard stereoscopy algorithms

Frédéric Devernay, Olivier Bantiche, Ève Coste-Manière

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Structured light on dynamic scenes using standard stereoscopy algorithms

Frédéric Devernay — Olivier Bantiche — Ève Coste-Manière

N° 4477

June 2002

THÈME 4

ISSN 0249-6399 ISRN INRIA/RR--4477--FR+ENG

Structured light on dynamic scenes using standard stereoscopy algorithms

Frédéric Devernay , Olivier Bantiche , Ève Coste-Manière

Thème 4 — Simulation et optimisation
de systèmes complexes
Projet Chir

Rapport de recherche n° 4477 — June 2002 — 16 pages

Abstract: Most 3D reconstruction systems based on structured light use a single camera for sensing, and a projector that emits one or several patterns over the observed surface. We use the following analogy between these systems and a classical two-camera stereoscopic system: the projector can be seen as one camera for which the image of the 3D scene is known (it is the emitted pattern), and the other camera sees the same scene from another position. Thus, we can use the large family of algorithms that are available for stereoscopy, with the advantage that, since we can choose the content of one of the two images, we get an easy and unique solution to the correspondence problem.

We use this analogy to calibrate a low-cost structured light system capable of acquiring dynamic scenes, and to recover 3D shape and texture from a single image using correlation-based stereoscopy.

Key-words: shape, structured light, stereoscopy, calibration

Lumière structurée sur scènes dynamiques par stéréoscopie

Résumé : La plupart des systèmes de reconstruction 3D basés sur la lumière structurée utilisent une seul caméra come capteur, et un projecteur qui émet un ou plusieurs motifs sur la surface observée. Nous utilisons l'analogie suivante avec un système stéréoscopique classique à deux caméras : le projecteur peut être vu comme une caméra pour laquelle l'image de la scène 3D est connue (c'est le motif projeté), et l'autre caméra voit la même scène depuis une autre position. Ainsi, on peut utiliser toute la famille d'algorithmes disponibles pour la stéréoscopie dense, avec un avantage important : puisqu'on peut choisir le contenu d'une des deux images, on peut facilement obtenir une solution unique au problème de mise en correspondance.

Nous utilisons cette analogie pour calibrer un système bon marché à lumière structurée, capable d'acquérir des scènes dynamiques, et pour reconstruire le géométrie 3D et la texture à partir d'une seule image, en utilisant la stéréoscopie par corrélation.

Mots-clés : forme, lumière structurée, stéréoscopie, calibrage

1 Introduction

1.1 Motivation

The 3D shape of a given surface can be digitized using either passive or active methods [8, 4]. Passive methods use one or several optical sensors (cameras), and work either by extracting depth information from a single image (as in range from focus or shape from shading) or from two or more images taken from different places, from which features are extracted and matched to get 3D shape by triangulation.

The methods that are based on solving the correspondence problem are also called stereoscopic methods. Though they are probably the most widely used passive 3D sensing methods, they suffer from several drawbacks due to their passive nature: features (which can be lines, points, or image areas) cannot be matched between images if there are no natural features on the 3D scene, as is the case with uniformly-colored objects. For this reason, the use of passive stereoscopy is practically limited to applications working in a featureful environment, such as remote sensing, indoor navigation, or planetary exploration. Range sensing methods that use a single image usually make even more assumptions about the scene and their use is even more restricted than for stereoscopy.

The active range sensing methods avoid the problems encountered with the passive 3D range sensing methods by either getting direct range information from the sensor (as with laser range finders), or by projecting a structured light pattern to solve the correspondence problem. In structured light systems, one camera of the stereoscopic system is replaced by a projector, and known patterns are projected onto the 3D scene. Once these patterns have been extracted from the camera image, triangulation between the projector and the camera gives the 3D geometry of the scene.

Our motivation for this work comes from the following observations: first, most image processing and computer vision methods involved in structured light systems, such as calibration of the projecting device or pattern extraction, are usually specifically designed for each of these systems; second, many structured light systems rely on projecting several patterns onto the 3D scene (with a few notable exceptions), and their use is thus limited to the acquisition of static 3D scenes, whereas most algorithms developed for *passive* 3D shape acquisition also work on dynamic scenes. Instead of using different methods for active (i.e. structured light) and passive (i.e. stereoscopic) systems, we propose to use a matching algorithm originally developed for passive stereoscopy in a structured light system.

1.2 Related work

Among the large collection of range-sensing techniques [8], the methods based on projecting a known light pattern on the 3D scene have been the subject of various studies for the last twenty years. Recent reviews compare the different ways of coding and extracting 3D information from structured light [1, 4].

The light projected onto the scene is usually coded using a spatio-temporal modulation of the illumination, or a projection of color-encoded stripes. The projection of time-varying

binary-coded patterns is still a widely used solution, mainly because of the simple image processing necessary for the extraction of the code from the illuminated images [9, 12]. The main drawback of these methods is that, since several patterns have to be projected on the same scene, they are limited to the 3D acquisition of static scenes. The methods based on the projection of color-encoded stripes or patterns may be used on dynamic scenes; however, they are still limited to the acquisition of uniformly colored scenes [3].

In order to recover correct 3D geometry, a structured light system must be calibrated. This involves calibrating both the camera and the pattern projection system. The geometric calibration of the projector can be done manually, or directly from the camera images [7, 2].

In structured light systems, texture must usually be acquired separately, using a uniformly lit image of the scene. One noticeable exception is the method proposed by Proesmans *et al.*, which is capable of acquiring both the 3D geometry *and* the texture of the scene from a single image of a projected grid [11, 10].

1.3 Proposed approach

We propose a system based on the projection of one pattern image, and a single image acquisition, so that this structured light system also works on a dynamic scene (as in [11]). Additionally, since there is a single projected pattern, the projector can either be a standard slide projector or a video projector. The advantage of using a video projector is that once the 3D reconstruction has been made, the projector can be used to draw information on the objects of the scene, thus serving as an augmented reality system.

The camera can either be a still camera or a video camera, depending on the needs. A color camera is useful if the photometric properties of the scene have to be recovered in addition to its geometric properties.

The algorithms used for 3D reconstruction are based on the following observation: the projected pattern can be thought of as the image seen from a virtual camera placed at the projector's location. Consequently, we can use passive stereoscopic methods, with the advantage that we can choose one image of the stereoscopic pair so that the solution to the correspondence problem is both easy to compute and unique.

Other passive vision methods can also be used in this system: for example, the camera and projector calibration is done by detecting the calibration features in the camera image and then mapping them to the virtual projector image using correspondences, so that we use the same calibration algorithm for the camera and the projector (sec. 2). In order to match the intensities between the projected pattern image and the camera image, we also must do the photometric calibration of the camera and the projector: we compute the transfer function from camera pixel exposure to pixel value, and from pattern pixel value to scene illumination.

Once the system is calibrated, matching between the projected pattern and the camera image is done by a standard correlation-based stereoscopic method, without any modification (sec. 3). Similarly, any area-based stereoscopic method can be used in this system, since there exists a locally linear relationship between the intensity in the original pattern and the intensity in the camera image in the neighborhood of corresponding points.

Actually, the pattern image is multiplied by the albedo of each 3D scene point, and ambient lighting is superimposed before it is reprojected onto the camera. By computing this locally linear photometric transform between the pattern image and the camera image, the scene ambient lighting and the albedo can be recovered from the same image that is used for 3D reconstruction (sec. 4).

2 Calibration

A structured light system, just as any stereoscopic system, must be calibrated. Usually, only geometric calibration is necessary, but due to the fact that the projected pattern's intensity undergoes many non-linear transforms, photometric calibration may also be necessary.

2.1 Photometric calibration

Photometric calibration of the whole system is preferable, since the stereo algorithms usually make the assumption that the relationship between the pixel values in the left and right images for corresponding points are at least locally linear. In the classical two-camera stereo setup, it is usually not necessary, since both images undergo the same transfer function before being acquired (usually a simple gamma transform). In our case, the camera image and the projected image may have very different local dynamic ranges: the local dynamic range in the projected image is almost uniform over the image, whereas the local dynamic range of the camera image depends on the surface albedo.

Though strongly advised, this calibration step is not always necessary (in fact, our first results were obtained without it), but it will improve the results (especially if one wants to compute the real scene photometric properties as described below, sec. 4).

2.1.1 Camera calibration.

To work in a common photometric referential, all the image pixel values can be converted to irradiance values, which are physical values independent of the acquisition device. Let us call E the irradiance value of a point seen in an image, Δt the exposure time (or inverse of shutter speed) of the camera, and a the aperture of the camera (usually expressed as a factor of the focal length). The reciprocity assumption says that halving E and doubling Δt will not change the resulting optical density on the sensor, so the pixel value Z will remain unchanged. The same holds for the aperture: halving E and doubling the square of the aperture a will not change the pixel value Z (the aperture *surface* is proportional to the square of the aperture). Thus, the reciprocity equation leads to:

$$Z = f(Ea^2\Delta t), \tag{1}$$

where f is the sensor response curve. Debevec and Malik [5] showed that this response curve can be computed from several pictures of the same scene taken with different values

of the shutter speed and/or aperture. They compute both the response curve and the high dynamic irradiance of each point by minimizing the following objective function:

$$\mathcal{O} = \sum_{i=1}^{N}\sum_{j=1}^{P}\left[g(Z_{ij}) - \ln E_i - \ln \Delta t_j - 2\ln a_j\right]^2 + \lambda \sum_{z=Z_{\min}+1}^{Z_{\max}-1} g''(z)^2, \tag{2}$$

where $g = \ln f^{-1}$, P is the number of shutter/aperture combinations (i.e. the number of images), and N is the number of image pixels used for response curve calibration. The second term is a smoothing term for the response curve. Since g is in fact a function of discrete values between $Z_{\min}$ and $Z_{\max}$, g'' can be computed by finite differences, and this becomes a linear least-squares problem which can be robustly solved using the singular value decomposition of the corresponding matrix. The MakeHDR[1] software which performs this calibration step is freely available in source form.

2.1.2 Projector calibration.

A given pixel value in the projected image results in an irradiance value on the corresponding camera pixel which can be written as:

$$E = E_0 + Ap(Z), \tag{3}$$

where E_0 is the ambient irradiance (with the projector off), A is a proportionality factor (which we call albedo), and p is the projector response curve which maps pixel values to illumination. E is known from the camera response curve, E_0 can be computed from a set of photographs using the previously described procedure with the projector shut off (or can be considered zero if there is no ambient lighting), Z is known, but A (which depends on the camera pixel) and p are unknown. By taking several images with different projected uniform images, we get the system of equations:

$$E_{ij} = E_{i0} + A_i p(v_j), \tag{4}$$

which can be solved by minimizing the following objective function:

$$\mathcal{O} = \sum_{i=1}^{N}\sum_{j=1}^{P}\left[q(v_j) + \ln A_i - \ln E'_{ij}\right] + \lambda \sum_{v=V_{\min}+1}^{V_{\max}-1} q''(V)^2, \tag{5}$$

where $q = \ln p$, $E'_{ij} = E_{ij} - Ei0$, P is the number of images taken with uniformly projected pixel values v_j, and N is the number of camera pixels taken for calibration. The projected images do not have to cover the whole camera image area: if there are shadow areas, A_i will be zero (or close to zero, because of radiosity) on these areas.

As a result of both photometric calibration processes, we get the camera response curve and the projector response curve. This gives us irradiance values for the camera image pixels,

[1] http://www.debevec.org/FiatLux/mkhdr/

and the illumination for each projected image pixel. Both values are linearly dependent, as shown in eq. 3, the gain and offset between these values being the surface albedo and the ambient irradiance. Thus, both images are locally linearly related if the surface albedo and ambient irradiance are slowly varying, i.e. everywhere except on scene contours and projector shadow boundaries.

We performed our experiments with a black and white video camera, a consumer digital camera, and a consumer LCD video projector. After calibration, the black and white camera showed a linear response curve, the consumer camera had a gamma-like response curve (with $\gamma \approx 2$), and the LCD projector had an inverse gamma-like response curve (with $\gamma \approx 2.6$).

Whenever used within the algorithms, all the images are transformed onto the same photometric reference (up to a linear transform), which means that the transfer function to real-world scene intensity must be applied to both the pattern image and the camera image.

2.2 Geometric calibration

After photometric calibration, there is a local linear relation between intensities in the pattern image and in the camera image. If we use a random projection pattern (as described below, sec. 3.1), any algorithm that uses normalized correlation for matching two images can be applied. For example, the fundamental matrix of the corresponding stereo system can be computed by correlation-based point matching and robust estimation [15].

To get the full Euclidean geometry of an optical stereo setup, the best way is to use the images of a calibration grid as seen by the two cameras. In our case, only the camera can see the image of the calibration grid: the virtual camera represented by the projector only sees the structured light pattern. Still, we can use the projected pattern to remap the calibration grid in the projector virtual image.

In fact, when the 3D scene is planar and there is no nonlinear distortion, there is a homographic relationship between the pattern image and the camera image. If we use a planar black and white calibration grid, this homography can be computed by point matching and robust estimation. We computed this homography by first extracting the points that have the highest response to the Harris operator, matching them by normalized correlation of the image and pattern intensities, and then estimating the homography from this set of matches using a robust method (like the least median of squares [13]). We deal with nonlinear distortions of the camera image by removing this distortion before the process (using the camera calibration parameters), and we suppose that the projector has no or little nonlinear distortion.

We can remap the pattern image onto the camera image of the planar calibration grid, and classify the black and the white areas of the camera image using the following method. From samples of one image of the calibration grid, we compute the mean and standard deviation (μ_w, σ_w) of the intensity in the white parts of the calibration grid. Similarly, we compute (μ_b, σ_b) for the black parts and (μ_t, σ_t) for the whole pattern image. Let $I(x, y)$ be the exposure value of a pixel in the camera images, $T(x, y)$ the value of a pixel in the

remapped texture image,

$$\alpha = \left| \frac{I(x,y) - \mu_w}{\sigma_w} - \frac{T(x,y) - \mu_t}{\sigma_t} \right|, \beta = \left| \frac{I(x,y) - \mu_b}{\sigma_b} - \frac{T(x,y) - \mu_t}{\sigma_t} \right|. \quad (6)$$

Within the white areas of the grid, we have $\alpha \ll \beta$, and within the black areas we have $\alpha \gg \beta$, so the scalar

$$G(x,y) = \frac{2}{\pi} \arctan\left(\frac{\beta}{\alpha}\right) \quad (7)$$

takes values close to 1 in white areas and 0 in black areas (fig. 1).

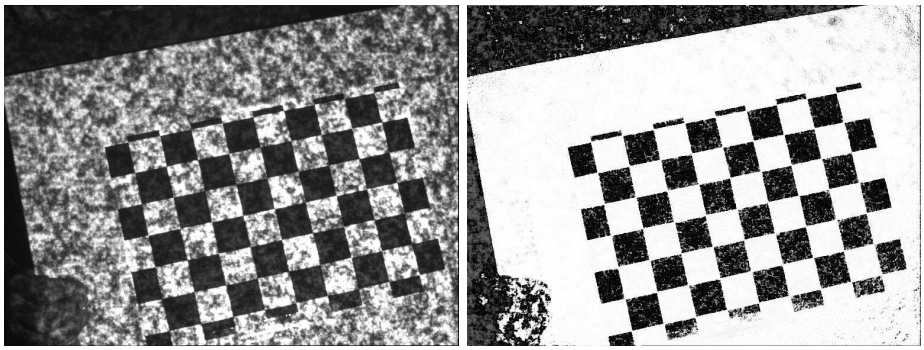

Figure 1: Simultaneous calibration of the camera and the projector: The original camera image of the calibration grid with the projected pattern $I(x,y)$ (left), and the subtracted image $G(x,y)$ where the image has been segmented using the pattern image (right). Objects that are off the plane (such as the hand holding the grid) give random values.

The image $G(x,y)$ can be used with any standard camera calibration method that uses a planar black and white grid, to calibrate both the camera and the projector. The grid features are first detected in the subtracted camera images $G(x,y)$, and can then be remapped to the projector virtual image using the inverse of the homography computed for pattern subtraction. Thus, the same camera calibration method can be used for the camera itself and the projector.

For our experiments, we chose a method that takes as input several views of a single planar grid seen at various orientations [14]. This method is freely available as a camera calibration toolbox for MATLAB[2], or as part of the Intel OpenCV[3] library. One interesting fact is that we found the optical center of the projector at the bottom of the image: since the projectors are usually put lower than the projection screen, this is made to avoid perspective distortion of the projected image.

[2] http://www.vision.caltech.edu/bouguetj/calib_doc/
[3] http://www.intel.com/research/mrl/research/opencv/

3 Stereoscopy and 3D reconstruction

3.1 Creating the projected pattern

Most structured light algorithms work by projecting a regular pattern (usually periodic or pseudo-periodic), which may be modulated to incorporate some code into this pattern. In our case, we want to use standard stereoscopy algorithms to solve the correspondence problem instead of an *ad hoc* method, and we must create the best possible conditions to get an easy and unique solution to the correspondence problem. In fact, the patterns used in other structured light systems almost represent the worst case for standard stereo: stereo algorithms will have difficulty in solving the correspondence problem whenever a local image pattern is repeated more than once along an epipolar line, or when an image edge is tangent to an epipolar line.

To avoid such situations, we chose to use a random grey-level pattern, in order to minimize the regularity of the camera image and the presence of straight edges. This pattern must also be able to deal with different resolutions of the projector and the camera, with low-resolution cameras (for desktop or real-time applications) and with the multi-scale stereo algorithms that use filtered versions of the images at different scales. A reasonable solution is a random white noise pattern, where random binary images are generated at different scales and summed up. We also include a phase shift between the scales to avoid horizontal or vertical artifacts in the pattern image (fig. 2).

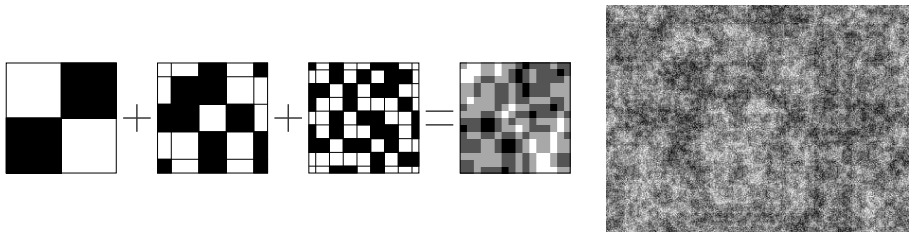

Figure 2: Generating the white noise pattern image using random binary images at different scales with a $\frac{1}{2}$ pixel phase shift between scales (left) and the resulting pattern image (right).

Of course, since the projected pattern is not a binary but a grey-level image, the transfer function from the pattern intensity to scene intensity should be calibrated, as described above, though our experience showed that this has little influence on the quality of stereo results (it is nonetheless necessary for scene unlighting, as described in section 4).

3.2 Matching the pattern and the camera image

Actually, the matching and reconstruction method used in this system is the simplest part to describe, since we can use any standard correlation-based stereo algorithms to solve the correspondence problem. Rectification is done using the camera and projector calibration

data, so that corresponding points lie on the same horizontal scanline in both rectified images. Because the surface albedo does a locally linear transform of the projected pattern illumination, we must use a normalized correlation criterion, such as zero-mean normalized cross correlation (which can be seen as the cosine of the angle between the two intensity vectors):

$$c_{x,y}(d) = \frac{\sum_{i,j}(I^c_{x+i,y+j} - \overline{I^c_{x,y}})(I^p_{x+d+i,y+j} - \overline{I^p_{x,y}})}{\sqrt{\sum_{i,j}(I^c_{x+i,y+j} - \overline{I^c_{x,y}})^2 \sum_{i,j}(I^p_{x+d+i,y+j} - \overline{I^p_{x,y}})^2}}, \tag{8}$$

where I^c is the intensity of the rectified camera image, I^p is the intensity of the rectified pattern image, all sums are made over a fixed-size window, and $\overline{I^c_{x,y}}$ denotes the mean intensity within a window centered at (x, y).

As pointed out before, any other intensity-based stereo algorithm can be applied, e.g. multi-scale approaches, or methods that also extract the normal to the surface from the images [6]. Once the disparity map $d(x, y)$ has been recovered, the 3D reconstruction of the scene by triangulation is easily done using the projector and camera calibration data.

4 Scene texture extraction

We now show that once the disparity map has been recovered by stereo, we can also compute some photometric properties of the scene, by using the fact that the projected pattern contains high spatial frequencies. As far as we know, there is only one other structured light method that extracts both the 3D geometry and the texture of the scene from a single image of a possibly dynamic scene [11, 10].

4.1 3D and photometric properties from a single image

Under the assumption that radiosity effects (i.e. self-lighting of the scene) are negligible, the irradiance of a pixel in the camera image is a linear function of the projected illumination at that point[4]. We call the value A_0 of this linear function at the origin the *ambient irradiance*, i.e. the irradiance of that point if the projector was turned off, and we call its slope A the *albedo* (these are intrinsic property of the surface, since these values depend on the position of the projector and the camera).

If we add the assumption that ambient irradiance and albedo are smoothly varying within the scene, the linear relationship between the projector illumination and the image irradiance still holds in a neighborhood. Hence, we can compute the ambient irradiance and the albedo by linear regression between the values of the illumination and irradiance for each camera pixel. In practice, this means that the projected illumination pattern must first be remapped to the camera image by applying successively the rectification of the pattern image, the disparity map, and the inverse of the rectification of the camera image. Ambient

[4]This still holds if the scene surface is not lambertian.

irradiance and albedo can then be estimated by weighted linear regression (in our examples, we use a Gaussian spatial weighting function $w_{ij} = \exp(-\frac{i^2+j^2}{2\sigma^2})$ with $\sigma = 2$ pixels):

$$(A_0, A) = \arg\min_{A_0,A} \sum_{i=-w}^{w} \sum_{j=-h}^{h} w_{ij} \left(I^c_{x+i,y+j} - (A_0 + AI^p_{x+i,y+j})\right)^2 \quad (9)$$

If the camera image is a color image, these values have to be computed separately for each channel.

4.2 Dealing with scene intensity discontinuities

The problem with the previous approach is that surface albedo discontinuities happen quite often in a real scene, and we would like to recover these discontinuities as well as the correct ambient irradiance and albedo around the discontinuities. The estimation of the photometric properties of one point should only use points in its neighborhood that have the same photometric properties, and other points should be discarded.

This classification of points in the neighborhood is done using a robust estimation method based on the least median of squares [13], which rejects outliers by testing random subsets of the original neighborhood. The method can also be easily adapted so that the centerpoint of the neighborhood is always part of the model and cannot be rejected.

4.3 Re-lighting the scene

Once we have the ambient irradiance and the albedo image, it is easy to re-light the scene with a uniform light that would be at the same position as the projector. The re-lit irradiance image $L(x, y)$ can be computed from the ambient irradiance image $A_0(x, y)$, the albedo image $A(x, y)$, and the light illumination λ using the simple formula:

$$L(x, y) = A_0(x, y) + \lambda A(x, y). \quad (10)$$

This re-lit irradiance image should be transformed by the camera transfer function in order to simulate the acquisition of a uniformly-lit image with that camera. A colored light can also be applied if the camera image is in color: the color is composed by using different values of λ for each channel.

5 Results

We tried the method on various objects and scenes. Scenes can be static or dynamic, since the method only uses one image.

5.1 Acquisition setup

The projector and camera photometric calibration can be done once and for all, since the response curves usually won't change over time (whereas the geometric calibration data do). As described above, The camera photometric calibration is done by taking several images of a normal scene with different values of the aperture and/or shutter speed (the images should go from under-exposed to over-exposed, and images containing high or low saturation are useful too). The projector calibration can be done in a dark room (to avoid ambient irradiance), by using the camera to take images of a scene lit with uniform gray images containing a known value from black to white.

A typical acquisition session starts by locating the scene volume, and placing the projector and the camera so that:

- the projector illuminates the part of the scene that has to be acquired,
- the volume where the projected pattern is in focus covers the volume of the scene,
- the camera field of view and range of focus covers the same parts of the scene.

A reasonable choice for the baseline between the camera and the projector is to get a factor of about 5 to 10 between the baseline and the camera-to-scene distance (a bigger baseline deteriorates the quality of the disparity map, whereas a smaller baseline gives less accuracy on the 3D reconstruction).

Figures 3 and 4 show various results on a human torso, using a 1024×768 LCD projector and a standard 640×480 monochrome video camera. Some details of the real texture (fig. $3d$) can be seen on the re-lit image (fig. $3c$). Figure 5 shows that the method works on static as well as dynamic scenes. The original images were acquired using a consumer digital camera.

5.2 System limitations

This structured light system, as any other 3D acquisition system, still has some limitations. These may come from the use of structured light as a method to solve the correspondence problem, or from the standard stereo algorithms used to recover the 3D shape.

If the scene is highly textured, i.e. has a high dynamic and contains high spatial frequencies, the texture may take over the projected pattern in the camera image, so that the pattern will become indistinguishable from the scene texture. There are similar limitations with other structured light systems (e.g. in a stripe projection system if the scene has stripes that are almost parallel to the structured light stripes).

Some limitations come from the stereoscopy algorithm: parts of the scene that present a high slope with respect to the camera, 3D structures that are smaller than the size of the correlation window, or occlusion contours may not be well rendered by the acquisition system. Since the projected pattern is computed to work at various scales, smaller structures or higher detail could be captured using a zoomable pan and tilt camera, which could focus on areas of interest. High slope areas should be acquired using a different viewing position, and merged with the rest of the scene.

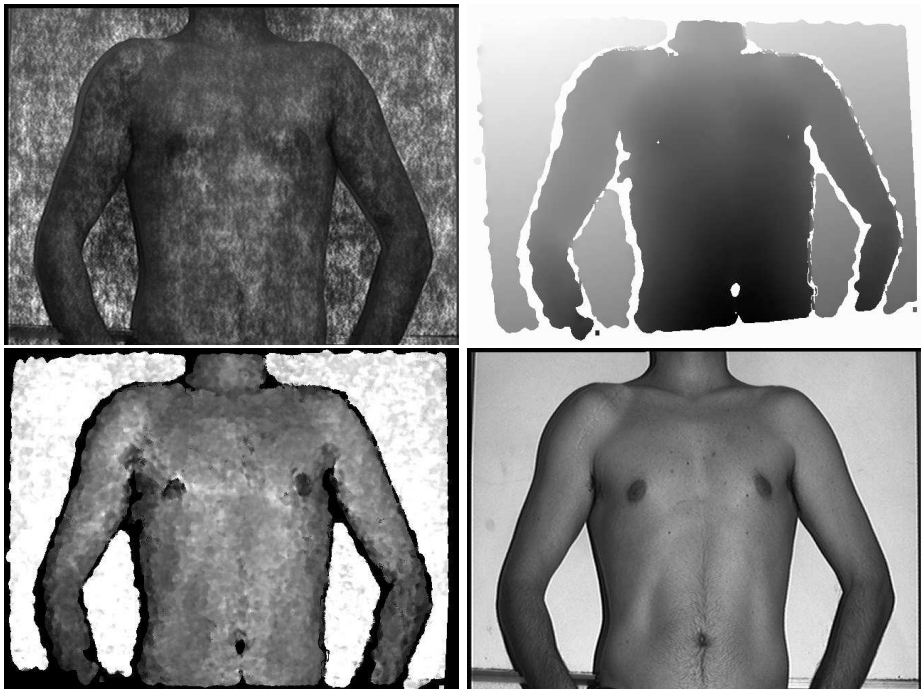

Figure 3: Left to right, top to bottom: *(a)* the original image captured with the projected pattern, *(b)* the corresponding disparity map obtained by correlation with the projected pattern, *(c)* the re-lit image obtained from image *a*, and *(d)* an image of the uniformly lit scene for comparison with *c*.

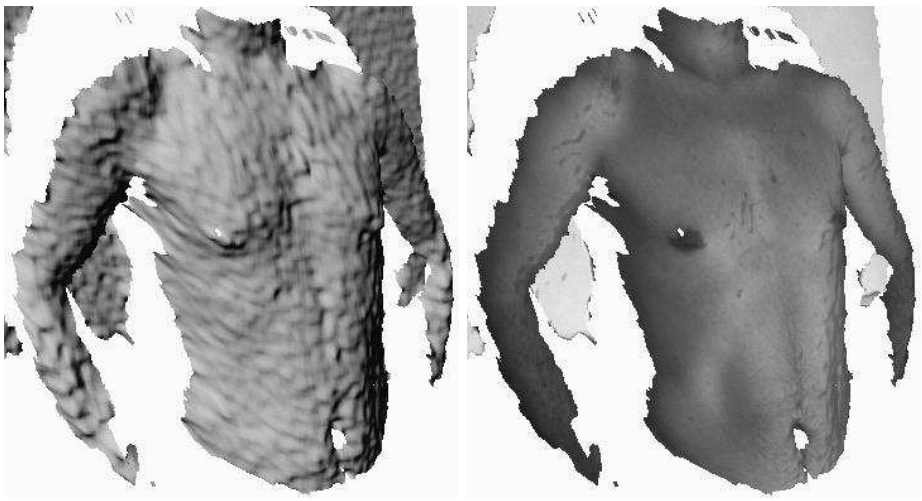

Figure 4: 3D reconstructions from the image of figure 3*a*: shaded view without texture, and with texture from figure 3*d*.

Figure 5: Shaded 3D reconstructions of various scenes: an elephant toy, a man with a moustache, two hands. This shows the variety of static or dynamic 3D scenes that can be acquired using this system.

Un-lighting and re-lighting gives an image that doesn't have the same quality as a real image of the uniformly lit scene, especially on textured scenes or on scenes that have many intensity discontinuities. If high quality scene texture is important, the texture should be acquired using an supplementary image of the uniformly lit scene.

6 Conclusion

We presented a structured light system that uses the analogy between structured light and stereoscopy to acquire the 3D shape and texture of a static or dynamic scene: the projector is similar to a camera that always sees the same image, i.e. the projected pattern, whereas the camera sees the same scene from a different position.

Our method works by projecting a white noise pattern onto the scene, taking a single image using a standard camera, and then using a standard correlation-based stereo algorithm to solve the correspondence problem and compute the 3D reconstruction. Since the incoming projector illumination is affected by the scene albedo, the local dynamic can be very different between the images, so both the geometric and photometric calibration of the camera and the projector were performed. The scene texture can also be extracted from the same image that was used for 3D acquisition, thus allowing acquisition of 3D shape and texture of a dynamic scene by using a single camera.

The system can be adapted to suit the user's requirements in terms of cost and usage: one can use a video camera or a digital still camera, since it works on both static and dynamic scenes; low or high camera resolutions will work, since the method does not need to extract special features such as a projected grid. Besides, the projector can either be a standard slide projector or a video projector, one advantage of the video projector being that it can be used after 3D acquisition to project other information on the scene, in order to perform augmented reality by painting on a free-from 3D surface.

Of course, there is still a lot of room for improvement in the system-specific method, in particular in the un-lighting and re-lighting algorithms. However, one of its biggest advantages is the use of standard vision algorithms for geometric calibration and for the stereoscopy part, so that this system will benefit from any improvements in these fields.

References

[1] J. Battle, E. Mouaddib, and J. Salvi. Recent progress in coded structured light as a technique to solve the correspondence problem: a survey. *Pattern Recognition*, 31(7):963–982, 1998.

[2] Jean-Yves Bouguet and Pietro Perona. 3D photography on your desk. In *Proc. of the Int. Conf. on Computer Vision*, Bombay, India, 1998.

[3] K. Boyer and A. Kak. Color-encoded structured light for rapid active ranging. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9(10):14–28, 1987.

[4] Brian Curless and Steve Seitz. 3D photography. ACM SIGGRAPH '00 Course Notes, July 2000. http://www.cs.cmu.edu/~seitz/course/3DPhoto.html.

[5] Paul E. Debevec and Jitendra Malik. Recovering high dynamic range radiance maps from photographs. *Computer Graphics (Proc. SIGGRAPH'97)*, 31(Annual Conference Series):369–378, 1997.

[6] Frédéric Devernay and Olivier Faugeras. Computing differential properties of 3-D shapes from stereoscopic images without 3-D models. In *Proceedings of the International Conference on Computer Vision and Pattern Recognition*, pages 208–213, Seattle, WA, June 1994. IEEE.

[7] R. B. Fisher, A. P. Ashbrook, C. Robertson, and N. Werghi. A low-cost range finder using a visually located, structured light source. In *Proc. 2nd Int. Conf. on 3-D Digital Imaging and Modeling*, pages 24–33, Ottawa, Canada, October 1999.

[8] R. Jarvis. Range sensing for computer vision. In A.K. Jain and P.J. Flynn, editors, *Three-Dimensional Objects Recognition Systems*. Elsevier Science Publishers, 1993.

[9] J. L. Postdamer and M. D. Altschuler. Surface measurement by space-encoded projected beam systems. *Computer Graphics and Image Processing*, 18(1):1–17, 1982.

[10] Marc Proesmans and Luc Van Gool. Reading between the lines – a method for extracting dynamic 3D with texture. In *Proc. ACM symp. on Virtual reality software and technology*, pages 95–102. ACM Press, New York, 1997.

[11] Marc Proesmans, Luc Van Gool, and A. Oosterlinck. One-shot active 3D shape acquisition. In *Proc. ICPR'96*, pages 336–340. IEEE, 1996.

[12] C. Rocchini, P. Cignoni, C. Montani, P. Pingi, and R. Scopigno. A low cost optical 3D scanner. *Computer Graphics Forum (Eurographics 2001 Conference Proc.)*, 20(3):299–308, September 2001.

[13] P.J. Rousseeuw and A.M. Leroy. *Robust Regression and Outlier Detection*. John Wiley & Sons, New York, 1987.

[14] Z. Zhang. Flexible camera calibration by viewing a plane from unknown orientations. In *Proceedings of the 7th International Conference on Computer Vision*, pages 666–673, Kerkyra, Greece, 1999. IEEE Computer Society, IEEE Computer Society Press.

[15] Z. Zhang, R. Deriche, O. Faugeras, and Q.T. Luong. A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry. *Artificial Intelligence Journal*, 78:87–119, October 1995.

Unité de recherche INRIA Sophia Antipolis
2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex (France)

Unité de recherche INRIA Lorraine : LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex (France)
Unité de recherche INRIA Rennes : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex (France)
Unité de recherche INRIA Rhône-Alpes : 655, avenue de l'Europe - 38330 Montbonnot-St-Martin (France)
Unité de recherche INRIA Rocquencourt : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex (France)

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-6399