

(51) International Patent Classification:
G01B 11/245 (2006.01) *G06T 1/00* (2006.01)(21) International Application Number:
PCT/JP2009/067136(22) International Filing Date:
24 September 2009 (24.09.2009)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
0818664.5 10 October 2008 (10.10.2008) GB(71) Applicant (for all designated States except US):
KABUSHIKI KAISHA TOSHIBA [JP/JP]; 1-1,
Shibaura 1-chome, Minato-ku, Tokyo, 1058001 (JP).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **HERNANDEZ,**
Carlos [—/GB]. **VOGIATZIS, Gero**ge [—/GB].
CIPOLLA, Roberto [—/GB].(74) Agents: **SUZUYE, Takehiko** et al.; c/o **SUZUYE &**
SUZUYE, 1-12-9, Toranomon, Minato-ku, Tokyo,
1050001 (JP).(81) Designated States (unless otherwise indicated, for every
kind of national protection available): AE, AG, AL, AM,
AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ,
CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO,
DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT,
HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP,
KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD,
ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI,
NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, SC, SD,
SE, SG, SK, SL, SM, ST, SV, SY, TJ, TM, TN, TR, TT,
TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.(84) Designated States (unless otherwise indicated, for every
kind of regional protection available): ARIPO (BW, GH,
GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM,
ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ,
TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE,
ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV,
MC, MK, MT, NL, NO, PL, PT, RO, SE, SI, SK, SM,
TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW,
ML, MR, NE, SN, TD, TG).

Published:

— with international search report (Art. 21(3))

(54) Title: IMAGING SYSTEM AND METHOD

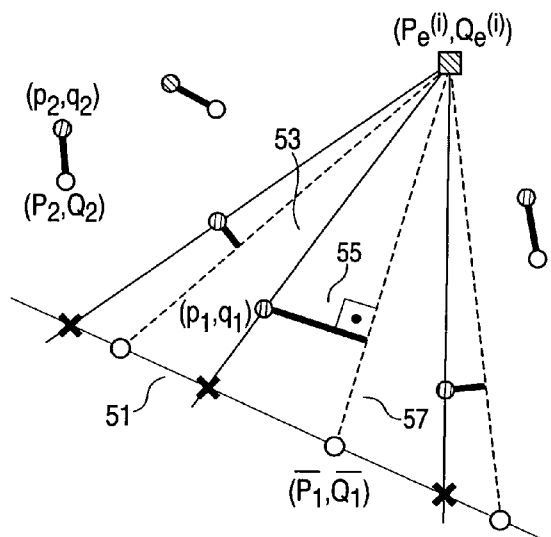


FIG. 4

(57) Abstract: An apparatus for determining
the gradients of the surface normals of an ob-
ject, the method comprising means to receive
data of three 2D images of the object, wherein
each image is taken under illumination from a
different direction, means to establish which
pixels of the image are in shadow such that
there is only data available from two images
from these pixels, means to determine a range
of possible solutions for the gradient of the sur-
face normal of a shadowed pixel using the data
available for the two images, and means to se-
lect a solution for the gradient using the inte-
grability of the gradient field over an area of
the object as a constraint and minimising a cost
function.

D E S C R I P T I O N

IMAGING SYSTEM AND METHOD

5 Technical Field

The present invention is concerned with the field of imaging systems which may be used to collect and display data for production of 3D images. The present invention may also be used to generate data for 2D and
10 3D animation of complex objects.

Background Art

The field of 3D image production has largely been hampered by the time which it takes to take the data to produce a 3D film. Previously, 3D films have generally
15 been perceived as a novelty as opposed to a serious recording format. Now, 3D image generation is seen as being an important tool in the production of CG images.

Photometric stereo is a well established 3D reconstruction technique. A sequence of images
20 (typically three or more) of a 3D scene are obtained from the same viewpoint and under illumination from different directions. From the intensity variation in each pixel it is possible to estimate the local orientation of the surface of the object that projects
25 onto that pixel.

By integrating all these surface orientations a very detailed estimate of the surface geometry can be

obtained. As any other reconstruction method,
photometric stereo faces several difficulties when
faced with real images. One of the most important of
these difficulties is the frequent presence of shadows
5 in an image. No matter how careful the arrangement of
the light sources, shadows are an almost unavoidable
phenomenon, especially in objects with complex
geometries.

Previously, shadows have been dealt with using
10 four or more images from four different illumination
directions. This over-determines the local surface
orientation and albedo which only requires three
degrees of freedom. This implies that it is possible
to use the residual of a least squares solution, to
15 determine whether shadowing has occurred. However when
there are only three 2D images there are no spare
constraints. Therefore the problem of detecting
shadows becomes more difficult. Furthermore, when a
pixel is in shadow in one of the three images most
20 methods simply discard it.

Disclosure of Invention

The present invention seeks to address the above
problems and in a first aspect provides a method for
determining the gradients of the surface normals of an
25 object, the method comprising:

receiving data of three 2D images of the object,
wherein each image is taken under illumination from a

different direction;

establishing which pixels of the image are in shadow such that there is only data available from two images from these pixels;

5 determining a range of possible solutions for the gradient of the surface normal of a shadowed pixel using the data available for the two images; and

selecting a solution for the gradient at a pixel using the integrability of the gradient field over an area of the object as a constraint and minimising a cost function.

10

The above estimates the surface geometry of a shadowed part of an object by enforcing the integrability of the gradient field.

15 In a preferred embodiment, the method of selecting a solution for the gradient comprises using the constraint that the surface of the object is a continuous piecewise polynomial.

The cost function may be minimised as a least squares problem. The cost function may represent the difference between the real value for each gradient and the potential solutions for the gradient over the area. The cost function may also be configured to take into account noise.

20

25 The area of the object over which the solution is calculated will generally comprise both shadowed and non-shadowed pixels. In a further preferred

embodiment, it will be assumed that there will be noise in the data for both shadowed and non-shadowed pixels.

Noise can cause unwanted artefacts in a generated three dimensional image. Therefore, in preferred
5 embodiment, a regularisation constraint is applied to compensate for noise in the acquired data.

In an embodiment, to compensate for noise the cost function is expressed in terms of a regularisation parameter which is used to compensate for noise. A
10 cost is then imposed on the regularisation parameter such that a pixel is forced to have a value which is close to the midpoint of its neighbours. The cost may be imposed in terms of the first order differential or higher order differential of the regularisation
15 parameter. The regularisation parameter may be related to the length along the straight line of possible solutions from a fixed point on the line.

The problem may be formulated by expressing the surface normals in terms of their gradients in a 2D
20 gradient space such that the range of possible solutions lie along a straight line in the 2D gradient space. The constraint for the least squares analysis may then be expressed in terms of a regularisation parameter, the variation in the regularisation
25 parameter between pixels to be less than a predetermined value and wherein the regularisation parameter is the length along the straight line of

possible solutions from a fixed point on the line.

For example, the range of possible solutions may be determined by expressing the normal as the albedo scaled surface normal $\mathbf{b}$:

5

$$\mathbf{b} = L^{-1}\mathbf{c}$$

wherein $\mathbf{c}$ is $[c_1 \ c_2 \ c_3]$ and c_1 , c_2 and c_3 represent the respective intensity of the pixel measured in each of the three 2D images and $L = [I_1 \ I_2 \ I_3]^T$ where I_1 , I_2 and I_3 are the vectors which represent the direction of the light source which illuminates the object for each of the three 2D images; wherein when the pixel in the i^{th} 2D image is in shadow

15

$\mathbf{b}$ is expressed as:

$$\mathbf{b} = L^{-1}D_i\mathbf{c} + \mu L^{-1}\mathbf{e}_i$$

20

where $\mathbf{e}_i$ is the i^{th} column of the 3×3 identity matrix, D_i is the identity matrix with a zero in the i^{th} position in the diagonal and μ is a scalar parameter representing the value that c_i would have if it were not in shadow;

25

projecting $\mathbf{b}$ into a 2D gradient space using the operator $P[x_1 \ x_2 \ x_3] = (x_1/x_3, x_2/x_3) = (p, q)$ such that for a point in shadow, the point (p, q) lies along the shadow line of the pixel defined as the line that joins $P[L^{-1}D_i\mathbf{c}]$ and $P[L^{-1}\mathbf{e}_i]$.

The above method may further comprise minimising the cost function is performed using a parameter w ,

where w is distance along the shadow line of point (p, q) from a fixed point on the shadow line, the method further comprising imposing a cost on ∇w .

5 The step of determining if a pixel is in shadow in i^{th} of the three images may be achieved by determining if the value of the intensity in the i^{th} image divided by the magnitude of the intensity vector c for all three images is below a predetermined value.

10 The above method may further comprise establishing whether a pixel is in shadow is by introducing a further constraint to reduce the chance of a pixel being determined to be in shadow and a neighbouring pixel not in shadow.

15 The above method can be used when an object is imaged from just three directions. Using photometric stereo on just three images may seem like an unreasonably hard restriction. There is however a particular situation when only three images are available. This technique is known as colour
20 photometric stereo and it uses three light sources with different light spectra. When the scene is photographed with a colour camera, the three colour channels capture three different photometric stereo images.

25 Since shape acquisition is performed on each frame independently, the method can be used on video sequences without having to change illumination between

frames. In this it is possible to capture the 3D shape of deforming objects such as cloth, or human faces.

The method may further comprise acquiring the image data from the object to be imaged or it may
5 operate on pre-recorded or remotely recorded data.

In a second aspect, the present invention provides an apparatus for determining the gradients of the surface normals of an object, the method comprising:

means to receive data of three 2D images of the
10 object, wherein each image is taken under illumination from a different direction;

means to establish which pixels of the image are in shadow such that there is only data available from two images from these pixels;

15 means to determine a range of possible solutions for the gradient of the surface normal of a shadowed pixel using the data available for the two images; and

means to select a solution for the gradient using the integrability of the gradient field over an area of
20 the object as a constraint and minimising a cost function.

The means may be a processor adapted to perform the above functions. Although the means receive data from 3 images, it may be configured to receive data
25 from 4 or more images. However, the above method and apparatus is primarily intended for determining the gradient of the surface normal of a pixel when it is

shadowed such that data is only available for that pixel from 2 images.

The apparatus may further comprise means to acquire the data of the three 2D images.

5 Such apparatus may be a colour photometric stereo apparatus, which further comprises:

 at least three light sources, irradiating the object from three different angles;

 a camera provided to collect radiation from the
10 three radiation sources which has been reflected from the object; and

 an image processor configured to generate a depth map of the three dimensional object,

 wherein each radiation source emits radiation at a
15 different frequency and the image processor is configured to distinguish between the reflected signal from the three different radiation sources.

 However, the present invention may also be applied to standard photometric stereo techniques where
20 multiple images are taken using light of the same frequency and the images from different illumination directions are captured in a sequential manner.

 Although the apparatus can stand alone, it may be incorporated in part of a 3D image generation apparatus
25 further comprising means for displaying a three dimensional moving image from the depth map.

 As previously explained, the object can be moving

and the camera may be a video camera.

The system may also be used in 2D or 3D animation where the system comprises means for moving a depth map generated from the determined gradients.

5 The system may also further comprise means for applying pattern to the depth map, the means configured to form a 3D template of the object from a frame of the depth map and determine the position of the pattern on the object of the frame and to deform the template with
10 the pattern to match subsequent frames. The template may be deformed using a constraint that the deformations of the template must be compatible with the frame to frame optical flow of the original captured data. Preferably the template is deformed
15 using the further constraint that the deformations be as rigid as the data will allow.

In summary, the present invention can be used for 3D reconstruction, medical imaging or cloth modelling.

20 The present invention will now be described with reference to the following preferred non-limiting embodiments in which:

Brief Description of Drawings

FIG. 1 is a schematic of an apparatus configured to perform photometric stereo.

25 FIG. 2 is a schematic of a system in accordance with an embodiment of the present invention configured to perform photometric stereo.

FIG. 3 is a flow diagram illustrating a method in accordance with an embodiment of the present invention.

FIG. 4 is a plot showing a 2D gradient space which is used to describe a method in accordance with an
5 embodiment of the present invention.

FIG. 5 shows a method of deriving gradients applied in practice on synthetic data. FIG. 5a is a schematic of a sphere having shaded regions which is to be imaged, FIG. 5b shows the sphere of FIG. 5a when
10 imaged using a technique in accordance with an embodiment of the present invention; and FIG. 5c shows the sphere of FIG. 5a which imaged using a technique in accordance with a further embodiment of the present invention using a regularisation constraint.

FIG. 6 is a 2D image of a face and a shadow mask of the face produced in accordance with an embodiment of the present invention. FIG. 6a is the 2D image of a face, and FIG. 6b is the shadow mask of the face of
15 FIG. 6a.

FIG. 7 is a schematic of a calibration board used in a method in accordance with an embodiment of the
20 present invention.

FIG. 8 shows an image of a real object which in this case is a frog. FIG. 8a shows three 2D images of
25 a frog obtained from different illumination directions, FIG. 8b shows the three images of FIG. 8a normalised in terms of their intensity; FIG. 8c shows the result of

applying a technique to determine the position of shadows in accordance with an embodiment of the present invention; FIG. 8d shows a shadow mask determined from the data of FIG. 8c and FIG. 8e shows a 3D

5 reconstruction of the frog of FIG. 8a performed in accordance with an embodiment of the present invention.

FIG. 9 shows video data of a white painted face illuminated by three coloured lights. FIG. 9a shows three 3D images generated ignoring shadows and FIG. 9b
10 shows the corresponding results for images generated in accordance with an embodiment of the present invention.

Best Mode for Carrying Out the Invention

FIG. 1 is a schematic of a system in accordance with an embodiment of the present invention used to
15 image object 1. The object is illuminated by three different light sources 3, 5 and 7.

In this embodiment, the system is either provided indoors or outside in the dark to minimise background radiation affecting the data. The three lights 3, 5
20 and 7 are arranged laterally around the object 1 and are vertically positioned at levels between floor level to the height of the object 1. The lights are directed towards the object 1.

The angular separation between the three light
25 sources 3, 5 and 7 is approximately 30 degrees in the plane of rotation about the object 1. Greater angular separation can make orientation dependent colour

changes more apparent. However, if the light sources are too far apart, concave shapes in the object 1 are more difficult to distinguish since shadows cast by such shapes will extend over larger portions of the object making data analysis more difficult. In a preferred arrangement each part of the object 1 is illuminated by all three light sources 3, 5 and 7.

Camera 9 which is positioned vertically below second light source 5 is used to record the object as it moves while being illuminated by the three lights 3, 5 and 7.

FIG. 2 shows an overview of the whole system. An object which may be moving is illuminated by the three light source 3, 5, and 7 described with reference to FIG. 1 and the image is captured by camera 9.

The output of camera 9 is then fed into processor 11 which processes the received data according to steps S103, S105, S107 and S109 of FIG. 3 to produce an image on display 13 of FIG. 2.

FIG. 3 is a flow chart showing a method in accordance with an embodiment of the present invention. In step S101, data using a photometric stereo technique is acquired. The data may be acquired real time and processed immediately or may be acquired by the system from previously recorded data.

As explained with reference to FIG. 1, a minimum of 3 lights is required to perform photometric stereo

with no extra assumptions. If only 2 images are available, then it was previously believed that a constant albedo must be assumed in order to obtain meaningful data. In the present embodiment it is
 5 assumed that there is a smoothly varying albedo. Whenever more lights are available, the light visibility problem becomes a labelling problem where each point on the surface has to be assigned to the correct set of lights in order to successfully
 10 reconstruct the surface.

It is assumed that the camera used to capture the data is orthographic (with infinite focal length) for simplicity, even though the extension to the more realistic projective case is straightforward.

15 In the case of orthographic projection, it is possible to align the world coordinate system so that the xy plane coincides with the image plane while the z axis corresponds to the viewing direction. The surface in front of the camera can then be parameterized as a
 20 height function $Z(x, y)$. If Z_x and Z_y are the two partial derivatives of Z , the normal can be defined as:

$$\mathbf{n} = \frac{1}{\sqrt{Z_x^2 + Z_y^2 + 1}} (Z_x Z_y - 1)^T$$

that is locally normal to the surface at (x, y) .

25 For $i = 1 \dots 3$ let $c_i(x, y)$ denote the pixel intensity of pixel (x, y) in the i^{th} image. It is assumed that in the i^{th} image the surface point $(x \ y \ Z(x, y))^T$ is

illuminated by a distant light source whose direction is denoted by the vector $\mathbf{I}_i$ and whose spectral distribution is $E_i(\lambda)$. It is also assumed that the surface point absorbs incoming light of various wavelengths according to the reflectance function $R(x, y, \lambda)$. Finally, it is assumed that the response of the camera sensor at each wavelength be given by $S(\lambda)$. Then the pixel intensity $c_i(x, y)$ is given by

$$c_i(x, y) = (\mathbf{I}_i^T \mathbf{n}) \int E(\lambda) R(x, y, \lambda) S(\lambda) d\lambda \quad (1)$$

The value of this integral is known as the surface albedo of point $(x, y, Z(x, y))^T$. This allows the albedo scaled normal vector to be defined as:

$$\mathbf{b} = \mathbf{n} \int E(\lambda) R(x, y, \lambda) S(\lambda) d\lambda$$

so that (1) becomes a simple dot product $c_i = \mathbf{I}_i^T \mathbf{b}$ (2)

$$c_i = \mathbf{I}_i^T \mathbf{b} \quad (2)$$

Photometric stereo methods use the linear constraints of (2) to solve for $\mathbf{b}$ in a least squares sense. From this partial derivatives of the height function are integrated to produce the function itself. In three-source photometric stereo, when the point is not in shadow with respect to all three lights, the three positive intensities c_i are measured and each of these gives a constraint on $\mathbf{b}$. If $\mathbf{L} = [\mathbf{I}_1 \ \mathbf{I}_2 \ \mathbf{I}_3]^T$ and $\mathbf{c} = [c_1 \ c_2 \ c_3]^T$ then the system has exactly one solution which is given by

$$\mathbf{L}^{-1} \mathbf{c} \quad (3)$$

If the point however is in shadow, say in the i^{th} image, then the measurement of c_i cannot be used as a constraint. Since each equation (2) describes a 3D plane, the intersection of the two remaining
 5 constraints is a 3D line.

If e_i is the i^{th} column of the 3×3 identity matrix and D_i is the identity matrix with a zero in the i^{th} position in the diagonal, the possible solutions
 10 for b are:

$$\mathbf{b} = L^{-1}D_i\mathbf{c} + \mu L^{-1}\mathbf{e}_i \quad (4)$$

where μ is a scalar parameter. This parameter represents the value that c_i would have, had the point not been in shadow in the i^{th} image.

15 Projecting b into a 2D gradient space using the operator $P[x_1 \ x_2 \ x_3] = (x_1/x_3, x_2/x_3) = (p, q)$ such that for a point in shadow, the point (p, q) lies along the shadow line of the pixel defined as the line that joins $P[L^{-1}D_i\mathbf{c}]$ and $P[L^{-1}\mathbf{e}_i]$.

20 To provide a solution, the scaled normals b_j into 2D space with coordinates p and q . The operator $P[x_1 \ x_2 \ x_3] = (x_1/x_3, x_2/x_3)$ is used to achieve this projection and the scaled normal b of surface point $(x, y, Z(x, y))$ projects to point $P[b] = (p, q)$. Thus, the
 25 coordinates p and q are equal to the two partial derivatives Z_x and Z_y respectively.

According to the image constraints and assuming no noise in the data, one of the following three cases can

occur:

Case 1. The surface point is in shadow in two or more images. In this case there is no constraint in $P[b]$ from the images.

5 Case 2. The surface point is not in shadow in any of the three images. In this case (p, q) coincides with $P[L^{-1}c] = (P, Q)$.

Case 3. The surface point is in shadow in exactly one image, say the i^{th} . In this case (p, q) must lie on the line that joins $P[L^{-1}D_i c] = (\bar{P}, \bar{Q})$ and $P[L^{-1}e_i] = (P_e^{(i)} Q_e^{(i)})$. This line shall be referred to as the shadow line 57 of the shaded pixel.

The above three cases can be thought of as shadow labels for the pixels. Which shadow label should be applied to each pixel is determined in step S103. How the shadow labels are assigned will be described later.

The description will now be concerned with case 3. For all pixels j which are occluded under the i^{th} image, the corresponding points in the 2D gradient space are $(\bar{P}_j, \bar{Q}_j)$ are on the line 51 described the.

20 equation $\mathbf{I}_i^T \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = 0$. Also, the shadow lines of all of the

pixels intersect at the point $(P_e^{(i)} Q_e^{(i)})$.

In the presence of noise in the data c , cases 2 and 3 above do not hold exactly as points (P, Q) and $(\bar{P}, \bar{Q})$ are corrupted: The point (p, q) is slightly different from (P, Q) for unoccluded pixels, and $(\bar{P}, \bar{Q})$

is not exactly on the line joining (p, q) and $(P_e^{(i)}Q_e^{(i)})$.

FIG. 4 shows the configuration for six pixels where pixel 1 is under shadow in the i^{th} image while pixel 2 is not in shadow in any image. The points (p_j, q_j) (dark dots) represent the partial derivatives of the height function at pixel j . For each point (p_j, q_j) there is a corresponding data point (white dot). Pixel 2 is unoccluded and hence (p_2, q_2) must be as close as possible to its data point (P_2, Q_2) . Pixel 1 however is occluded so (p_1, q_1) must be as close as possible to its shadow line 57. This is the line 57 joining its data point $(\bar{P}_j, \bar{Q}_j)$ and the intersection point $(P_e^{(i)}Q_e^{(i)})$.

Once a range of potential solutions has been determined in line with S105, the method progresses to step S107 where a solution is selected using a constraint. The constraint is in the form of an integrability condition.

The (p, q) coordinates are not independent for each pixel as they represent the partial derivatives of a scalar field and as a result they must satisfy an integrability condition. By assuming that the height function is a continuous piecewise polynomial, the integrability condition takes the form of a convolution between p and q . If for example $Z(x, y)$ is a linear interpolation between height values at pixel centres $Z(i, j)$ integrability can be expressed as:

$$p(i+1, j) - p(i, j) = q(i, j+1) - q(i, j)$$

To obtain the solution, values (p, q) for each pixel that minimize the discrepancy between the noisy data and the model while satisfying the integrability condition are obtained.

As mentioned above, under noise in the image data c , the 2D points (P, Q) and (p, q) are not perfectly consistent with the model. For non-shadowed pixels, the difference between the model and data can be measured by the point-to-point square difference term

$$\varepsilon = (p - P)^2 + (q - Q)^2$$

In the case of the shadowed pixels however there are a number of possible ways to quantify the non-collinearity of (p, q) , $(\bar{P}, \bar{Q})$ and $(P_e^{(i)} Q_e^{(i)})$. The point $(\bar{P}, \bar{Q})$ contains noise and hence the distance which should be measured is from $(\bar{P}, \bar{Q})$ to the intersection of the line 53 joining (p, q) , and $(P_e^{(i)} Q_e^{(i)})$ with the line

$$51 \quad \mathbf{I}_i^T \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = 0 \quad \text{which leads to the term:}$$

20

$$\|(\bar{P}, \bar{Q}) - \mathbf{P}[(I - \mathbf{m}_i \mathbf{I}_i^T) \mathbf{b}]\|^2$$

where $\mathbf{m}_i$ is the i^{th} vector of L^{-1} . The above expression is non linear with respect to p and q and thus not suitable as a constraint for a least squares estimate.

25

Therefore, an alternative distance 55 expression is minimised which is the distance from (p, q) to the line 57 joining $(\bar{P}, \bar{Q})$ and $(P_e^{(i)} Q_e^{(i)})$. The term this

corresponds to is:

$$\varepsilon_0^{(i)} = \frac{\left((\bar{Q} - Q_e^{(i)}) (p - P_e^{(i)}) - (\bar{P} - P_e^{(i)}) (q - Q_e^{(i)}) \right)^2}{(\bar{P} - P_e^{(i)})^2 + (\bar{Q} - Q_e^{(i)})^2}$$

where the quantity being squared is linear with
 5 respect to p and q.

Thus, once all pixels have been labelled with
 their shadow conditions in accordance with step S103,
 the cost function which is to be minimised in order to
 solve the whole problems is:

$$\sum_{j \in S} \varepsilon_j + \sum_{j \in S_1} \bar{\varepsilon}_j^{(1)} + \sum_{j \in S_2} \bar{\varepsilon}_j^{(2)} + \sum_{j \in S_3} \bar{\varepsilon}_j^{(3)}$$

where S contains all of the non-shadowed pixels
 while S_i contains pixels shaded in the i^{th} image.

The above cost function is quadratic terms in p
 15 and q, thus finding the minimum of this quantity
 subject to the integrability condition is a constrained
 linear least squares problem that can be solved by a
 variety of methods.

In step S109, a 3D image is generated from the
 20 determined gradients. Knowledge of the gradients
 allows a depth map of the object to be determined. The
 depth map can be used to directly generate a 3D image
 or it can be used to generate data for animation.

For example, if the above method is used to
 25 capture data of a complex moving object such as cloth,
 hair or the like to produce data for a moving mesh.
 The moving mesh of can then be attached to an

articulated skeleton.

Skinning algorithms are well known in the art of computer animation. For example, a smooth skinning algorithm may be used in which each vertex v_k is
 5 attached to one of more skeleton joints and a link to each joint j is weighted by $w_{j,k}$. The weights control how much the movement of each joint affects the transformation of a vertex:

$$10 \quad v'_k = \sum_i w_{i,k} S_i^{t-1} v_k^{t-1}, \quad \sum_i w_{i,k} = 1$$

The matrix S_i^t represents the transformation from the joint's local space to world space at time instant t .

The mesh was attached to the skeleton by first
 15 aligning a depth pattern of a fixed dress or other piece of clothing or the like with a fixed skeleton and for each mesh vertex a set of nearest neighbours on the skeleton. The weights are set inversely proportional to these distances. The skeleton is then animated
 20 using publicly available mocap data (Carnegie-mellon mocap database <http://mocap.cs.cmu.edu>). The mesh is animated by playing back a previously captured cloth sequence of cloth movement.

FIG. 5 shows the above method of deriving
 25 gradients applied in practice on synthetic data. FIG. 5a shows a synthetic sphere, the sphere is illuminated from three directions and the light regions indicate the regions where pixels are shadowed in one

of the three images.

FIG. 5b shows a 3D image of the sphere generated using the above method. The general shape of the shaded areas is well produced providing that the areas are surrounded by unshaded regions. The unshaded regions act as boundary conditions for the shadowed regions and give the problem a unique solution. The presence of scratch artefacts are seen in the shadowed regions in FIG. 5b. These are caused by the point-to-line distances which do not introduce enough constraints in the cost function. The point (p, q) can move significantly in a direction parallel to the corresponding shadow line only to gain a slight decrease in the overall cost. This results in violent perturbations in the resulting height function that manifest themselves as deep scratches running perpendicular to the shadow lines.

FIG. 5c shows an image generated from the data of FIG. 5a in accordance with an embodiment of the present invention where a regularisation scheme is used.

The regularizing criterion is consistent with the linear least squares framework described above such that no non-linear constraints can be enforced and it suppresses noise while preserving as much of the data as possible.

One possible choice for a regularization criterion is minimizing the Laplacian of the height field $\nabla^2 z$.

This is known to have good noise reduction properties and to produce smooth well behaved surfaces with low curvature. However, the Laplacian is isotropic so it tends to indiscriminately smooth along all possible
 5 directions.

There are other anisotropic alternatives to the Laplacian such as anisotropic filtering which is known for image processing.

An improved method of achieving regularisation in
 10 the specific case is to enforce regularization in (p, q) space along the direction of the shadow line for each shadowed pixel.

This is achieved by modifying the point to line distance term as follows:

$$15 \quad \hat{\epsilon}^{(i)} = \left(p - w\bar{P} - (1-w)P_e^{(i)} \right)^2 + \left(q - w\bar{Q} - (1-w)Q_e^{(i)} \right)^2 \quad (5)$$

This introduces a new variable w per shaded pixel that specifies a location along the shadow line of that pixel. The term is still quadratic with respect to p, q and w but this now allows regularizing of the
 20 solution in such a way that noise can be compensated. The variable w is related to parameter μ of (4) by:

$$w = \frac{e_3^T L^{-1} D_i \mathbf{c}}{e_3^T L^{-1} D_i \mathbf{c} + \mu e_3^T L^{-1} e_i} \quad (6)$$

25 As explained previously μ represents the value of c_i that would have been measured had the pixel not been in shadow in that image. The method puts a cost on the length of ∇w inside the shaded regions, the cost may

also be applied to a second or higher order term. As w is a proxy for μ , this corresponds to introducing smoothness in $I_i^T b$. This allows for the elimination of scratch artefacts while letting b have variability in the directions perpendicular to I_i , the results from the regularisation procedure are shown in FIG. 5c.

Step S103 of FIG. 3 is concerned with determining a shadow mask which allows determination of which pixels are shadowed.

In photometric stereo with four or more images it is possible to detect shadows by computing the scaled normal that satisfies the constraints in a least squares sense. If the residual of this least squares calculation is high, this implies that the pixel is either in a shadow or in a highlight. However, with three images this becomes impossible as the three constraints can always be satisfied exactly, leaving a residual of zero.

Recently, Chandraker, M et al "Shadowcuts: Photometric stereo with shadows" Proceedings of Computer Vision and Pattern Recognition, 17-22 June 2007 p 1-8 proposed a graph-cut based scheme for labelling shadows in photometric stereo with four or more images. Based on the constraint residual, a cost for assigning a particular shadow label to each pixel was calculated. This cost was then regularized in an MRF framework where neighbouring pixels were encouraged

to have similar shadow labels.

The basic characteristic of a shadow region is that pixel intensities inside it are dark. However this can also occur because of dark surface albedo. To
5 remove the albedo factor pixel intensities are divided with the magnitude of the intensity vector c . The cost for deciding that a pixel is occluded in the i -th image is $c_i/\|c\|$. This still leaves the possibility that a pixel is mistakenly classified whose normal is nearly
10 perpendicular to the i th illumination direction I_i . However in that case the pixel is close to being in a self shadow so the risk from misclassifying it is small. The cost for assigning a pixel to the non-shadowed set is given by

15
$$\sqrt{3} - \min_i \frac{c_i}{\|c\|}.$$

The costs are regularised in an MRF framework under a Potts model pairwise cost. This assigns a fixed penalty for two neighbouring pixels being given
20 different shadow labels. The MRF is optimized using the Tree Re-weighted message passing algorithm (Kolmogorov et al. "Convergent Tree-reweighted message passing for energy minimisation" IEEE Trans. Pattern Anal. Mach. Intell. 28 (2006) 1568-1583.)

25 FIG. 6a shows a face which is to be imaged and FIG. 6b shows the shadow mask determined using the above technique.

The above description has used three images to

generate a 3D image. A particularly useful method of obtaining these three images is to use colour photometric stereo. Here, an object to be images is illuminated from 3 different directions with radiation of different wavelengths and a cameras is used which can distinguish between the three wavelengths so that the three images can be taken at the same time. Typically, the three colours of red, green and blue are used as these colours can be reflected from the object and detected with minimal mixing from a single RGB image.

As colour photometric stereo is applied on a single image, one can use it on a video sequence without having to change illumination between frames.

In colour photometric stereo each of the three camera sensors (R, G, B) can be seen as one of the three images of classic photometric stereo. The pixel intensity of pixel (x, y) for the i-th sensor is given by:

$$c_i(x, y) = \sum_j (\mathbf{I}_j^T \mathbf{n}) \int E_j(\lambda) R(x, y, \lambda) S_i(\lambda) d\lambda \quad (7)$$

Note that now the sensor sensitivity S_i and spectral distribution E_j are different per sensor and per light source respectively. To be able to determine a unique mapping between RGB values and normal orientation a monochromatic surface is assumed. Therefore $R(x, y, \lambda) = \alpha(x, y) \rho(\lambda)$ where $\alpha(x, y)$ is the monochromatic albedo of the surface point and $\rho(\lambda)$ is

the characteristic chromaticity of the material. Let

$$v_{ij} = \int E_j(\lambda) \rho(\lambda) S_i(\lambda) d\lambda$$

and

5

$$\mathbf{v}_j = (v_{1j} v_{2j} v_{3j})^T$$

The scaled normal is:

$$\mathbf{b} = \alpha(x, y) \mathbf{n}$$

10

Then the vector of the three sensor responses at a pixel is given by

$$\mathbf{c} = [\mathbf{v}_1 \mathbf{v}_2 \mathbf{v}_3] [\mathbf{l}_1 \mathbf{l}_2 \mathbf{l}_3]^T \mathbf{b}.$$

Essentially each vector $\mathbf{v}_j$ provides the response
 15 measured by the three sensors when a unit of light from
 source j is received by the camera. If matrix $[\mathbf{v}_1 \mathbf{v}_2$
 $\mathbf{v}_3]$ is known then:

$$\hat{\mathbf{c}} = [\mathbf{v}_1 \mathbf{v}_2 \mathbf{v}_3]^{-1} \mathbf{c}.$$

20

The values of $\hat{\mathbf{c}}$ can be treated in exactly the same way as described above for standard photometric stereo.

The system may be calibrated in a number of ways. A calibration board of the type shown in FIG. 7 may be used. The calibration board 21 comprises a square of
 25 cloth 23 and a pattern of circles 25. Movement of the board 21 allows the homography between the camera 9 and the light sources 3, 5 and 7 to be calculated. Calculating the homography means calculating the light source directions relative to the camera. Once this
 30 has been done, zoom and focus can change during filming as these do not affect the colours or light directions.

The cloth 23 also allows the association between colour and orientation to be measured.

By measuring the RGB response corresponding to each orientation of the material the following matrix
5 may be estimated:

$$M = [\mathbf{v}_1 \ \mathbf{v}_2 \ \mathbf{v}_3] [\mathbf{l}_1 \ \mathbf{l}_2 \ \mathbf{l}_3]^T$$

The results from the initial calibration routine where an image is captured for various known
10 orientations of the board does not need to be performed for every possible board orientation as nearest neighbour interpolation can be used to determine suitable data for all orientations. It is possible to capture data from just 4 orientations in order to
15 provide calibration data for a 3x4 matrix. Good calibration data is achieved from around 50 orientations. However, since calibration data is easily collected it is possibly to obtain data from thousands of orientations.

20 In another calibration method, a mirror sphere is used to estimate light directions $\mathbf{l}_1$ $\mathbf{l}_2$ and $\mathbf{l}_3$. Secondly, three sequences of the object moving in front of the camera are captured. In each sequence, only one of the three lights is switched on at a time. In the
25 absence of noise and if the monochromatic assumption was satisfied, the RGB triplets acquired are multiples of $\mathbf{v}_j$ when light j was switched on.

A least squares fit can then be performed to

obtain the directions v_1 , v_2 and v_3 . To obtain the relative lengths of the three vectors the ratios of the lengths of the RGB vectors can be used. The length of v_j is set to the maximum length in RGB space, of all the triplets when light j was switched on.

FIG. 8 shows an image of a real object which in this case is a frog. FIG. 8a shows three images taken from three different illumination directions. In the experiment, the frog was imaged from five different directions, but only data from three images were used to produce the final three-dimensional image.

FIG. 8b shows the three images of figures 6a normalised in intensity using the expression:

$$c_i / \|c\|.$$

FIG. 8c shows the calculated non-shadowing cost of

$$\sqrt{3} - \min_i \frac{c_i}{\|c\|}.$$

FIG. 8d is the shadow mask calculated using the data shown in figures 8b and 8c. The light regions of the shadow mask show the regions where one of the images is occluded. These shaded regions are accommodated using the technique discussed above.

FIG. 8e is the generated three dimensional image of the object.

FIG. 9 shows video data of a white painted face illuminated by three coloured lights. Three reconstructed frames from a video sequence where shadows are not taken into account are shown in column

9a. Column 9b shows the corresponding three frames
where shadows are taken in account as explained above.
It can be seen that the nose reconstruction is
dramatically improved when correctly processing the
5 shadows, even though only two lights are visible in the
shadowed regions.

C L A I M S

1. An apparatus for determining the gradients of the surface normals of an object, the method comprising:

5 means to receive data of three 2D images of said object, wherein each image is captured under illumination from a different direction;

means to establish which pixels of the image are in shadow such that there is only data available from two images from these pixels;

means to determine a range of possible solutions for the gradient of the surface normal of a shadowed pixel using the data available for the two images; and

15 means to select a solution for the gradient using the integrability of the gradient field over an area of the object as a constraint and minimising a cost function.

2. An apparatus according to claim 1, wherein said apparatus further comprises means to acquire data, said means to acquire data comprising:

at least three light sources, irradiating the object from three different angles;

a camera provided to collect radiation from said three radiation sources which has been reflected from said object; and

25 an image processor configured to generate a depth map of the three dimensional object,

wherein each radiation source emits radiation at a different frequency and said image processor is configured to distinguish between the reflected signal from the three different radiation sources.

5 3. An apparatus according to claim 2, wherein said object is moving and wherein said camera is a video camera.

 4. An apparatus according to any one of claims 1 to 3, further comprising means to generate an image
10 from the determined surface normals.

 5. A method for determining the gradients of the surface normals of an object, the method comprising:

 receiving data of three 2D images of said object, wherein each image is taken under illumination from a
15 different direction;

 establishing which pixels of the image are in shadow such that there is only data available from two images from these pixels;

 determining a range of possible solutions for the
20 gradient of the surface normal of a shadowed pixel using the data available for the two images; and

 selecting a solution for the gradient at a pixel using the integrability of the gradient field over an area of the object as a constraint and minimising a
25 cost function.

 6. A method according to claim 5, wherein said method of selecting a solution for the gradient

comprises uses the constraint that the surface of the object is a continuous piecewise polynomial.

7. A method according to either of claims 5 or 6, wherein said method of selecting a solution is
5 performed as a least squares analysis.

8. A method according to any one of claims 5 to 7, wherein the cost function is expressed in terms of a regularisation parameter which is used to compensate for noise.

10 9. A method according to claim 8, wherein a cost is imposed on said regularisation parameter.

10. A method according to any one of claims 5 to 9, wherein the range of possible solutions is determined by expressing the surface normals in terms
15 of their gradients in a 2D gradient space such that the range of possible solutions lie along a straight line in said 2D gradient space.

11. A method according to claim 10, wherein the regularisation parameter is the length along the said
20 straight line of possible solutions from a fixed point on said line.

12. A method according to claim 5, wherein the range of possible solutions is determined by expressing the normal as the albedo scaled surface normal $\mathbf{b}$:
25

$$\mathbf{b} = L^{-1} \mathbf{c}$$

wherein $\mathbf{c}$ is $[c_1 \ c_2 \ c_3]$ and c_1 , c_2 and c_3 represent the respective intensity of the pixel

measured in each of the three 2D images and $L=[I_1 \ I_2 \ I_3]^T$ where I_1 , I_2 and I_3 are the vectors which represent the direction of the light source which illuminates the object for each of the three 2D images;
 5 wherein when the pixel in the i^{th} 2D image is in shadow b is expressed as:

$$\mathbf{b} = L^{-1} D_i \mathbf{c} + \mu L^{-1} \mathbf{e}_i$$

where $\mathbf{e}_i$ is the i^{th} column of the 3x3 identity matrix, D_i is the identity matrix with a zero in the
 10 i^{th} position in the diagonal and μ is a scalar parameter representing the value that c_i would have if it were not in shadow; and

projecting b into a 2D gradient space using the
 15 operator $P[x_1 \ x_2 \ x_3] = (x_1/x_3, x_2/x_3) = (p, q)$ such that for a point in shadow, the point (p, q) lies along the shadow line of the pixel defined as the line that joins $P[L^{-1} D_i \mathbf{c}]$ and $P[L^{-1} \mathbf{e}_i]$.

13. A method according to claim 12, wherein
 20 minimising the cost function is performed using a parameter w , where w is distance along the shadow line of point (p, q) from a fixed point on the shadow line, the method further comprising imposing a cost on ∇w .

14. A method according to any one of claims 5 to
 25 13, wherein establishing whether a pixel is in shadow in i^{th} of said three images is achieved by determining if the value of the intensity in the i^{th} image divided by the magnitude of the intensity vector c for all

three images is below a predetermined value.

15. A method according to claim 14, wherein establishing whether a pixel is in shadow is further determined by introducing a further constraint to
5 reduce the chance of a pixel being determined to be in shadow and a neighbouring pixel not in shadow.

16. A method according to any one of claims 5 to 15, wherein the data for the three 2D images is acquired by using three radiation sources, irradiating
10 the object from three different angles and wherein each radiation source emits radiation at a different frequency such that the data from said three radiation sources can be acquired at the same time.

17. A method of generating an image, the method
15 comprising: determining the gradients of the surface normals according to any one of claims 5 to 16; and generating an image from data of said surface normals.

18. A method according to claim 17, wherein the generated image is a three dimensional image.

20 19. A method of animating cloth, the method comprising:

imaging cloth according to the method of any one of claims 5 to 18 and animating a depth map generated from said determined gradients.

25 20. A computer running a computer program configured to perform the method of any one of claims 5 to 19.

1/5

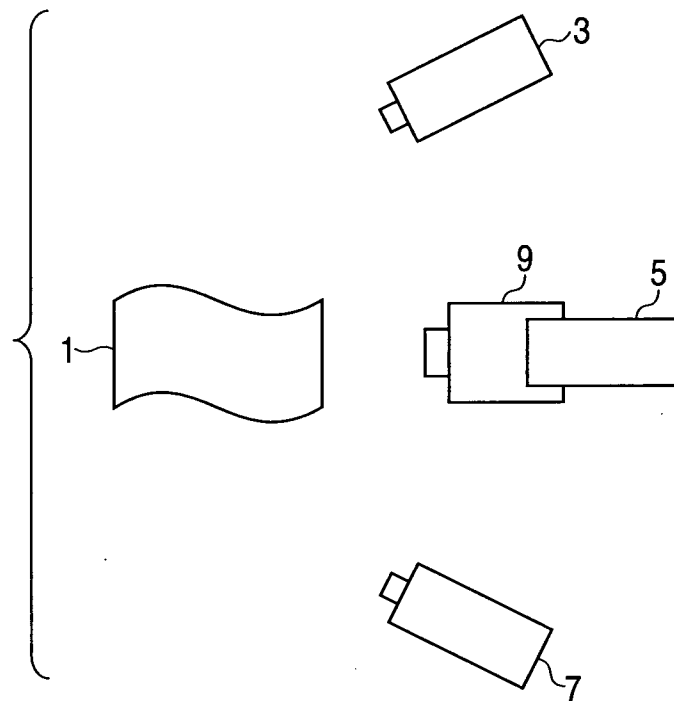


FIG. 1

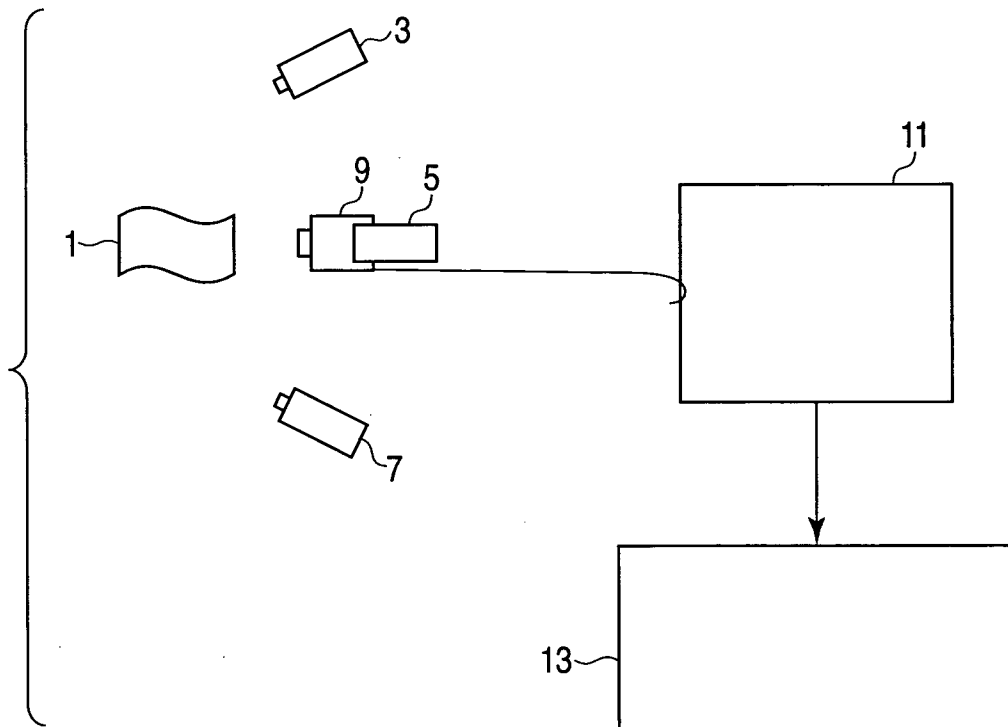


FIG. 2

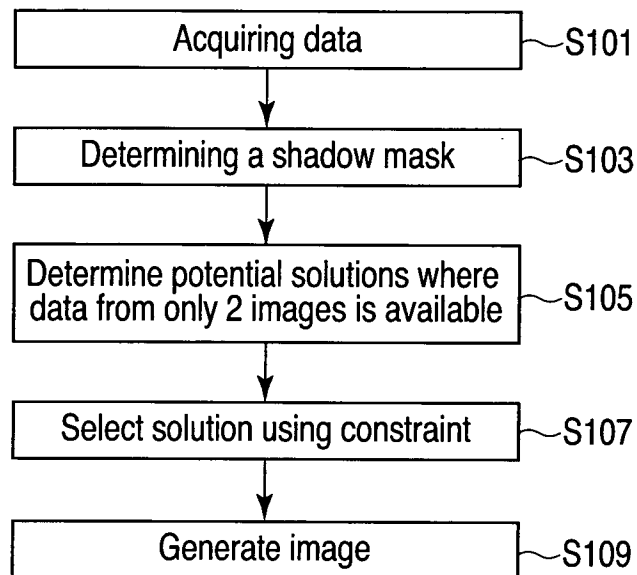


FIG. 3

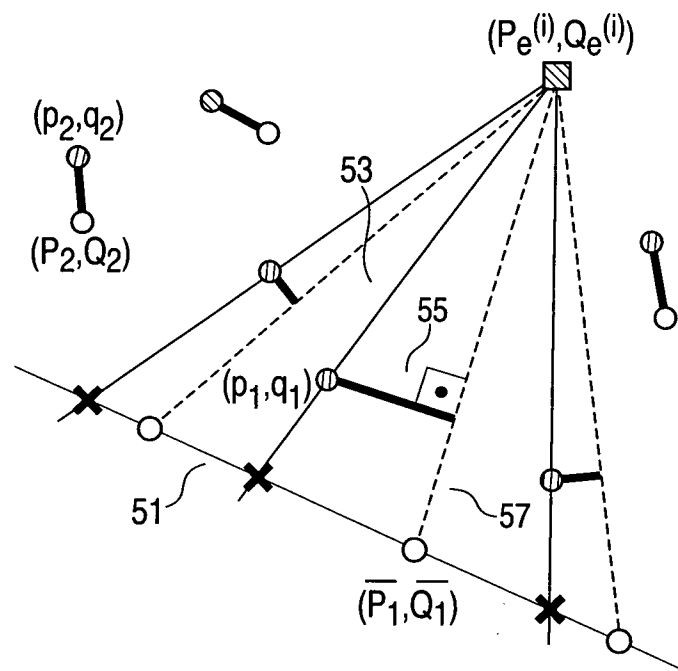


FIG. 4

3/5

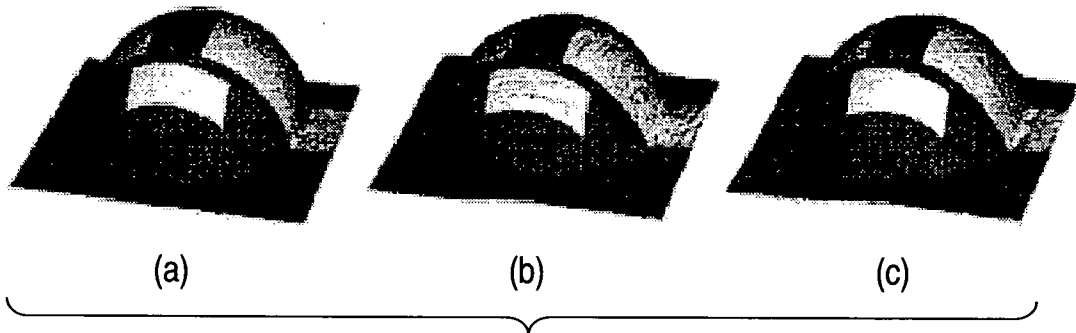


FIG. 5

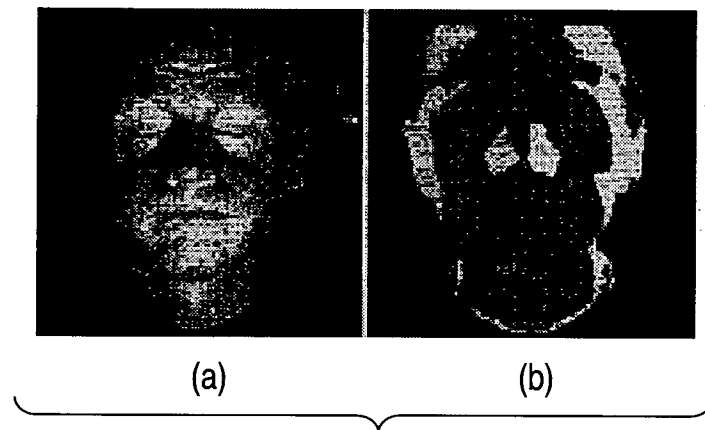


FIG. 6

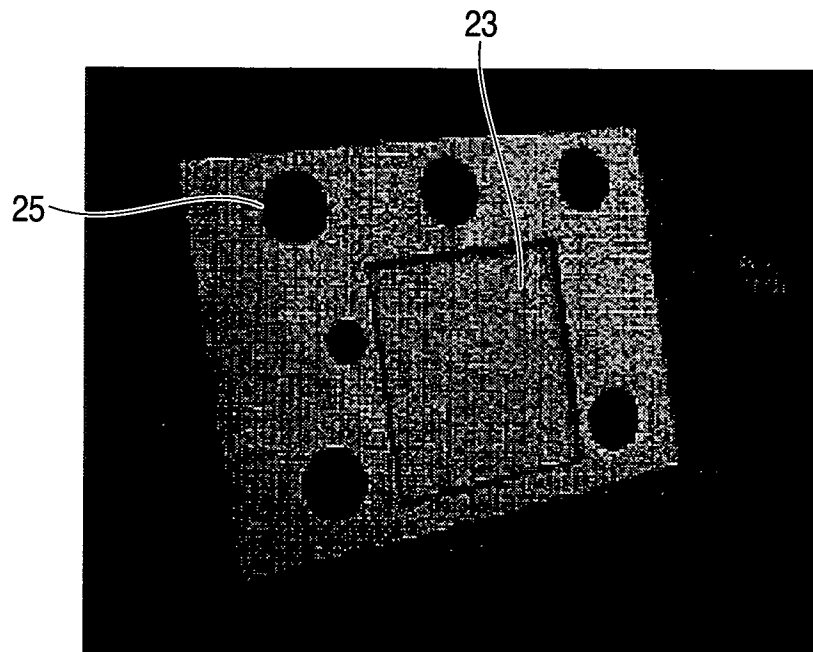


FIG. 7

4/5

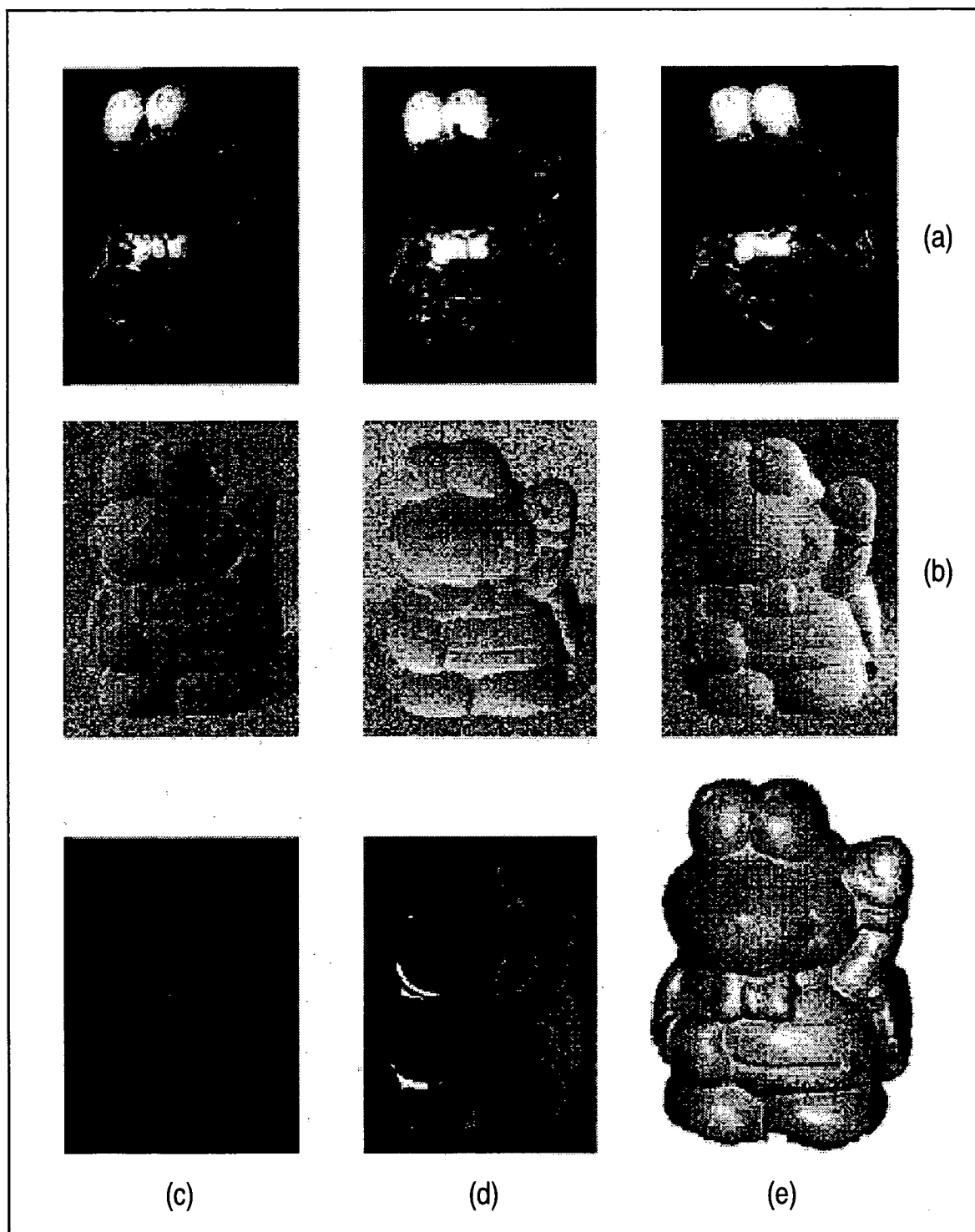


FIG. 8

5/5

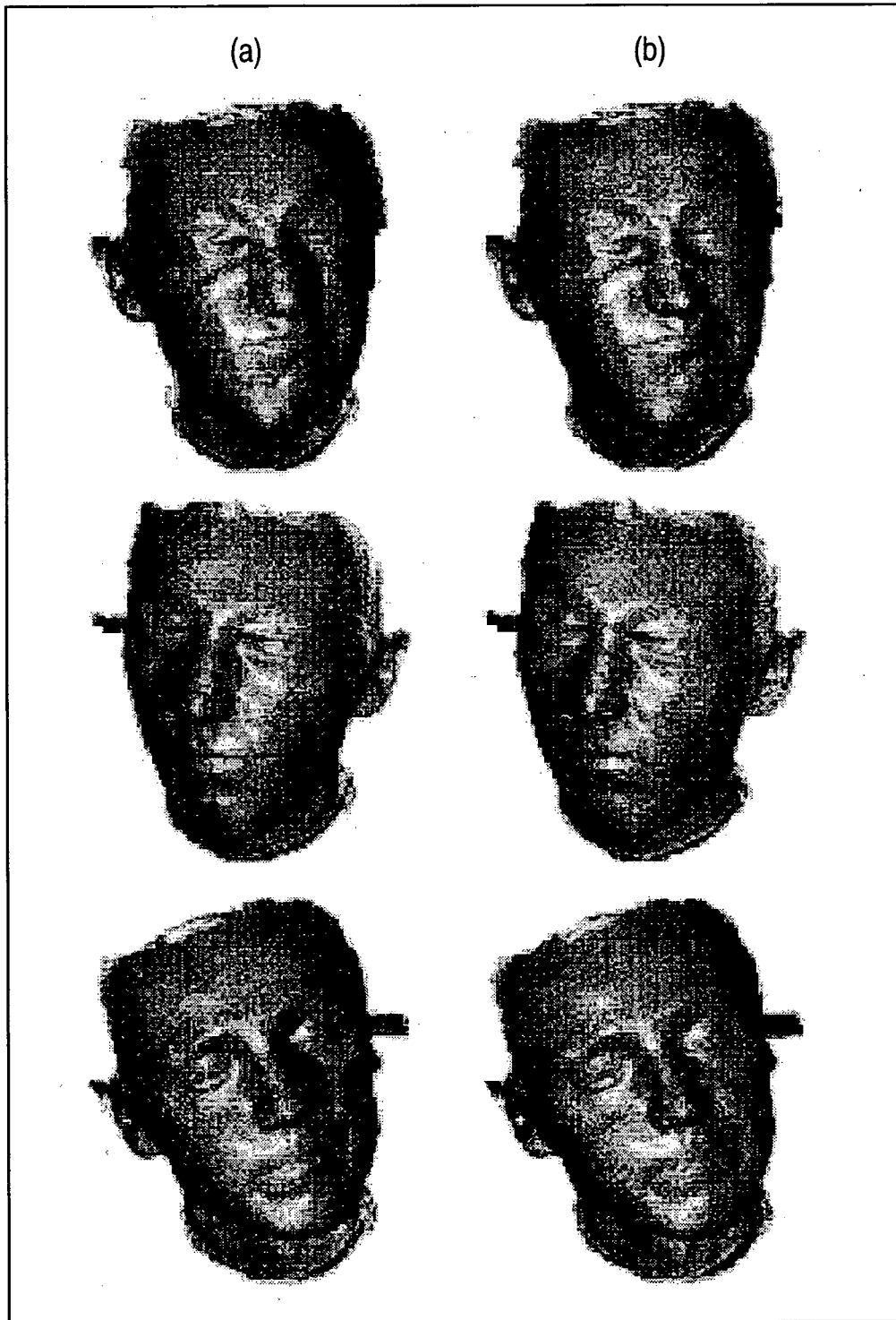


FIG. 9

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2009/067136

A. CLASSIFICATION OF SUBJECT MATTER

Int.Cl. G01B11/245 (2006.01) i, G06T1/00 (2006.01) i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

Int.Cl. G01B11/245, G06T1/00, H04N5/225

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Published examined utility model applications of Japan 1922-1996
 Published unexamined utility model applications of Japan 1971-2009
 Registered utility model specifications of Japan 1996-2009
 Published registered utility model applications of Japan 1994-2009

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

JSTPlus (JDreamII)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P, X	HERNANDEZ, Carlos, Shadows in Three-Source Photometric Stereo, Computer Vision-ECCV2008: 10th European Conference on Computer Vision, Marseille, France, October 12-18, 2008, Proceedings, Part I, 2008.11, Vol. 5302, p. 290-303	1-20
A	JP 6-105826 A (Shiseido Co., Ltd.) 1994.04.19, whole of document (Family:none)	1-20
A	EP 1814083 A1 (OMRON CORPORATION) 2007.08.01, whole of document & JP 2007-206797 A & US 2007/176927 A1 & CN 101013028 A	1-20



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

21.12.2009

Date of mailing of the international search report

28.12.2009

Name and mailing address of the ISA/JP

Japan Patent Office

3-4-3, Kasumigaseki, Chiyoda-ku, Tokyo 100-8915, Japan

Authorized officer

Shingo Ushida

Telephone No. +81-3-3581-1101 Ext. 3258

2S

3100

INTERNATIONAL SEARCH REPORT

International application No.
PCT/JP2009/067136

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JP 63-93076 A (Sony Corporation) 1988.04.23, whole of document (Family:none)	1-20
A	JP 2004-213159 A (Shigeru Ando) 2004.07.29, whole of document (Family:none)	1-20