Structured Light Projection
1. Introduction to Structured Lighting [1]

- The projection of **light patterns** into a scene is called **structured lighting**.
- The light patterns are projected onto the objects which lie in the field of view of the camera.
- The distance of an object to the camera or the location of an object in space can be determined through analyzing the **observed light patterns** in the images.

- A general idea of structured lighting consists in intersecting the ray $l$, formed by 3D point $(X,Y,Z)$ and image point $(x,y)$, with an additional ray $l'$ or an additional plane $\pi$ which leads to a unique reconstruction of the object point $(X,Y,Z)$.

http://www.stockeryale.com/i/lasers/structured_light.htm
1. Introduction to Structured Lighting [1]

- Structured lighting methods can be regarded as a modification of static binocular stereo. **One of the cameras is replaced by a light source which projects the light pattern into the scene.**
- The correspondence problem in the stereo vision does not exist any more since the triangulation is carried out by intersecting the projection ray (camera) and the light ray or plane (light source).
  ➔ **Correspondence between a bit of the light pattern and a pixel of the captured image is definite.**
2. Light Spot Projection, 2D Object Scene [1]

- A single light beam which is modeled as a ray (or line) is projected into the scene.
- The angle $\alpha$ and base distance $b$ is given by the calibration. The angle $\beta$ is defined by the projection geometry.

![Diagram of light spot projection](image.png)

**Figure 9.1:** Triangulation with a light spot projector.
2. Light Spot Projection, 2D Object Scene [1]

Figure 9.1: Triangulation with a light spot projector.

\[
\frac{d}{\sin(\alpha)} = \frac{b}{\sin(\gamma)}.
\]

From \( \gamma = \pi - (\alpha + \beta) \) and \( \sin(\pi - \gamma) = \sin(\gamma) \) it follows that

\[
\frac{d}{\sin(\alpha)} = \frac{b}{\sin(\pi - \gamma)} = \frac{b}{\sin(\alpha + \beta)}.
\]

Thus, the distance \( d \) is given as

\[
d = \frac{b \cdot \sin(\alpha)}{\sin(\alpha + \beta)}.
\]

Cartesian coordinates is carried out by

\[
X_0 = d \cdot \cos(\beta) \quad \text{and} \quad Z_0 = h = d \cdot \sin(\beta).
\]
2. Light Spot Projection, 3D Object Scene [1]

- A camera centered $XYZ$-coordinate system and an image plane lying at $Z=f$ are assumed. The considered object point $\mathbf{P}=(X_0, Y_0, Z_0)$ is projected onto a point $\mathbf{p}=(x, y)$ in the image plane.
- The optical center of the light source is assumed to be located on the $X$-axis.
2. Light Spot Projection, 3D Object Scene [1]

Using $p=(x,y)$,

$$\frac{X_0}{x} = \frac{Z_0}{f} = \frac{Y_0}{y} = k \quad \rightarrow \quad (kx, ky, kf)$$

Using pre-calibrated light source,

$$\tan(\alpha) = \frac{Z_0}{b - X_0} \quad \rightarrow \quad Z_0 = \tan(\alpha)(b - X_0)$$

By finding the cross-point,

$$kf = \tan(\alpha)(b - kx) \quad \rightarrow \quad k(f + x \tan(\alpha)) = b \tan(\alpha) \quad \rightarrow \quad k = \frac{b \tan(\alpha)}{f + x \tan(\alpha)}$$

By replacing $k$ of point coordinates with the acquired equation,

$$X_0 = \frac{xb \tan(\alpha)}{f + x \tan(\alpha)} \quad \quad \quad Y_0 = \frac{yb \tan(\alpha)}{f + x \tan(\alpha)} \quad \quad \quad Z_0 = \frac{fb \tan(\alpha)}{f + x \tan(\alpha)}$$
2. Light Spot Stereo Analysis [1]

- The method of light spot stereo analysis is based on the combination of the light spot technique with a method of static stereo analysis. A laser spot is projected at a chosen location onto the object surface, and this spot is acquired with two cameras from different directions.
2. Light Spot Stereo Analysis [1]

• The correspondence analysis is considerably simplified in both images compared to the static stereo analysis. The image of the laser spot on the object surface is assumed to be the brightest point in both images.
• The influence of the scene illumination on the laser point detection can be reduced by using special laser light filters for the image acquisition and by using image subtraction.
• Compared to the light spot technique, the light spot stereo analysis does not need any calibration of laser deflection system.
3. Light Stripe Projection [1]

- The **light stripe projection technique** or **light striping technique** represents an extension of the light spot projection technique. This technique projects a light plane into the object scene. The idea is to intersect the projection ray of the examined image point with the light plane.
- Therefore, a larger set of depth values can be recovered from a single image which results in a faster reconstruction process compared to the single spot techniques.

![Diagram of light stripe projection technique](image)
3. Light Stripe Projection [1]

The equation of light plane $\pi$

$$ n(X - P_0) = 0 $$

Where,

$$ n = \begin{bmatrix} \sin \alpha \cdot \sin \rho \\ \sin \alpha \cdot \cos \rho \\ -\cos \alpha \end{bmatrix} \quad P_0 (0, -b, 0) $$

The equation of $\text{Op}$

$$ \frac{X}{x} = \frac{Z}{f} = \frac{Y}{y} \quad \rightarrow \quad Q = (k \cdot x, k \cdot y, k \cdot f) $$

The intersection

$$ k = \frac{b \cdot \tan \alpha \cdot \cos \rho}{f - \tan \alpha (x \cdot \sin \rho + y \cdot \cos \rho)} $$

Therefore,

$$ X = \frac{x \cdot b \cdot \tan \alpha \cdot \cos \rho}{f - \tan \alpha (x \cdot \sin \rho + y \cdot \cos \rho)} \quad Y = \frac{y \cdot b \cdot \tan \alpha \cdot \cos \rho}{f - \tan \alpha (x \cdot \sin \rho + y \cdot \cos \rho)} \quad Z = \frac{f \cdot b \cdot \tan \alpha \cdot \cos \rho}{f - \tan \alpha (x \cdot \sin \rho + y \cdot \cos \rho)} $$

E-mail: hogijung@hanyang.ac.kr
http://web.yonsei.ac.kr/hgjung
3. Light Stripe Projection, 3D Laser Scanner [3]

http://www.youtube.com/watch?v=SPywgDBjM1Y
05:09

Structured Lighting

- Principle: Based on the relation between a single camera and a light source which projects a known pattern on the measuring scene.

- The correspondence problem is reduced:
  - Single dot:
    - No correspondence problem.
    - Scanning both axis.
  - Single slit:
    - No correspondence problem.
    - Scanning the axis orthogonal to the slit.
  - Stripe patterns:
    - No scanning.
    - Correspondence problem among slits.
  - Grid, multiple dots:
    - No scanning.
    - Correspondence problem among all the imaged features (points, dots, segments, ...).

- The matching between the projected pattern and the captured one can be uniquely solved codifying the pattern.

Pattern encoding/decoding (I)

- A pattern is **encoded** when after projecting it onto a surface, a set of regions of the observed projection can be easily matched with the original pattern. Example: pattern with two-encoded-columns

Pixels in red and yellow are directly matched with the pattern columns

- The process of matching an image region with its corresponding pattern region is known as **pattern decoding** → similar to searching correspondences

- **Decoding** a projected pattern allows a large set of correspondences to be easily found thanks to the *a priori* knowledge of the light pattern.

Pattern encoding/decoding (II)

- Two ways of encoding the correspondences: single and double axis codification ⇒ it determines how the triangulation is calculated.

**Decoding the pattern** means locating points in the camera image whose corresponding point in the projector pattern is *a priori* known.

### Coded structured light patterns: a classification proposal

<table>
<thead>
<tr>
<th>AXIS CODIFICATION</th>
<th>SCENE APPLICABILITY</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Single Axis</td>
<td>• Static Scenes</td>
</tr>
<tr>
<td>- Row-coded patterns</td>
<td>- Projection of a set of patterns.</td>
</tr>
<tr>
<td>- Column-coded patterns</td>
<td></td>
</tr>
<tr>
<td>• Both Axis</td>
<td>• Moving Scenes</td>
</tr>
<tr>
<td></td>
<td>- Projection of a unique pattern.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>PIXEL DEPTH</th>
<th>CODING STRATEGY</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Binary</td>
<td>• Periodical</td>
</tr>
<tr>
<td>• Grey Levels</td>
<td>- The codification of the tokens is repeated periodically.</td>
</tr>
<tr>
<td>• Colour</td>
<td>• Absolute</td>
</tr>
<tr>
<td></td>
<td>- Each token is uniquely encoded</td>
</tr>
</tbody>
</table>

Coded structured light patterns: a classification proposal

<table>
<thead>
<tr>
<th>TIME-MULTIPLEXING</th>
<th>Binary codes</th>
<th>n-ary codes</th>
<th>Gray code + Phase shifting</th>
<th>Hybrid methods</th>
</tr>
</thead>
<tbody>
<tr>
<td>Every point is encoded by a sequence of intensities. <strong>Time coding</strong></td>
<td>Posdamer et al., Inokuchi et al., Minou et al., Trobina, Valkenburg and McIvor, Skocaj and Leonardis, Rocchin et al., ...</td>
<td>Caspi et al., Horn and Kiryati, Osawa et al., ...</td>
<td>Bergmann, Sansoni et al., Wiora, Gühring, ...</td>
<td>K. Sato, Hall-Holt and Rusinkiewicz, Wang et al., ...</td>
</tr>
<tr>
<td><strong>SPATIAL CODIFICATION</strong></td>
<td>Non-formal codification</td>
<td>De Bruijn sequences</td>
<td>M-arrays</td>
<td></td>
</tr>
<tr>
<td>Every point is encoded by surrounding intensities. <strong>Window coding</strong></td>
<td>Maruyama and Abe, Durdle et al., Ito and Ishii, Boyer and Kak, Chen et al., ...</td>
<td>Hügli and Maître, Monks et al., Vuylsteke and Oosterlinck, Salvi et al. Lavoi et al., Zhang et al., ...</td>
<td>Morita et al., Petriu et al., Kiyasu et al., Spoelder et al., Griffin and Yee, Davies and Nixon, Morano et al., ...</td>
<td></td>
</tr>
<tr>
<td><strong>DIRECT CODIFICATION</strong></td>
<td>Grey levels</td>
<td>Colour</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Every point is encoded by its unique intensity. <strong>Direct coding</strong></td>
<td>Carrihill and Hummel, Chazan and Kiryati, Hung, ...</td>
<td>Tajima and Iwakawa, Smutny and Pajdla, Geng, Wust and Capson, T. Sato, ...</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Time Multiplexing**

- The time-multiplexing paradigm consists of projecting a series of light patterns so that every encoded point is identified with the sequence of intensities that receives.
- The most common structure of the patterns is a sequence of stripes increasing its length by the time → single-axis encoding.

**Advantages:**
- High resolution → a lot of 3D points
- High accuracy (order of µm)
- Robustness against colorful objects (using binary patterns)

**Drawbacks:**
- Static objects only
- Large number of patterns → High computing time

Example: 3 binary-encoded patterns which allows the measuring surface to be divided in 8 sub-regions.

Projected over time:

- Pattern 1
- Pattern 2
- Pattern 3

E-mail: hogijung@hanyang.ac.kr  
http://web.yonsei.ac.kr/hgjung

Time Multiplexing: Binary Codes

- Every encoded point is identified by the sequence of intensities that receives
- \( n \) patterns must be projected in order to encode \( 2^n \) stripes

Example: 7 binary patterns proposed by Posdamer & Altschuler

Codeword of this pixel: 1010010 \( \rightarrow \) identifies the corresponding pattern stripe

Time Multiplexing: N-ary Codes

- n-ary codes reduce the number of patterns by increasing the number of projected intensities (grey levels/colours) → increases the basis of the code

- The number of patterns, the number of grey levels or colours and the number of encoded stripes are strongly related → fixing two of these parameters the remaining one is obtained

Using a 4-ary code, 3 patterns are used to encode 64 stripes (Horn & Kiryati)

Using a binary code, 6 patterns are necessary to encode 64 stripes

Time-multiplexing: Gray code + Phase shifting

- A sequence of binary patterns (Gray encoded) are projected in order to divide the object in regions

- An additional periodical pattern is projected

- The periodical pattern is projected several times by shifting it in one direction in order to increase the resolution of the system \(\rightarrow\) similar to a laser scanner

Gühring’s line-shift technique

Example: three binary patterns divide the object in 8 regions

Without the binary patterns we would not be able to distinguish among all the projected slits

Every slit always falls in the same region

Time-multiplexing: hybrid methods

• In order to decode an illuminated point it is necessary to observe not only the sequence of intensities received by such a point but also the intensities of few (normally 2) adjacent points.

• The number of projected patterns reduces thanks to the spatial information that is taken into account.

Hall-Holt and Rusinkiewicz technique:

4 patterns with 111 binary stripes

Edges encoding: 4x2 bits (every adjacent stripe is a bit)

• The redundancy on the binary codification is eliminated.

Edge codeword: 10110101

Pattern 1

Pattern 2

Pattern 3

Pattern 4

Space Codification

- Spatial codification paradigm encodes a set of points with the information contained in a neighborhood (called window) around them.
- The codification is condensed in a unique pattern instead of multiplexing it along time.
- The size of the neighborhood (window size) is proportional to the number of encoded points and inversely proportional to the number of used colors.
- The aim of these techniques is to obtain a one-shot measurement system ⇒ moving objects can be measured.

Drawbacks:
- Discontinuities on the object surface can produce erroneous window decoding (occlusions problem).
- The higher the number of used colours, the more difficult to correctly identify them when measuring non-neutral coloured surfaces.

Advantages:
- Moving objects supported.
- Possibility to condense the codification into a unique pattern.

Space Codification: De Bruijn sequences

• A De Bruijn sequence (or pseudorandom sequence) of order \( m \) over an alphabet of \( n \) symbols is a circular string of length \( n^m \) that contains every substring of length \( m \) exactly once (in this case the windows are unidimensional).

\[
1000010111101001 \quad m=4 \text{ (window size)}
\]

\[
n=2 \text{ (alphabet symbols)}
\]

• Formulation:
Given \( P=\{1,2,\ldots,p\} \) set of colours.

• We want to determine \( S=\{s_1,s_2,\ldots,s_n\} \) sequence of coloured slits.

  Node: \( \{ijk\} \in VR_p^3 \)

  Number of nodes: \( p^3 \) nodes.

  Transition \( \{ijk\} \rightarrow \{rst\} : j = r, k = s \)

• The problem is reduced to obtain the path which visits all the nodes of the graph only once (a simple variation of the Salesman’s problem).
  – Backtracking based solution.
  – Deterministic and optimally solved by Griffin.

Space Codification: De Bruijn sequences

Example:

\[ p = 2 \]
\[ \text{VR}_p^3 = 2^3 = 8 \]

Path: (111),(112),(122),(222),(221),(212),(121),(211).

Slit color sequence: 111,2,2,2,1,2,1,1 \implies \text{Maximum 10 slits.}

‘1’ \rightarrow \text{Red}

‘2’ \rightarrow \text{Green}

Space Codification: De Bruijn sequences

• The De Bruijn sequences are used to define coloured slit patterns (single axis codification) or grid patterns (double axis codification)

• In order to decode a certain slit it is only necessary to identify one of the windows in which it belongs to

Zhang et al.: 125 slits encoded with a De Bruijn sequence of 5 colors and window size of 3 slits

Salvi et al.: grid of 29×29 where a De Bruijn sequence of 3 colors and window size of 3 slits is used to encode the vertical and horizontal slits

Spatial codification: M-arrays

- An m-array is the bidimensional extension of a De Bruijn sequence. Every window of \( w \times h \) units appears only once. The window size is related with the size of the m-array and the number of symbols used.

\[
\begin{array}{cccccccc}
0 & 0 & 1 & 0 & 1 & 0 \\
0 & 1 & 0 & 1 & 1 & 0 \\
1 & 1 & 0 & 0 & 1 & 1 \\
0 & 0 & 1 & 0 & 1 & 0 \\
\end{array}
\]

Example: binary m-array of size 4×6 and window size of 2×2

Morano et al. M-array represented with an array of coloured dots

M-array proposed by Vuylsteke et al. Represented with shape primitives

Direct Codification

- Every encoded pixel is identified by its own intensity/colour
- Since the codification is usually condensed in a unique pattern, the spectrum of intensities/colours used is very large
- Additional reference patterns must be projected in order to differentiate among all the projected intensities/colours:
  - Ambient lighting (black pattern)
  - Full illuminated (white pattern)
  - ...

- Advantages:
  - Reduced number of patterns
  - High resolution can be in theory achieved (all points are coded)

- Drawbacks:
  - Very noisy in front of reflective properties of the objects, non-linearities in the camera spectral response and projector spectrum ⇒ non-standard light emitters are required in order to project single wave-lengths
  - Low accuracy (order of 1 mm)
Direct Codification: Grey Level

- Every encoded point of the pattern is identified by its intensity level

Carrihill and Hummel Intensity Ratio Sensor: fade from black to white

- Every slit must be projected using a single wave-length

Requirements to obtain high resolution

- Cameras with large depth-per-pixel (about 11 bits) must be used in order to differentiate all the projected intensities

Direct Codification: Color

- Every encoded point of the pattern is identified by its colour

Tajima and Iwakawa rainbow pattern
(the rainbow is generated with a source of white light passing through a crystal prism)

T. Sato patterns capable of cancelling the object colour by projecting three shifted patterns
(it can be implemented with an LCD projector if few colours are projected → drawback: the pattern becomes periodic in order to maintain a good resolution)

http://eia.udg.es/~qsalvi/codedLight.mpeg

02:03
The digital micromirror array (DMD) in a DLP projector is capable of switching mirrors "on" and "off" at high speeds ($10^6$/s). An off-the-shelf DLP projector, however, effectively operates at much lower rates (30-60Hz) by emitting smaller intensities that are integrated over time by a sensor (eye or camera) to produce the desired brightness value. Our key idea is to exploit this temporal dithering of illumination, as observed by a high-speed camera. The dithering encodes each brightness value uniquely and may be used in conjunction with virtually any active vision technique.
At the heart of every DLP® projection system is an optical semiconductor known as the DLP® chip, which was invented by Dr. Larry Hornbeck of Texas Instruments in 1987.

A DLP® chip's micromirrors are mounted on tiny hinges that enable them to tilt either toward the light source in a DLP® projection system (ON) or away from it (OFF)-creating a light or dark pixel on the projection surface.

The bit-streamed image code entering the semiconductor directs each mirror to switch \textbf{on and off up to several thousand times per second}. When a mirror is switched on more frequently than off, it reflects a light gray pixel; a mirror that's switched off more frequently reflects a darker gray pixel.
Televisions, home theater systems and business projectors using DLP® technology rely on a single chip configuration like the one described above.

White light passes through a color filter, causing red, green, blue and even additional primary colors such as yellow cyan, magenta and more to be shone in sequence on the surface of the DLP® chip. The switching of the mirrors, and the proportion of time they are 'on' or 'off' is coordinated according to the color shining on them. Then the sequential colors blend to create a full-color image you see on the screen.

http://www.dh.aist.go.jp/~shun/research/dlp/vid/025–v01.avi

05:01

- The projector is assumed to have a narrow depth of field (wide aperture) while the camera is assumed to have a wide depth of field (small aperture).
- Using the lens law, the diameter $D$ of the blur circle on the camera's image plane can be written as

$$D = \pm 2 f_c r \left(\frac{1}{u} - \frac{1}{u_f}\right)$$

where $f_c$ is the camera focal length, $r$ is the radius of the projector lens, $u_f$ is the distance of the focal plane from the lens, and $u$ is the distance of the surface from the lens.

Figure 3: System used to acquire images for refocusing.

\[ D = \pm 2f_c r \left( \frac{1}{u} - \frac{1}{u_f} \right) \]

\begin{align*}
&\text{u=50cm} & \text{u=65cm} & \text{u=80cm} \\
&\text{u=95cm} & \text{u=110cm} & \text{u=125cm} \\
\end{align*}

\begin{figure}
\centering
\includegraphics[width=\textwidth]{fig4.png}
\caption{Geometric and radiometric properties of projected dots. (top) Camera images of a square dot of 3 x 3 pixels projected onto different depths. (center) The dot width \( D_w \) and radiance \( I_w \) were measured from the images and compared to the values predicted by our models. (bottom) The radiance variation of a projected dot, within a chosen working range of the system, may be controlled by changing the parameters of the setup, such as the distance \( u_f \) of the focal plane or the width \( w \) of the dots.}
\end{figure}

Figure 2: The steps involved in the refocusing method. (a) Acquired image; (b) image after removal of the projected dots; (c) sparse depth map estimated from the removed dots; (d) color over-segmentation of the dot-removed image in (b); (e) merging of segmented regions using the sparse depth map in (c); (f) depth map after boundary refinement using a matting algorithm; (g-i) refocused images with different depths of field; and (j) refocused image for an image taken with new lighting.

05:37
7. KINECT [7]

Figure 1-6. Kinect external component identification—Output: A) IR (infrared) structured-light laser projector, B) LED indicator, and K) motor to control tilt-in base. Input: F-I) Four microphones, C-D) two cameras (RGB and IR), and E) one accelerometer.
Fig. 3.2 Matricial active triangulation flow: pixel $p_A$ (green dot) is coded in the pattern. The pattern is projected to the scene and acquired by $C$. The 3D point associated to $p_A$ is $P$ and the conjugate point of $p_A$ (green dot) in $I_K$ is $p_C$ (blue dot). The correspondence estimation algorithm (red dashed arrow) estimates the conjugate points.
7. KINECT [8]

Fig. 1.7 Active triangulation by a system made by a camera C (blue) and a light projector A (green).

If the active system is calibrated and rectified, \( p_C \) has coordinates \( p_C = [u_C = u_A + d, v_C = v_A] \). \[ z = \frac{b|f|}{d} \]

Fig. 3.3 Example of \( I_K, \hat{D}_K \) and \( \hat{Z}_K \) acquired by the Kinect™ range camera.
7. KINECT [8]

Light Coding Techniques

Since for a calibrated and rectified setup conjugate points lie on horizontal lines, the coding problem can be independently formulated for each row in order to keep as low as possible the cardinality of the total number of possible code-words.
Light Coding Techniques: Artifacts Affecting the Projected Patterns

- Slanted surface: perspective distortion
- Surface color
7. KINECT [8]

Light Coding Techniques: Artifacts Affecting the Projected Patterns

- Strong external illumination,
- Reflective surface

- Occlusion
7. KINECT [8]

Light Coding Techniques

Fig. 3.5 Examples of coding strategies: a) direct coding; b) time-multiplexing coding; c) spatial-multiplexing coding.
7. KINECT [8]

Light Coding Techniques

The Kinect™ range camera adopts a spatial-multiplexing approach that allows to robustly capture dynamic scenes at high frame-rate (30 [fps]).


In a number of real situations, conjugate points can be simply detected from the covariance maximum between a window centered around the specific pixel \( p_C \) in \( I_K \) and all the possible conjugates \( p^i_A \) in the same row of the projected pattern, as shown by the example of Figure 3.8.

![Figure 3.8](image-url) Covariance of the pattern acquired by the Kinect™ for \( p^i_A \) with coordinates \( p^i_A = [19, 5]^T \) and \( p^j_A = [n, 5]^T \) and \( n \in \{1, \ldots, 200\} \).
7. KINECT [8]

Light Coding Techniques

Fig. 3.7 Pattern acquired by the Kinect™ IR camera.
7. KINECT [8]

Human Pose Estimation for Kinect

Kinect for Xbox 360 and Windows makes you the controller by fusing 3D imaging hardware with markerless human-motion capture software. Our group investigates such software. Mixing computer vision, graphics, and machine learning techniques, we look at how to build algorithms that can learn to recognize human poses quickly and reliably.

Images


E-mail: hogijung@hanyang.ac.kr
http://web.yonsei.ac.kr/hgjung
References