Depth estimation using structured light flow
— analysis of projected pattern flow on an object's surface —

Ryo Furukawa
Hiroshima City University, Japan

Ryusuke Sagawa
AIST, Japan

Hiroshi Kawasaki
Kyushu University, Japan

Abstract

Shape reconstruction techniques using structured light have been widely researched and developed due to their robustness, high precision, and density. Because the techniques are based on decoding a pattern to find correspondences, it implicitly requires that the projected patterns be clearly captured by an image sensor, i.e., to avoid defocus and motion blur of the projected pattern. Although intensive researches have been conducted for solving defocus blur, few researches for motion blur and only solution is to capture with extremely fast shutter speed. In this paper, unlike the previous approaches, we actively utilize motion blur, which we refer to as a light flow, to estimate depth. Analysis reveals that minimum two light flows, which are retrieved from two projected patterns on the object, are required for depth estimation. To retrieve two light flows at the same time, two sets of parallel line patterns are illuminated from two video projectors and the size of motion blur of each line is precisely measured. By analyzing the light flows, i.e. lengths of the blurs, scene depth information is estimated. In the experiments, 3D shapes of fast moving objects, which are inevitably captured with motion blur, are successfully reconstructed by our technique.

1. Introduction

3D shape reconstruction techniques by projecting special light to objects (*i.e.*, active lighting techniques) have been important research topics. Since they have significant advantages, *e.g.*, a textureless object can be robustly reconstructed with dense and high accuracy irrespective of light conditions, commercial 3D scanners are primarily based on active lighting techniques. Among two important active lighting techniques, such as structured light and photometric stereo, structured light techniques, which encode the positional information of projector pixels into projected patterns, have been widely researched and developed due to their robustness, high precision, and density.

Recently, capturing 3D shapes of moving objects from a moving sensor has become increasingly important, such as measuring a scene from self-driving cars, etc., and solution is strongly desired. The one-shot scanning technique, which

is one of structured light techniques, is a promising technique because it required just single image and unaffected by motion. In general, one-shot scanning techniques embed the pattern's positional information into a small area of the pattern for decoding [11, 6, 10, 20], and thus, the projected pattern tends to be complicated to increase robustness and density. If object/sensor motion is faster than an assumption, the reflected patterns are easily blurred out, resulting in unstable and inaccurate reconstruction.

To solve such motion blur problem, one may consider applying an algorithm developed for deblurring photographs. However, although the phenomena look similar, the optical processes are completely different[1] and such algorithms cannot be applied to structured light. In this paper, we analyze the information that is implied in the motion of a projected pattern (hereafter "*light flow*") to recover the shape. In fact, it is shown that *light flows* are explained by three factors such as depth, normal, and velocity of the object surface and extracting one of them from a single *light flow* is impossible, but possible with two *light flows*. In this paper, we propose a technique to decouple the three values from multiple *light flows*. It is also revealed that *light flows* includes only little depth information, and thus, accurate detection of motion blur is required for practical depth estimation. To achieve this, we project the pattern of parallel lines onto the objects so that a blur size is precisely measured as the width of the band of motion blur of each line.

The contributions of this paper is as follows: (1) Depth from *light flows*, which does not require a decoding process, or a matching process, is presented. (2) Shape recovery of "fast motion," where motion speed is faster than the shutter speed causing blurred patterns, is achieved. (3) Two algorithms using two different patterns are proposed to reconstruct shapes of fast moving objects. By using our technique, multiple fast moving objects, which are almost impossible to scan with state of the art techniques, can be reconstructed, for example, rotating blades of a fan are reconstructed by using a normal camera and a video projector.

---

[1]Imagine the case where a planar board moves perpendicular direction to its surface normal. It is obvious that no blur effect appears on projected pattern.
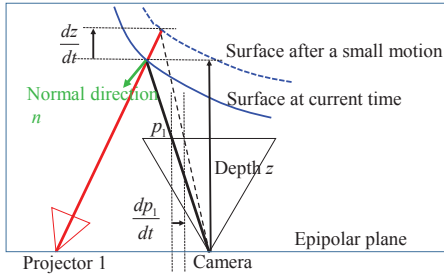
Figure 1. Motion of the object surfaces and patterns. Note that his figure represents a 2D section of the scene by an epipolar plane.

2. Related works

There are two major shape recovery techniques using active light, such as photometric stereo and structured light. Photometric stereo recovers the surface normal of each pixel using multiple images captured by a camera while changing the light source direction [5, 4, 3, 9]. Although photometric stereo can recover surface normals, they cannot recover absolute 3D distances; thus, their applicability is limited. The structured light technique has been used for practical applications [16, 22, 12]. Two primary approaches to encode positional information into patterns are temporal and spatial encoding. Because temporal encoding requires multiple images, it is not suitable for capturing moving objects [17, 18]. Spatial encoding requires only a single image and is possible to capture fast-moving objects [11, 6, 10, 20]. For example, some commercial devices realize 30 fps at VGA resolution [11]. Another technique can scan the shape of a rotating blade [6]. Although such techniques can capture fast moving objects, they still require the patterns to be captured clearly. If the pattern is blurred due to unexpected fast motion, the accuracy is reduced or reconstruction fails. Kawasaki *et al.* needed to use a high-speed camera to measure a rotating blade [6]. In contrast, our technique uses a different approach, *i.e.*, shapes are reconstructed from motion blur of a projected pattern. This helps to decrease the difficulty for scanning fast moving objects and allows off-the-shelf devices to be used for such extreme conditions.

Another limitation of structured light techniques is that they are highly dependent on the locally decoded pattern information. If decoding the positional information fails for some reason, such as noise, specularity, blur, etc., shape reconstruction will subsequently fail. To avoid this limitation, some techniques are based on geometric constraints rather than decoding [15, 13, 8, 20] or active usage of defocus effect by coded aperture [7]. One problem with such techniques is that, because they are heavily dependent on geometric constraints or special devices, strong motion blur cannot be handled.

Our technique uses relationship between two overlapped

patterns and such techniques have been researched. The Moire method is a well-known technique where a high-frequency pattern is projected on an objected and is observed by a high-frequency gate, which generates Moire patterns [19, 2]. Recently, a technique that uses the intersection of multiple parallel line patterns was proposed to achieve an entire scene reconstruction [1]. Both techniques assume a static scene and recover shape from geometric constraints, whereas our technique recovers shape from a temporal gradient. To the best of our knowledge, no such technique has been proposed previously.

3. Analysis of projected pattern flow

3.1. Overview

In this section, we briefly overview information that can be obtained from apparent motion of projected patterns. We assume that the projectors and the camera are relatively static and calibrated. If the object moves, the observed patterns move along epipolar lines. For the target scene, we assume that the regions around the measured 3D points can be regarded as locally planer and depth of the points are changing. Otherwise, we would not be able to observe continuous motions of patterns moving with the scene motions.

Without losing generality, the relationships between the motions of the target surfaces and the patterns can be considered within the epipolar planes. Fig. 1 shows the relationship.

Let the apparent position of the pattern be p, and the apparent motion of the pattern (*i.e.*, light flow) be $\frac{dp}{dt}$. $\frac{dp}{dt}$ depends on the surface depth z, surface normal n, and the depth velocity $\frac{dz}{dt}$ as shown in Fig. 1. $\frac{dp}{dt}$ is proportional to $\frac{dz}{dt}$, and represented as

$$\frac{dp}{dt} = F(z, n)\frac{dz}{dt}, \tag{1}$$

where F is a nonlinear function that can be defined from the projection model and calibration parameters of the projectors and cameras (*i.e.*, the information of the epipolar geometry of the projector-camera pair is included in F). By limiting the geometry within a epipolar plane, the normal direction at a point on the surface can be represented by a scalar variable, thus, $F(z, n)$ is a 2D real function.

Here, we assume that we have N projectors. The light flows can be observed for each projector.

In case of $N = 3$, where we use three patterns from three projectors, the observed light flows are $\frac{dp_i}{dt}$ where $i = 1, 2, 3$, and we obtain three equations

$$\frac{dp_i}{dt} = F_i(z, n)\frac{dz}{dt}, (i = 1, 2, 3) \tag{2}$$

where F_i is the same function with Eqn. 1 for pattern i. By

eliminating $\frac{dz}{dt}$, we obtain two equations:

$$\frac{dp_1}{dt} \Big/ \frac{dp_2}{dt} = F_1(z,n)/F_2(z,n)$$
$$\frac{dp_2}{dt} \Big/ \frac{dp_3}{dt} = F_2(z,n)/F_3(z,n). \qquad (3)$$

The unknown variables are z and n, and F_1, F_2, F_3 are known functions since they are defined from calibration parameters. Since we have two equations with two unknowns, we can estimate z and n by numerically solving these equations.

Note that, from Eqn. 3, geometrical information are obtained from "ratio" of the pattern motions $\frac{dp_i}{dt} \big/ \frac{dp_j}{dt} (i \neq j)$, rather than $\frac{dp_i}{dt}$ themselves.

In case of $N = 2$, we obtain only one equation by eliminating $\frac{dz}{dt}$, which is $\frac{dp_1}{dt} \big/ \frac{dp_2}{dt} = F_1(z,n)/F_2(z,n)$. Since we have one equation with two unknowns, we cannot determine z, n for this case.

However, if we use additional information such as knowledges of the projected patterns, we can estimate z. As described in the following sections, we propose to use uniformly spaced parallel patterns from "two" projectors to estimate depths. Moreover, we also show a special case of $N = 2$, where the two patterns are projected from a "single" projector. In the case of single projector, we should use non-uniformly spaced parallel pattens.

In case of $N = 1$, we use only one pattern, we obtain only $\frac{dp_1}{dt} = F_1(z,n)\frac{dz}{dt}$. Obviously, we cannot solve $z, n, \frac{dz}{dt}$ from the form.

As explained here, we can estimate depths if we can observe three patterns from three projectors. However, to achieve this, three patterns should be decoupled from captured images, which is not an easy task. One solution may be using three color channels, however, crosstalks between color channels are problematic. On the other hand, decoupling two patterns are relatively stable, since crosstalks between red-blue channels are small. Thus, in this paper, we propose a technique which requires only two colors, *i.e.* two projection patterns, for shape reconstruction.

3.2. Depth estimation with two-pattern case

In this section, we explain depth estimation with two patterns using uniformly spaced parallel lines. To achieve this, we estimate the *light flows* "in the projectors' image coordinates" instead of in the camera's image coordinates. Then, we can avoid considering normals of the surfaces as shown in the the following discussions.

First, we assume a pixel p while observing a target object from a camera. We define the ray from the camera corresponding to p as $\mathbf{r}$, the intersection between the object and $\mathbf{r}$ as $\mathbf{s}$, and the depth of $\mathbf{s}$ as z. Thus,
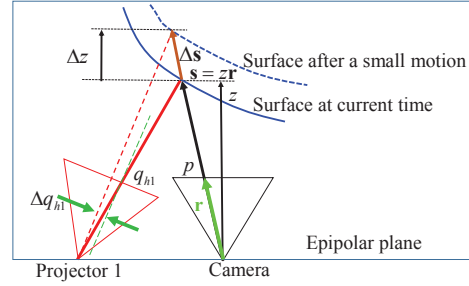
$$\mathbf{s} = z \, \mathbf{r}. \qquad (4)$$

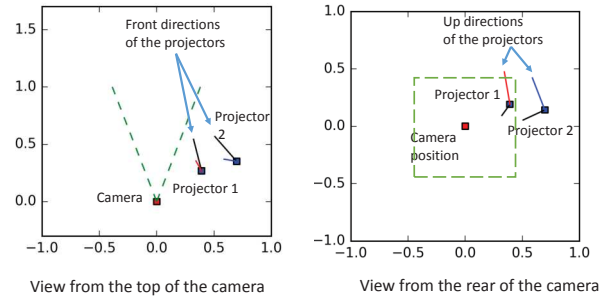Figure 2. Light flows observed with respect to the pattern coordinates q_h.

Figure 3. Positions of the camera and the projectors.

We assume that the object is moving with respect to the camera; thus, depth z changes. Let a small displacement of z be Δz. Then, the position of $\mathbf{s}$ is expressed as follows:

$$\mathbf{s} + \Delta\mathbf{s} = (z + \Delta z)\mathbf{r}. \qquad (5)$$

Fig. 2 shows the relationships between the symbols used in this section.

The projector can be geometrically regarded as an "inverse camera," *i.e.*, the relationship between a 3D coordinate $\mathbf{s}$ of point s and the 2D coordinate on the pattern pixel coordinate $\mathbf{q}$ that illuminates s can be formulated in the same as the projections of the camera. This can be expressed as follows:

$$\mathbf{q} = \mathbf{f}(\mathbf{R}\mathbf{s} + \mathbf{t}), \qquad (6)$$

where $\mathbf{s}$ is the coordinates of s in the camera coordinate system; $\mathbf{R}$ and $\mathbf{t}$ are the rotation matrix and the translation vector of the rigid transformation from the camera coordinates to the projector coordinates, respectively; and $\mathbf{f}$ is a perspective projection function that is the same as the normalized camera. We assume that the pattern image illuminated from the projector is parallel lines that is near the vertical direction, and the epipolar lines are assumed to be near the horizontal direction. Thus, the pattern intensity only varies with the horizontal coordinate, which we define as q_h. Then,

$$q_h = f_h(\mathbf{R}\mathbf{s} + \mathbf{t}) = f_h(z\mathbf{R}\mathbf{r} + \mathbf{t}), \qquad (7)$$

where $f_h((x,y,z)^{\mathrm{T}}) \equiv (-x/z)$.
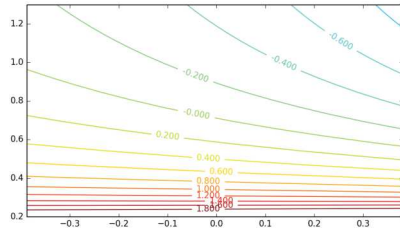
(a): $h(z)$ for the setup shown in Fig. 3.
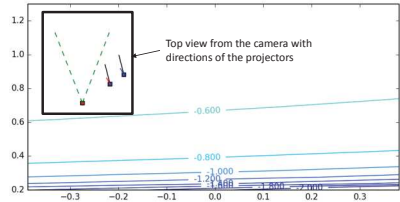
(b): Bad setup where gradient of $h(z)$ becomes small.

Figure 4. Function from the depth to the log ratio of *light flows* ($h(z)$ of Eqn. 12) : (a) The experimental setup used in this paper. The horizontal axis is the x position of the camera in the coordinates of the normalized camera. (b) A bad setup where projectors 1 and 2 are placed in parallel.

Let the relationship between the depth z and q_h be substituted by a function g. Thus,

$$q_h = f_h(z\mathbf{R}\mathbf{r} + \mathbf{t}) = g(z). \tag{8}$$

Here assume that pattern coordinate q_h is changed by Δq_h for a small displacement Δz of z. Δq_h can be observed as *light flow*. Then, the ratio between the changes can be approximated by the derivative of g:

$$\frac{\Delta q_h}{\Delta z} \approx g'(z). \tag{9}$$

Function $g'(z)$ changes depending on the depth z.

Since two projectors are assumed in this analysis, there are two pattern coordinates for each projector, such as q_{h1} and q_{h2}, and two functions for each projector, such as g_1 and g_2. Thus,

$$\frac{\Delta q_{h1}}{\Delta z} \approx g_1'(z), \frac{\Delta q_{h2}}{\Delta z} \approx g_2'(z) \tag{10}$$

can be reduced to

$$\frac{\Delta q_{h1}}{\Delta q_{h2}} \approx \frac{g_1'(z)}{g_2'(z)}. \tag{11}$$

Here, function h is defined as follows:

$$h(z) \equiv \log\left(\frac{g_1'(z)}{g_2'(z)}\right), \tag{12}$$

where $h(z)$ is a function of z and does not depend on Δz. If $h(z)$ is a monotonic function in the domain of the working distances of z, there is an inverse function h^{-1} for this domain, and depth z can be estimated from the ratio of the two *light flows* (Δq_{h1} and Δq_{h2}) as follows:

$$z \approx h^{-1}\left(\log\left(\frac{\Delta q_{h1}}{\Delta q_{h2}}\right)\right). \tag{13}$$

Function $h(z)$ depends on the pixel position p, *i.e.*, for different pixel positions, h becomes different functions. Thus, let $h(z)$ for a pixel position p be expressed as $h_p(z)$. Similarly to F, the information of the epipolar geometry is included in h, where the explicit dependency is based on Eqn. 8.

An example of the camera setup is shown in Fig. 3 (this setup is also used in the later experiments), and $h_p(z)$ for this example setup is shown in Fig. 4(a), where p is on a horizontal line in a camera image (including the image center) and the vertical axis represents z. As can be seen, $h(z)$ actually varies depending on z and h_p^{-1} can be defined for each pixel. In real implementation, we can sample pairs of z and the function value $h_p(z)$ as $\{(z, h_p(z))|z \in \{z_1, z_2, \cdots, z_n\}\}$ and h_p^{-1} can be approximated by interpolation of the samples. Then, depth estimation can be processed efficiently.

Note that the positional setup of the projector and the camera affects severely to the sensitivity of the depth estimation. For example, an example of a bad setup is shown in Fig. 4(b), where the two projectors are placed in parallel. Although this setup look similar to the setup of Fig. 3, the gradient of $h_p(z)$ becomes much smaller, thus, depth estimation using h_p^{-1} becomes much sensitive to observation noises. Specifically, if the projectors and the camera are placed in perfect fronto-parallel configuration, $h_p(z)$ becomes constant and h_p^{-1} does not exist.

In the proposed method, the required input value is the ratio of the *light flows* $\Delta q_{h1}/\Delta q_{h2}$, whereas the absolute values of q_{h1} or q_{h2} are not required. If the pattern image is repetitive, the depth can be estimated from only the relative local changes as described later. This means that we do not need to encode the absolute pixel position into the pattern image and this is an important advantage of our method.

The *light flows* are observed as the motion on the image planes of the projectors in Fig. 2, rather than the motion on the image plane of the camera as shown in Fig. 1. Observing *light flows* on the projectors' image planes is possible if we use knowledge of the projected pattern, such as uniformly spaced parallel lines. As described later, we use the uniform intervals as "scales" on the projectors' image planes. Also, in the discussion of this section, normal directions of the object surface are not considered. This is because the relationships between the light flows and the object motion are considered on the fixed "ray" from the camera along $\mathbf{r}$.

$h_p(z)$ does not depend on the intrinsic parameters of the projector-camera pair, since the displacements Δq_{h1} and Δq_{h2} are represented in normalized coordinates. For real measurements, displacements in the units of projector pixels Δq_{hp1} and Δq_{hp2} are observed instead. The ratio of these values can be converted by

$$\frac{\Delta q_{h1}}{\Delta q_{h2}} = \frac{f_2}{f_1}\frac{\Delta q_{hp1}}{\Delta q_{hp2}}, \qquad (14)$$

where f_1 and f_2 are the focal lengths of the projectors.

4. Experimental systems

4.1. System configurations

As experimental systems, we have constructed a two-projector system and single-projector system. For two-projector system, the projectors and the camera are positioned so that the camera can observe overlapping patterns projected by the projectors. The two projectors project uniformly-spaced parallel lines and because discrimination of the two patterns is required for reconstruction, different colors are used. The setup with two projectors and a camera is shown in Fig. 7(b). For single-projector system, one projector projects two non-uniformly-spaced parallel line sets with different color. The setup with single projector and a camera is shown in Fig. 7(a).

For both systems, the projected patterns are static and do not change; thus, synchronization is not required. The camera and projector(s) are assumed to be calibrated (*i.e.*, the intrinsic parameters of the devices and their relative positions and orientations are known). Depths are estimated by the light flow analysis as already described.

4.2. Light flow estimation with two projectors

In the analysis of section 3.2, pattern motion on the projectors' image planes (Δq_{h1} and Δq_{h2}) were used. To observe these values, we propose to use a uniformly spaced line pattern with constant intervals as "scale" for measuring the pattern motion. In this method, the length of blur is retrieved from the observed image as the width of the band of blurred lines. Then, the ratio of the blur on the projector plane can be calculated by dividing the length by the line intervals on the same image. Since the line intervals are constant on the projector's image plane, the ratio becomes a pattern motion on the projector's image plane.

The blur of the parallel lines are observed only for each lines, thus, the resolution of this approach is relatively sparse. However, it has a great potential to reconstruct extremely fast motion using only a single image.

The method is as follows: The projected pattern is a set of parallel vertical lines. The captured lines projected onto the object move with the motion of the object. The exposure time of the camera is assumed to be set so that the apparent

motion of the vertical line of the image is observed as motion blurs of the lines.

Let the width of the motion blur of the line defined as Δr. Δr cannot be used directly as Δq_h in Eqn. 9, because Δq_h is a displacement on the projected pattern image, whereas Δr is a displacement on the captured camera image. If vertical lines at uniform intervals are projected, Δr can be normalized by the apparent intervals on the camera image and can be used as Δq_h. If the apparent motion blur of the line on a local patch in the camera image is Δr and the interval between the lines on the same patch is B, then Δq_h can be approximated as follows:

$$\Delta q_h \approx \Delta r / B. \qquad (15)$$

In the current implementation, the blur is detected just by simple adaptive binarization. Then, Δr and B of Eqn. 15 are estimated with sub-pixel accuracy by localizing crossing positions of the profile with the threshold levels of the adaptive binarization (the positions marked by red circles in Fig. 5(e)).

By projecting vertical lines from two projectors in different colors (*e.g.*, red and blue) and by estimating Δq_{h1} and Δq_{h2} from the different color channels, depth estimation from motion blur becomes possible. Fig. 5 shows the pattern images in (a, b), projected line patterns with motion blurring in (c), an apparent line interval (*i.e.*, B) and a size of blur (*i.e.*, Δr) in (d), and the *light flow* estimation by interpolating $\Delta r / B$.

If the target surfaces have object boundaries, the line patterns cannot be detected outside the boundaries. Thus, blurred patterns are disconnected and values of Δr or B become abnormal at those points. Since the assumption of smooth surfaces are not fulfilled there, we remove these regions. In the current implementation, we specify the upper and lower limits for the intervals B and label pixels as outliers where B exceeds this range. The outlier points are removed from the 3D reconstruction. The boundary points are removed from the reconstruction process using this technique.

The line directions of the pattern are not required to be perpendicular to scan lines. This is because although blur widths Δr and line intervals B along scan lines are affected by the apparent line directions, the ratio $\Delta r / B$ are not affected by the directions.

In this method, the spatial resolution of the *light flows* is as coarse as the apparent line intervals, but fast-moving objects that are only observable as motion blur can be measured.

About the precision of the method, we can conduct a coarse analysis on depth precision based on Fig. 4(a). Let Δr and B be 10 and 30 pixels, which is a typical setup of the experiments shown later. Also, let precisions of those values be $1/4$ subpixels. Then, the errors of $\log(\Delta r)/B$ is ap-
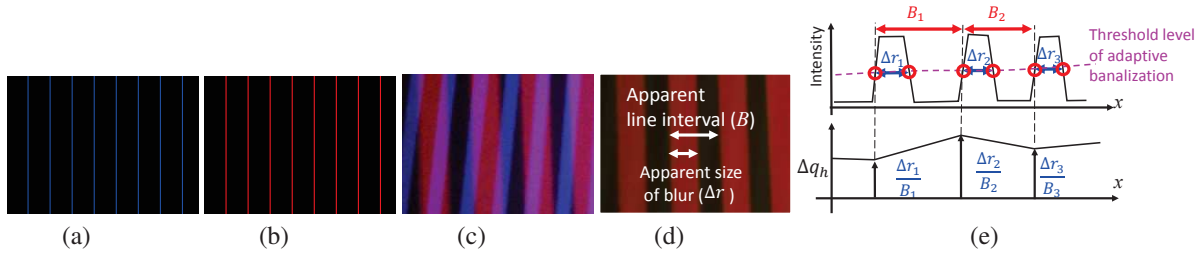
Figure 5. *Light flow* estimation from uniform line patterns: (a,b) line pattens, (c) projected line patterns with motion blurring, (d) an apparent line interval and a size of blur, (e) flow estimation using line intervals and the blur sizes.
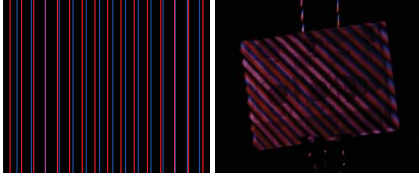

Figure 6. Left: Modulated nonuniform parallel line patterns and Right: captured image with strong motion blur.
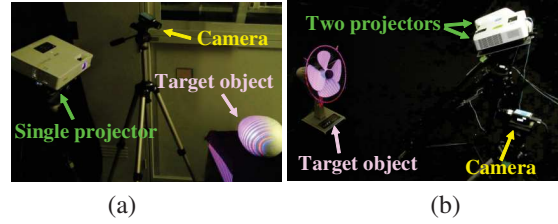

Figure 7. Two examples of the system configuration. (a) single projector projects non-uniformly spaced lines with two colors and (b) two projectors project uniformly spaced lines to make overlapping pattern on the object.

proximately 0.049 (we take $\log$ because h^{-1} takes log values). In Fig. reffig:depthfigratiofunc(a), this is about 0.02-0.03m depth errors at 0.5m distance, and about 0.06m depth errors at 1.0m distance.

4.3. Light flow estimation with a single projector

We have explained depth estimation method with two patterns projected from two projectors. Now, we consider the special case of two patterns, where the two patterns are projected from a single projector. In this case, depth estimation becomes impossible because of degenerate conditions. Specifically, the function $h(z)$ becomes constant functions for all the camera pixels, thus, h^{-1} cannot be defined.

Even in this case, by applying different modulation for these patterns, depth estimation becomes possible. For example, we modulate the pattern intervals so that the intervals becomes wide by two times at the right side with respect to the left side (Fig. 6). This modulation is achieved with the mapping of the horizontal coordinate p_0 by $p_0^{new} = 2^{p_0} - 1$, where range of p_0 is $[0, 1]$. On the contrary, for the other pattern, intervals are modulated by the mapping reversed for the left and right side: $p_1^{new} = 2 - 2^{1-p_1}$. With this setup, in calculation of $h(z)$, the inverse mappings $p_0 = \log_2(p_0^{new} + 1)$, $p_1 = 1 - \log_2(2 - p_1^{new})$, are applied. Then, non-constant $h(z)$ can be obtained. With this technique, we can estimate depth using two patterns projected from a single projector with the same algorithm as the two-projectors case.

5. Experiments

5.1. Evaluation with planar board

The first experiment was conducted with two-projector system (Fig. 7(b)). The camera resolution was 1600×1200
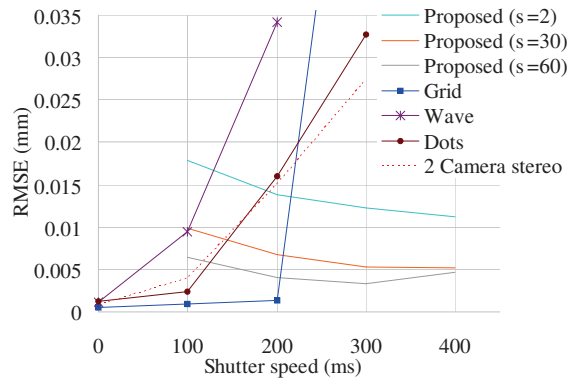

Figure 8. Accuracy evaluation for captured plane with line patterns. Vertical axis are RMSEs and horizontal axis are shutter speeds. Since *light flows* are approximated by piecewise linear interpolation as Fig. 5(e), they are smoothed with Gaussian filters with $\sigma = 2, 30, 60$. The size of the board is $600 \times 400mm$.

pixels, projector resolution was 1280×800 pixels, and the baselines between the camera and the two projectors were both approximately 400mm.

For evaluation, we attached a target board onto the motorized stage, and captured it with the camera while the board was moving under different conditions, such as different shutter speed and different projection pattern. The RMSE was calculated after plane fitting to the estimated depth. Results are shown in Fig.8.

Since the spatial accuracy of the flow estimation with the line pattern is an approximation and is affected by the apparent sizes of the blur bands and the apparent line intervals, we smoothed the flow estimation with Gaussian kernels with σ

values of 2, 30, and 60 (the apparent line intervals were around 40 pixels in the experiment.) In the results, we can clearly observe that the increase in the shutter speed improves the RMSE. This is because the longer shutter speed makes the longer bands for blur of lines, which helps to increase the accuracy on estimating the ratio of *light flows*. Although we also captured data with the shutter speed of 500 ms, the bands began to overlap each other and the depth estimation failed. For the shutter speed of 400 ms, filtering with $\sigma = 60$ did not improve the result than the case of $\sigma = 30$. This can be because the spatial resolution in the case of 400 ms shutter speed was worse than the case of 300 ms.

Fig. 10 shows examples of the actual captured image and reconstructed depth maps with our method. As can be seen, depth are correctly reconstructed just from relative information of flows without coded positions. We also applied the line based technique to textured materials, such as newspaper and RMSE was about 30mm for the shutter speed of 300 ms. This is because our image processing is based on simple thresholding and easily affected by textures; solution is our important future work.

We also recover the moving planar board with four other methods for comparison, such as grid based reconstruction [21], wave based reconstruction [14], random dot based reconstruction, which is equivalent to Kinect [11], and stereo-based reconstruction using two cameras. As clearly shown in the graph, longer shutter speed (equivalent to fast motion) drastically decreases the accuracy, and no technique can recover meaningful shapes when shutter speed exceeds 300 ms.

The above performance decrease of the conventional methods is caused by motion blurred patterns. Fig. 9 show examples of the projected patterns, the real captured images, and the line detection results at the shutter speed of 300 ms. As can be seen, the line detection results become unstable with severely blurred pattern, resulting in failure reconstruction. As for the two camera stereo case, since light flow is view-dependent, which is apparently shown in (i) and (j), it fails to find correspondences for reconstruction (the sizes and directions of the blur are different).

5.2. Arbitrary shape reconstruction

We applied our method to object with curved surfaces and fast motion as shown in Figs. 11, 12, and 13.

The first target was a rotating fan using the uniformly spaced line pattern (Fig. 11). As can be seen, the blades are blurred even with the fastest shutter speed (1ms) as shown in (a), whereas a projected pattern made a clear band of blur for each line as shown in (b,c). Using the detected bands and their width, *light flows* are estimated as (d,e), and depths are correctly reconstructed as shown in (f) using the calculated ratio of the *light flows*. Note that, in the flow val-

ues shown in (d,e) at the "front" parts of the blades (marked by red and black ellipses in (a) and (d)) are larger than the other parts. Because the fan blades are curved so that the air can be accelerated mostly by the front parts of the blade, the changing velocity at these parts are the highest. In this example, boundaries of the fan blades are removed from the reconstruction as outliers as described in Sec. 4.2.

The second target was a thrown ball using the same setup. Similar to the blades, ball has strong blur on captured image (Fig. 12(a)), whereas the projected pattern made a clear band of blur for each line ((b,c)). Using the detected bands, depths are correctly reconstructed, as shown in overlapped three frames of depth maps in (d). From (d), the motion of the ball moving from the camera can be clearly observed.

Finally, two balls were thrown and shapes are reconstructed by a single projector setup as shown in Fig. 7(a). Fig. 13(a) shows that projected lines were strongly blurred on the target objects, however two color bands are robustly extracted by our algorithm as shown in (b) and (c). Then, shapes are correctly reconstructed as shown in (d). In the next frame, two balls are reconstructed at further distance as shown in (e).

6. Conclusion

In this paper, we have proposed techniques to reconstruct the shape of fast moving objects which are captured with motion blur of a projected pattern using the ratio of *light flows* of two projections. With the proposed technique, the distances can be directly calculated from local displacement information. Encoding and decoding of global positional information from the pattern is not required, which is usually a difficult task. We have presented two types of setups, *i.e.*, single and two projectors configurations, to efficiently estimate the *light flow* and the depth of an object. Our experimental results demonstrate that depth is actually recovered by the ratio of the *light flows* on several moving objects, such as planar board, rotating fan and thrown ball.

There are, of course, limitations in the proposed method, such as the spatial resolution becoming as coarse as the line intervals. However, this is the first technique which achieves depth estimation only from a flow of projected pattern, which is observed as a blur with the best of our knowledge. In addition, if we use a precision device like laser projector with phase based analysis which can realize high precision pixel detection, it can improve the accuracies and spatial resolution in theory, which will be our future research.
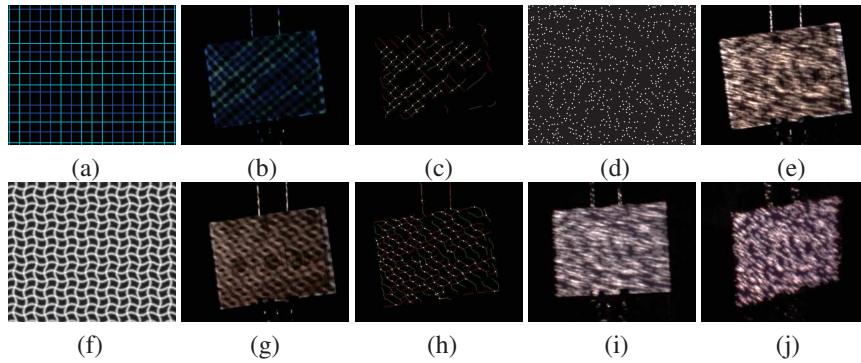
Figure 9. (a) grid pattern, (b) captured image of grid pattern, (c) line detection result of grid pattern, (d) random dot pattern, (e) captured image of random dots, (f) wave pattern, (g) captured image of wave pattern, (h) line detection result of wave pattern, (i) captured image of random dots from the first camera and (j) from the second camera. All the images are captured with shutter speed 300 ms.
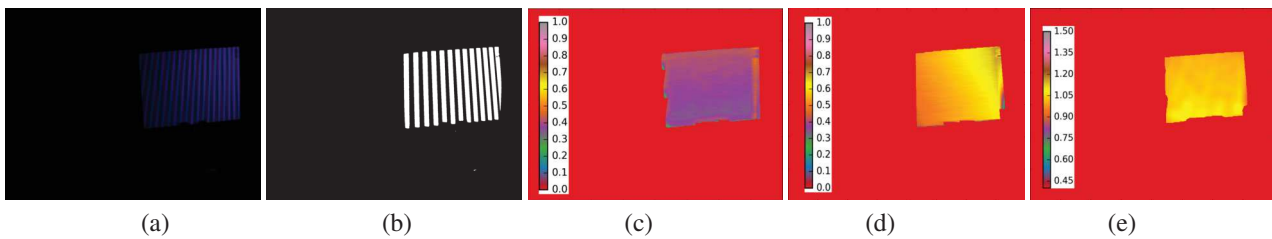

Figure 10. Capturing plane with line pattern for evaluation: (a) captured image, (b) blurred bands (c,d) (c,d) flow estimation of Δq_{h1} and Δq_{h2}, and (e) estimated depth.
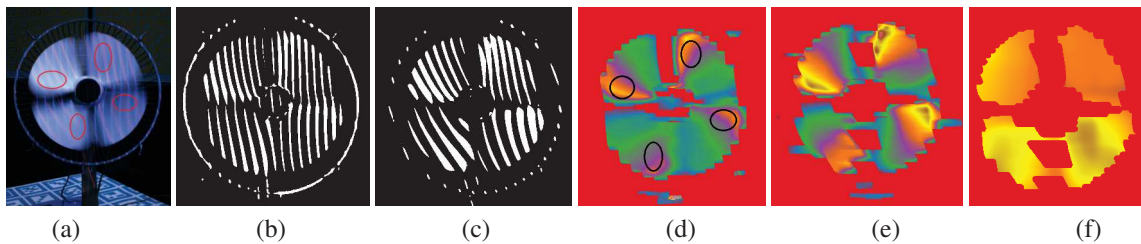

Figure 11. Depth estimation of a rotating fan (uniformly spaced line pattern): (a) captured image, (b,c) blurred bands (d,e) flow displacements for two projectors, and (f) estimated depth. Color mapping of (d,e,f) is the same as Fig. 10(c,d,e).
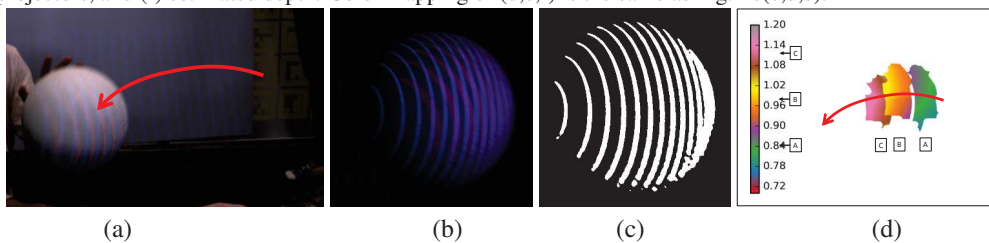

Figure 12. Depth estimation of a thrown ball (uniformly spaced line pattern):(a) the ball and the trajectory of the throw, (b) the captured image, (c) blurred bands, (d) the estimated depth maps for 3 consequent frames overlapped. Note that, since the image of (a) is blurred, normal structured light methods would fail.
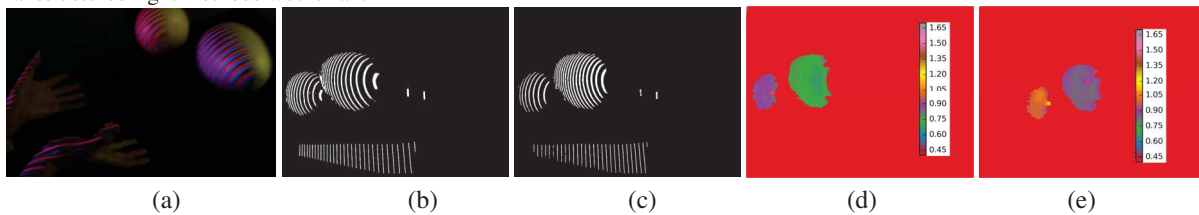

Figure 13. Depth estimation of two ball throwing (non-uniformly spaced line pattern):(a) capturing scene of the throwing, (b) extracted red band, (c) extracted blue band, (d) the estimated depth maps for frame 1 and (e) the estimated depth maps for frame 2.

References

[1] R. Furukawa, R. Sagawa, H. Kawasaki, K. Sakashita, Y. Yagi, and N. Asada. One-shot entire shape acquisition method using multiple projectors and cameras. In *4th Pacific-Rim Symposium on Image and Video Technology*, pages 107–114. IEEE Computer Society, 2010. 2

[2] B. Han, D. Post, and P. Ifju. Moiré interferometry for engineering mechanics: current practices and future developments. *The Journal of Strain Analysis for Engineering Design*, 36(1):101–117, 2001. 2

[3] T. Higo, Y. Matsushita, and K. Ikeuchi. Consensus photometric stereo. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1157–1164, June 2010. 2

[4] B. K. P. Horn. Shape from shading. chapter Obtaining Shape from Shading Information, pages 123–171. MIT Press, Cambridge, MA, USA, 1989. 2

[5] K. Ikeuchi. Determining surface orientations of specular surfaces by using the photometric stereo method. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (6):661–669, 1981. 2

[6] H. Kawasaki, R. Furukawa, , R. Sagawa, and Y. Yagi. Dynamic scene shape reconstruction using a single structured light pattern. In *CVPR*, pages 1–8, June 23-28 2008. 1, 2

[7] H. Kawasaki, S. Ono, Y. Horita, Y. Shiba, R. Furukawa, and S. Hiura. Active one-shot scan for wide depth range using a light field projector based on coded aperture. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 3568–3576, 2015. 2

[8] T. P. Koninckx and L. Van Gool. Real-time range acquisition by adaptive structured light. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 28(3):432–445, 2006. 2

[9] L. Maier-Hein, A. Groch, A. Bartoli, S. Bodenstedt, G. Boissonnat, P.-L. Chang, N. Clancy, D. Elson, S. Haase, E. Heim, J. Hornegger, P. Jannin, H. Kenngott, T. Kilgus, B. Muller-Stich, D. Oladokun, S. Rohl, T. dos Santos, H.-P. Schlemmer, A. Seitel, S. Speidel, M. Wagner, and D. Stoyanov. Comparative validation of single-shot optical techniques for laparoscopic 3-d surface reconstruction. *IEEE Transactions on Medical Imaging*, 33(10):1913–1930, October 2014. 2

[10] Mesa Imaging AG. SwissRanger SR-4000, 2011. http://www.swissranger .ch/index.php. 1, 2

[11] Microsoft. Xbox 360 Kinect, 2010. http://www.xbox.com/en-US/kinect. 1, 2, 7

[12] M. O'Toole, S. Achar, S. G. Narasimhan, and K. N. Kutulakos. Homogeneous codes for energy-efficient illumination and imaging. *ACM Trans. Graph.*, 34(4):35:1–35:13, July 2015. 2

[13] M. Proesmans and L. Van Gool. One-shot 3d-shape and texture acquisition of facial data. In *Audio-and Video-based Biometric Person Authentication*, pages 411–418. Springer, 1997. 2

[14] R. Sagawa, H. Kawasaki, R. Furukawa, and S. Kiyota. Dense one-shot 3D reconstruction by detecting continuous regions with parallel line projection. In *Proc. 13th IEEE International Conference on Conputer Vison(ICCV 2011)*, pages 1911–1918, 2011. 7

[15] R. Sagawa, Y. Ota, Y. Yagi, R. Furukawa, N. Asada, and H. Kawasaki. Dense 3D reconstruction method using a single pattern for fast moving object. In *ICCV*, 2009. 2

[16] J. Salvi, J. Pages, and J. Batlle. Pattern codification strategies in structured light systems. *Pattern Recognition*, 37(4):827–849, 4 2004. 2

[17] K. Sato and S. Inokuchi. Three-dimensional surface measurement by space encoding range imaging. *Journal of Robotic Systems*, 2:27–39, 1985. 2

[18] Y. Taguchi, A. Agrawal, and O. Tuzel. Motion-aware structured light using spatio-temporal decodable patterns. *Computer Vision–ECCV 2012*, pages 832–845, 2012. 2

[19] H. Takasaki. Moiré topography. *Appl. Opt.*, 12(4):845–850, Apr 1973. 2

[20] A. Ulusoy, F. Calakli, and G. Taubin. One-shot scanning using de bruijn spaced grids. In *Proc. The 2009 IEEE International Workshop on 3-D Digital Imaging and Modeling*, 2009. 1, 2

[21] A. O. Ulusoy, F. Calakli, and G. Taubin. One-shot scanning using de bruijn spaced grids. In *The 7th IEEE Conf. 3DIM*, pages 1786–1792, 2009. 7

[22] J. Wang, A. C. Sankaranarayanan, M. Gupta, and S. G. Narasimhan. Dual structured light 3d using a 1d sensor. In *Computer Vision - ECCV 2016 - 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part VI*, pages 383–398, 2016. 2