PAPER *Special Section on Information Theory and Its Applications*

Optimal Grid Pattern for Automated Camera Calibration Using Cross Ratio*

Chikara MATSUNAGA†, *Nonmember*, **Yasushi KANAZAWA**††, *and* **Kenichi KANATANI**†††, *Members*

SUMMARY With a view to virtual studio applications, we design an optimal grid pattern such that the observed image of a small portion of it can be matched to its corresponding position in the pattern easily. The grid shape is so determined that the *cross ratio* of adjacent intervals is different everywhere. The cross ratios are generated by an optimal Markov process that maximizes the accuracy of matching. We test our camera calibration system using the resulting grid pattern in a realistic setting and show that the performance is greatly improved by applying techniques derived from the designed properties of the pattern.

key words: cross ratio, Markov process, error analysis, reliability evaluation, bootstrap, virtual studio

1. Introduction

Visually presenting 3-D shapes of real objects is one of the main goals of many Internet applications such as network cataloging and virtual museums, and generating virtual images by embedding graphics objects in real scenes or real objects in graphics scenes, known as *mixed reality* or *augmented reality*, is one of the central themes of image and media applications today. For such applications, we need to know the 3-D position of the camera that we use and its internal parameters. Thus, *camera calibration* is a first step in all vision and media applications.

The standard method is *pre-calibration*: we determine the internal parameters of the camera by taking images of marker patterns in a controlled environment in advance [12]. Recently, techniques for computing the 3-D position and the internal parameters of the camera at the same time by observing multiple images of a natural scene have been studied intensively [9]. Such a technique, known as *self-calibration*, may be very useful in unknown environments such as outdoors. However, it cannot be applied if the scene changes as the camera moves.

Manuscript received January 21, 2000.
Manuscript revised June 19, 2000.

†The author is with the Broadcast Division, FOR-A Co. Ltd., 2-3-3 Ohsaku, Sakura, Chiba 285-0802 Japan.
††The author is with the Department of Knowledge-based Information Engineering, Toyohashi University of Technology, Toyohashi, Aichi 441-8580 Japan.
†††The author is with the Department of Computer Science, Gunma University, Kiryu, Gunma 376-8515 Japan.

*This work was in part supported by the Ministry of Education, Science, Sports and Culture, Japan under the Grant-in-Aid for Scientific Research C(2) (No. 11680377).

In this paper, we focus on *virtual studio* applications [2], [11]: we take images of moving objects such as persons and superimpose them in a graphics background in real time by computing the 3-D positions and the internal parameters of a moving camera. Since the camera parameters, such as zooming, as well as the scene itself may change in the course of the camera motion, we cannot pre-calibrate or self-calibrate the camera.

This difficulty can be avoided by placing an easily distinguishable planar pattern in the scene (Fig. 1): we detect the image of the pattern in each frame, compute the 3-D position and the internal parameters of the camera from it [3], and remove the image of the pattern by image segmentation. We call this process *simultaneous calibration*.

Unlike pre-calibration, for which manual intervention is allowed, simultaneous calibration must be automated completely. In particular, we must match a given image of a portion of the pattern to its corresponding position in the pattern. In this paper, we design a grid pattern in such a way that the *cross ratio* of adjacent intervals is different everywhere. Since the cross ratio is invariant to perspective projection [5], [8], observed grid points can be matched to their absolute positions easily by comparing the cross ratios.

We *optimize* the grid shape so that the accuracy of matching is maximized in the presence of noise. Introducing a statistic model of image noise, we generate the grid intervals by an optimal Markov process that maximizes the accuracy of matching. Since the pattern is theoretically designed by statistical analysis, we can derive many effective recognition techniques from the analysis by which the pattern is designed. We test our

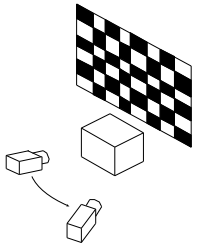


Fig. 1 Simultaneous calibration of a moving camera.

camera calibration system using the designed pattern in a realistic setting and show that the performance is greatly improved by applying such techniques.

2. Cross Ratio

The cross ratio can be defined in many different ways [3], [10]. Here, we define the cross ratio τ_i of four numbers $\{x_{i-1}, x_i, x_{i+1}, x_{i+2}\}$ by

$$\tau_i = \frac{x_i - x_{i-1}}{x_{i+1} - x_{i-1}} \Big/ \frac{x_{i+2} - x_i}{x_{i+2} - x_{i+1}} = \frac{1}{(1 + l_i/l_{i+1})(1 + l_i/l_{i-1})}, \quad (1)$$

where we have defined the ith interval width l_i by

$$l_i = x_{i+1} - x_i. \quad (2)$$

Our task is to generate a sequence $\{x_i\}$ in such a way that we can easily find the number i for which $\{x_{i-1}, x_i, x_{i+1}, x_{i+2}\}$ have a specified cross ratio τ_i. This is achieved by determining the sequence $\{x_i\}$ so that each of the cross ratios $\{\tau_i\}$ is as different from others as possible.

If the cross ratios $\{\tau_i\}$ are given, the sequence $\{x_i\}$ is determined as follows. Suppose we have already generated $\{x_0, ..., x_i, x_{i+1}\}$. If τ_i is given, the next number x_{i+2} is determined from Eq. (1) in the form

$$x_{i+2} = x_{i+1} + \frac{\tau_i(1 + \gamma_i)}{1 - \tau_i(1 + \gamma_i)} l_i, \quad (3)$$

where we have defined the ith *adjacency ratio* γ_i by

$$\gamma_i = \frac{x_{i+1} - x_i}{x_i - x_{i-1}}. \quad (4)$$

However, the cross ratios $\{\tau_i\}$ cannot be given arbitrarily, because the resulting sequence $\{x_i\}$ must have certain properties.

First of all, $\{x_i\}$ must be an increasing sequence. Also, it should be as homogeneous over the entire sequence as possible. Furthermore, intervals of very small separation cannot be discerned in the camera image, so the ratio of the minimum width $l_{\min}$ to the average interval width l_0 must be specified. Since the absolute scale of the pattern does not affect our analysis, we can normalize l_0 to be 1 without losing generality.

It is very difficult to find a sequence $\{\tau_i\}$ such that $\{x_i\}$ is a homogeneous increasing sequence with the average interval being 1 and the minimum width being $l_{\min}$ as specified. Our strategy here is to generate the sequence $\{\tau_i\}$ *stochastically*, according to a probability distribution defined in such a way that the resulting sequence $\{x_i\}$ has the desired properties. Moreover, as we will show shortly, we can *optimize* the probability distribution so that the matching performance is maximized by analyzing the statistical properties of image noise.

3. Error Analysis

Suppose x_{i-1}, x_i, x_{i+1}, and x_{i+2} have errors Δx_{i-1}, Δx_i, Δx_{i+1}, and Δx_{i+2}, respectively. The errors in the intervals l_{i-1}, l_i, and l_{i+1} are

$$\Delta l_{i-1} = \Delta x_i - \Delta x_{i-1}, \quad \Delta l_i = \Delta x_{i+1} - \Delta x_i, \quad \Delta l_{i+1} = \Delta x_{i+2} - \Delta x_{i+1}. \quad (5)$$

It follows from Eq. (1) that the cross ratio τ_i has the following error to a first approximation:

$$\begin{aligned}\Delta\tau_i &= \frac{\partial\tau_i}{\partial l_{i-1}}\Delta l_{i-1} + \frac{\partial\tau_i}{\partial l_i}\Delta l_i + \frac{\partial\tau_i}{\partial l_{i+1}}\Delta l_{i+1} \\ &= -\frac{\partial\tau_i}{\partial l_{i-1}}\Delta x_{i-1} + \left(\frac{\partial\tau_i}{\partial l_{i-1}} - \frac{\partial\tau_i}{\partial l_i}\right)\Delta x_i \\ &\quad + \left(\frac{\partial\tau_i}{\partial l_i} - \frac{\partial\tau_i}{\partial l_{i+1}}\right)\Delta x_{i+1} + \frac{\partial\tau_i}{\partial l_{i+1}}\Delta x_{i+2}. \quad (6)\end{aligned}$$

If the noise is an independent Gaussian variable of mean 0 and standard deviation σ, the variance of $\Delta\tau_i$ is

$$V[\tau_i] = \sigma^2\left(\left(\frac{\partial\tau_i}{\partial l_{i-1}}\right)^2 + \left(\frac{\partial\tau_i}{\partial l_{i-1}} - \frac{\partial\tau_i}{\partial l_i}\right)^2 + \left(\frac{\partial\tau_i}{\partial l_i} - \frac{\partial\tau_i}{\partial l_{i+1}}\right)^2 + \left(\frac{\partial\tau_i}{\partial l_{i+1}}\right)^2\right), \quad (7)$$

where

$$\begin{aligned}\frac{\partial\tau_i}{\partial l_{i-1}} &= \frac{\tau_i l_i}{l_{i-1}(l_{i-1} + l_i)}, \\ \frac{\partial\tau_i}{\partial l_i} &= -\frac{\tau_i^2(l_{i-1}/l_i + 2 + l_{i+1}/l_i)}{(l_{i-1}/l_i)(l_{i+1}/l_i)}\frac{1}{l_i}, \\ \frac{\partial\tau_i}{\partial l_{i+1}} &= \frac{\tau_i}{l_{i+1}/l_i(1 + l_{i+1}/l_i)}\frac{1}{l_i}. \quad (8)\end{aligned}$$

Hence, the standard deviation $\sqrt{V[\tau_i]}$ of τ_i has the following expression:

$$\begin{aligned}\sqrt{V[\tau_i]} &= \frac{\sigma}{l_i} s_{\gamma_i}(\tau_i), \\ s_{\gamma_i}(\tau_i) &= \sqrt{A_i^2 + (A_i - B_i)^2 + (B_i - C_i)^2 + C_i^2}, \\ A_i &= \frac{\gamma_i^2\tau_i}{1 + \gamma_i}, \quad B_i = -\frac{\tau_i^2(1 + \gamma_i(2 + D_i))}{D_i} \\ C_i &= \frac{\tau_i}{D_i(1 + D_i)}, \quad D_i = \frac{(1 + \gamma_i)\tau_i}{1 - (1 + \gamma_i)\tau_i}. \quad (9)\end{aligned}$$

4. Optimal Conditional Probability

If l_i and γ_i are given, the standard deviation $\sqrt{V[\tau_i]}$ is a function of τ_i. It follows that the matching error is minimized if we generate the cross ratio τ densely in the domain over which the standard deviation is small and sparsely in the domain over which it is large. This means that we should define the probability density $p_{l,\gamma}(\tau)$ conditioned on l and γ to be inversely proportional to $s_\gamma(\tau)$, i.e.,

$$p_{l,\gamma}(\tau) = \frac{C_{l,\gamma}}{s_\gamma(\tau)}, \tag{10}$$

where $C_{l,\gamma}$ is the normalization constant. From the normalization condition $\int_{\tau_a}^{\tau_b} p_{l,\gamma}(\tau)d\tau = 1$, we obtain

$$C_{l,\gamma} = 1 \Big/ \int_{\tau_a}^{\tau_b} \frac{d\tau}{s_\gamma(\tau)} . \tag{11}$$

From Eq. (3), the condition that the expected length of $x_{i+2} - x_{i+1}$ be 1 is written as

$$\int_{\tau_{ai}}^{\tau_{bi}} \frac{\tau(1+\gamma_i)l_i}{1-\tau(1+\gamma_i)} p_{l_i,\gamma_i}(\tau)d\tau = 1. \tag{12}$$

Substituting Eqs. (10) and (11) into this, we obtain

$$\frac{1+1/l}{1+\gamma} \int_{\tau_a}^{\tau_b} \frac{(1+\gamma)\tau - 1/(1+l)}{1-(1+\gamma)\tau} \frac{d\tau}{s_\gamma(\tau)} = 0. \tag{13}$$

It follows that if we define

$$f_{l,\gamma}(x) = \int_{\tau_a}^{x} \frac{\tau - 1/(1+l)(1+\gamma)}{\tau - 1/(1+\gamma)} \frac{d\tau}{s_\gamma(\tau)}, \tag{14}$$

the upper bound τ_b of the domain $[\tau_a, \tau_b]$ is determined for a given lower bound τ_a as the solution of the equation $f_{l,\gamma}(x) = 0$. It has two solutions, one of which is τ_a itself. We denote the other solution by $\tau_{l,\gamma}(\tau_a)$.

5. Optimal Sequence

Since measurement error in one position affects two consecutive interval widths, three consecutive adjacency ratios, and four consecutive cross ratios, the desired distribution of the cross ratio τ depends on the interval width l and the adjacency ratio γ defined by the preceding positions. Hence, the resulting sequence $\{x_i\}$ is a *Markov process.*

Given x_0, ..., x_i, x_{i+1}, we generate the ith cross ratio τ_i according to the conditional probability (10) over the domain $[\tau_{ai}, \tau_{bi}]$ determined from Eqs. (1) and (13) in the form

$$\tau_{ai} = \frac{1}{(1+l_i/l_{\min})(1+\gamma_i)}, \quad \tau_{bi} = \tau_{l_i,\gamma_i}(\tau_{ai}), \tag{15}$$

where τ_{bi} is the solution of $f_{l_i,\gamma_i}(x) = 0$ such that $x \neq \tau_{ai}$. The integral in Eq. (14) can be numerically evaluated (say, by the trapezoidal rule), and the solution can be obtained by a numerical scheme (e.g., Newton iterations). Then, x_{i+2} is computed by Eq. (3), and we repeat this procedure. Let us call the resulting sequence $\{x_i\}$ the *optimal sequence* for short.

For comparison, we generate the ith interval l_i independently and uniformly over a domain $[l_{\min}, l_{\max}]$ centered at 1 (i.e., $l_{\max} = 2 - l_{\min}$) and define $x_{i+1} = x_i + l_i$. Let us call the resulting independent additive process $\{x_i\}$ the *random sequence* for short.

Figures 2(a) and (b) show an instance of the random sequence and the optimal sequence, respectively,

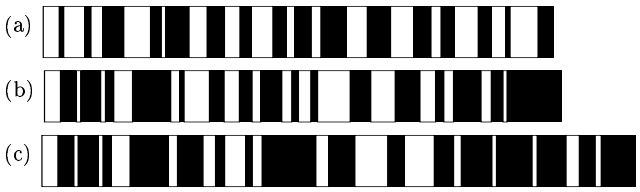


Fig. 2 (a) A random sequence. (b) An optimal sequence. (c) An optimal sequence with buffer zones.

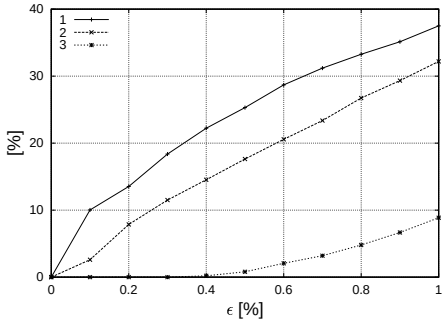


Fig. 3 The error ratio of matching. 1. Random sequence. 2. Optimal sequence. 3. Optimal sequence with buffer zones.

for $l_{\min} = 1/4$ (Fig. 2(c) is an instance of the optimal sequence with buffer zones to be explained later). We added independent Gaussian noise of mean 0 and standard deviation ϵ% of the average interval width to each position and computed the cross ratios of all four consecutive positions. We matched each position with the position that has the closest cross ratio. We repeated this 100 times using different noise each time and plotted the average error ratio in Fig. 3. We can observe that the optimal sequence can reduce the error to about 70% as compared with the random sequence.

6. Absolute Bounds on Distribution

We now investigate stationary properties of our Markov process. We first evaluate the absolute upper and lower bounds on the cross ratio, the adjacency ratio, and the interval width.

It is easily seen from Eq. (1) that the cross ratio τ_i takes its minimum when l_i/l_{i+1} and l_i/l_{i-1} both take their maximums. If we let $l_{\max}$ and $l_{\min}$ be the maximum and minimum of the interval width, respectively, the cross ratio τ_i takes its minimum when $l_{i-1} = l_{i+1} = l_{\min}$ and $l_i = l_{\max}$; it takes its maximum when $l_{i-1} = l_{i+1} = l_{\max}$ and $l_i = l_{\min}$. Hence, if $l_{\max}$ is given ($l_{\min}$ is a parameter set by the user), the maximum and the minimum of the cross ratio are, respectively,

$$\begin{aligned} \tau_{\min} &= \frac{1}{(1+l_{\max}/l_{\min})^2}, \\ \tau_{\max} &= \frac{1}{(1+l_{\min}/l_{\max})^2}. \end{aligned} \tag{16}$$

Since $\gamma_{\min} = l_{\min}/l_{\max}$ and $\gamma_{\max} = l_{\max}/l_{\min}$, the above equations are rewritten as follows:

$$\tau_{\min} = \frac{1}{(1+\gamma_{\max})^2}, \quad \tau_{\max} = \frac{1}{(1+\gamma_{\min})^2}. \tag{17}$$

Suppose $l_i = l_{\min}$ and $\gamma_i = \gamma_{\min}$. Equation (1) implies that τ_i ranges over

$$\frac{1}{2(1+\gamma_{\min})} \leqq \tau_i \leqq \tau_{\max} \tag{18}$$

as $l_{\min} \leqq l_{i+1} \leqq l_{\max}$. Hence,

$$\tau_{\max} = \tau_{l_{\min},\gamma_{\min}}\left(\frac{1}{2(1+\gamma_{\min})}\right). \tag{19}$$

Equation. (3) is rewritten as

$$l_{i+1} = \frac{\tau_i(1+\gamma_i)l_i}{1-\tau_i(1+\gamma_i)}. \tag{20}$$

Letting $l_i = l_{\min}$, $l_{i+1} = l_{\max}$, $\gamma_i = \gamma_{\min}$, and $\tau_i = \tau_{\max}$ in the expression for l_{i+1}/l_i, we can write the minimum adjacency ratio $\gamma_{\min} = l_{\min}/l_{\max}$ in the form

$$\gamma_{\min} = \frac{1-(1+\gamma_{\min})\tau_{\max}}{(1+\gamma_{\min})\tau_{\max}}. \tag{21}$$

If this expression is substituted into the right-hand side of Eq. (19), we obtain an equation to determine $\tau_{\max}$ by iterations: starting from an initial guess of $\gamma_{\min}$, say, $\gamma_{\min} = l_{\min}$, we compute $\tau_{\max}$ by Eq. (19), compute $\gamma_{\min}$ by Eq. (21), and repeat this process until the iterations converge.

The remaining upper and lower bounds are given as follows ($l_{\min}$ is a parameter set by the user):

$$l_{\max} = \frac{l_{\min}}{\gamma_{\min}}, \quad \gamma_{\max} = \frac{1}{\gamma_{\min}},$$
$$\tau_{\min} = \frac{1}{(1+1/\gamma_{\min})^2}. \tag{22}$$

7. Transition Probabilities

Now that we have determined the absolute upper and lower bounds $\tau_{\min}$, $\tau_{\max}$, $\gamma_{\min}$, $\gamma_{\max}$, $l_{\min}$, and $l_{\max}$, we regard the cross ratio as having a common domain $[\tau_{\min}, \tau_{\max}]$ by letting the density be 0 for $\tau_{\min} \leqq \tau_i \leqq \tau_{ai}$ and $\tau_{bi} \leqq \tau_i \leqq \tau_{\max}$. Similarly, we regard the adjacency ratio and the interval width as having common domains $[\gamma_{\min}, \gamma_{\max}]$ and $[l_{\min}, l_{\max}]$, respectively. We now derive their transition probabilities.

If l_i, γ_i, and τ_i are given, Eq. (20) gives the differential dl_{i+1} of l_{i+1} in the form

$$dl_{i+1} = \frac{1+\gamma_i}{(1-(1+\gamma_i)\tau_i)^2} l_i d\tau_i. \tag{23}$$

Let $F(l_{i+1}|l_i, \gamma_i)$ be the transition probability density of l_{i+1}. The probability that l_{i+1} is in the interval $[l_{i+1}, l_{i+1}+dl_{i+1}]$ is $F(l_{i+1}|l_i,\gamma_i)dl_{i+1}$, which should be equal to $p_{l_i,\gamma_i}(\tau_i)d\tau_i$. Since $\tau_i = 1/(1+l_i/l_{i+1})(1+\gamma_i)$, we obtain

$$F(l_{i+1}|l_i,\gamma_i) = \frac{(1-(1+\gamma_i)\tau_i)^2}{(1+\gamma_i)l_i} p_{l_i,\gamma_i}\left(\frac{1}{(1+l_i/l_{i+1})(1+\gamma_i)}\right). \tag{24}$$

Equation (20) is rewritten for the adjacency ratio $\gamma_i = l_i/l_{i-1}$ in the form

$$\gamma_{i+1} = \frac{\tau_i(1+\gamma_i)}{1-\tau_i(1+\gamma_i)}, \tag{25}$$

and hence

$$d\gamma_{i+1} = \frac{1+\gamma_i}{(1-(1+\gamma_i)\tau_i)^2} d\tau_i. \tag{26}$$

Since $\tau_i = 1/(1+1/\gamma_{i+1})(1+\gamma_i)$ from Eq. (1), the transition probability density of the cross ratio is given by

$$G(\gamma_{i+1}|l_i,\gamma_i) = \frac{(1-(1+\gamma_i)\tau_i)^2}{1+\gamma_i} p_{l_i,\gamma_i}\left(\frac{1}{(1+1/\gamma_{i+1})(1+\gamma_i)}\right). \tag{27}$$

8. Stationary Probabilities

Now, we derive stationary probabilities, for which we omit the indices i, $i+1$, etc. The stationary probability densities $f(l)$ and $g(\gamma)$ of the interval width l and the adjacency ratio γ are defined via the relationships

$$f(l') = \int_{l_{\min}}^{l_{\max}} \int_{\gamma_{\min}}^{\gamma_{\max}} F(l'|l,\gamma) f(l) g(\gamma) d\gamma dl,$$
$$g(\gamma') = \int_{l_{\min}}^{l_{\max}} \int_{\gamma_{\min}}^{\gamma_{\max}} G(\gamma'|l,\gamma) f(l) g(\gamma) d\gamma dl. \tag{28}$$

Figures 4 and 5 plot $f(l)$ and $g(\gamma)$ computed by simultaneously iterating Eqs. (28) for $l_{\min} = 1/4$. The integration was conducted by the trapezoidal rule. We also computed a sequence $\{x_i\}$ of length 1,000 according to the Markov process that we have defined. The resulting

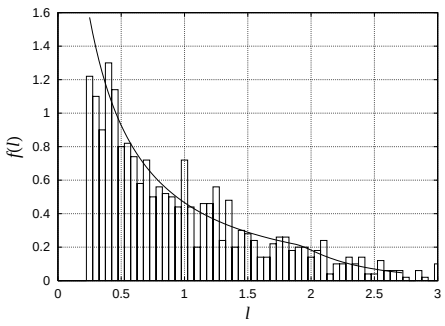


Fig. 4 Distribution of the interval width.

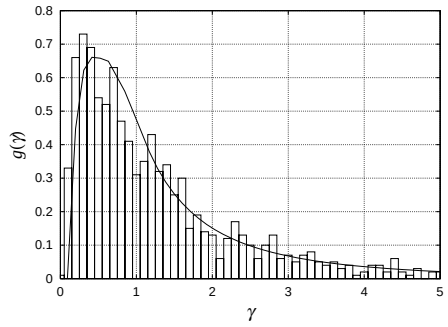


Fig. 5 Distribution of the adjacency ratio.

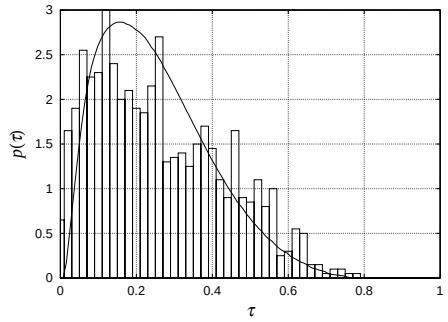


Fig. 6 Distribution of the cross ratio.

histograms are superimposed on the same scale in the figures. We can see that they agree with the theoretical prediction very well.

The stationary probability density of the cross ratio τ is defined via the relationship

$$p(\tau) = \int_{l_{\min}}^{l_{\max}} \int_{\gamma_{\min}}^{\gamma_{\max}} p_{l,\gamma}(\tau) f(l) g(\gamma) d\gamma dl. \qquad (29)$$

Figures 6 shows the computed density and the experimental histogram. They agree reasonably well.

9. Upper Bound on Matching Capability

Even if the distribution of the cross ratio is optimally defined, the resulting sequence may still contain values that are very close to each other as long as the generation is stochastic. This causes deterioration of the matching capability in the presence of image noise. So, we introduce a constraint that no two cross ratios be very close to each other.

If the standard deviation of the noise in $\{x_i\}$ is σ, the standard deviation $s_{l,\gamma}(\tau)$ of τ conditioned on l and γ is given by Eqs. (9):

$$s_{l,\gamma}(\tau) = \frac{\sigma}{l} s_\gamma(\tau). \qquad (30)$$

Each time we generate a cross ratio τ, we define buffer zones of width $s_{l,\gamma}(\tau)$ on both sides of τ and forbid subsequent cross ratios to occur in those zones.

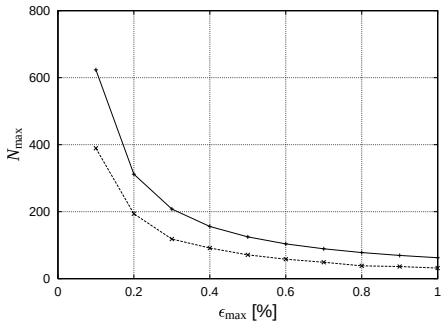


Fig. 7 The upper bound on the number of cross ratios.

If we continue this process indefinitely, the domain $[\tau_{\min}, \tau_{\max}]$ is ultimately covered with such prohibition zones, so no new values can be generated any longer. In other words, there is an upper bound on the number of the cross ratios that can be generated.

The expectation of Eq. (30) over all possible values of l and γ is given by

$$\begin{aligned} s(\tau) &= \int_{l_{\min}}^{l_{\max}} \int_{\gamma_{\min}}^{\gamma_{\max}} s_{l,\gamma} f(l) g(\gamma) d\gamma dl \\ &= \sigma \int_{l_{\min}}^{l_{\max}} \frac{f(l)}{l} dl \int_{\gamma_{\min}}^{\gamma_{\max}} s_\gamma(\tau) g(\gamma) d\gamma. \end{aligned} \qquad (31)$$

Consider an ideally packed state in which the cross ratios are distributed over $[\tau_{\min}, \tau_{\max}]$ at an interval $s(\tau)$. Since the number of cross ratios per unit length is $1/s(\tau)$, we obtain an upper bound on the number of available cross ratios in the from

$$N_{\max} = \frac{1}{\sigma} \int_{\tau_{\min}}^{\tau_{\max}} \frac{d\tau}{s(\tau)}. \qquad (32)$$

Of course this is an overestimation, since such an ideal state does not occur in practice.

The solid line in Fig. 7 plots the value $N_{\max}$ that can be obtained when the standard deviation σ of noise is $\epsilon_{\max}\%$ of the average interval width. The dashed line in Fig. 7 plots the number of actually generated cross ratios; we took the average of ten independent trials. We can see that only about 50% of the theoretical upper bound can be generated in practice.

Figure 2(c) shows an instance of such a sequence with buffer zones for $l_{\min} = 1/4$ and $\epsilon_{\max} = 1\%$. As we see from Fig. 3, the existence of buffer zones dramatically reduces the matching error.

10. Optimal Grid Pattern

Generating two sequences $\{x_i\}$ and $\{y_j\}$ independently, we can define a grid pattern with vertices $\{(x_i, y_j)\}$. Figure 8 shows one example for $l_{\min} = 1/3$ and $\epsilon_{\max} = 1\%$. It is painted like a checkerboard with dark and light blue colors for the convenience of chromakey applications.

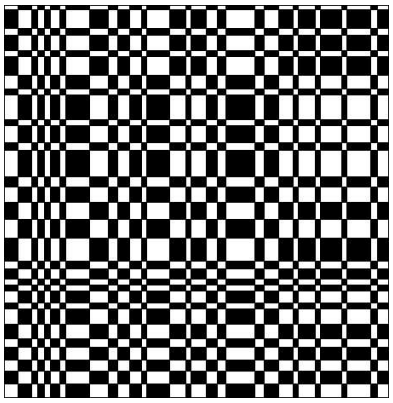


Fig. 8 An optimal grid pattern (checkerboard type).

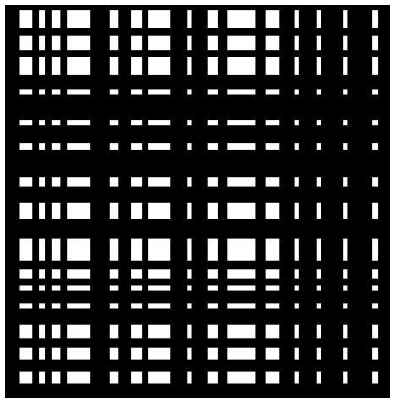


Fig. 9 An optimal grid pattern (framework type).

In order to compute the cross ratios in both directions, we need to observe at least a 3×3 block, from which the cross ratios are computed in four ways. Let τ_x and τ_y be the averages of the four values for the x and y directions, respectively. The absolute position of that block in the pattern is determined by finding integers i and j such that $|\tau_x - \tau_{x(i)}|$ and $|\tau_y - \tau_{y(j)}|$ are minimized, where $\{\tau_{x(i)}\}$ and $\{\tau_{y(j)}\}$ are the cross ratio sequences associated with $\{x_i\}$ and $\{y_j\}$, respectively.

This process does not consider the error behavior of the cross ratio. Since the standard deviation of the cross ratio can be evaluated by Eqs. (9), a statistically optimal method is the *maximum likelihood estimation*: we minimize the squared Mahalanobis distance

$$J(i,j) = \frac{l_{x(i)}^2 |\tau_x - \tau_{x(i)}|^2}{s_{\gamma_{x(i)}}(\tau_{x(i)})^2} + \frac{l_{y(j)}^2 |\tau_y - \tau_{y(j)}|^2}{s_{\gamma_{y(j)}}(\tau_{y(j)})^2}, \quad (33)$$

where $\{l_{x(i)}\}$ and $\{l_{y(j)}\}$ are the interval sequences defined from $\{x_i\}$ and $\{y_j\}$, and $\{\gamma_{x(i)}\}$ and $\{\gamma_{y(j)}\}$ are the similarly defined adjacency ratio sequences.

For a checkerboard pattern, a 3×3 block has two possibilities for its coloring. This information can be used to reduce the search space for minimizing Eq. (33). Another possibility for coloring the pattern is to alternate colors for neighboring rows and columns (Fig. 9). Let us call it a *framework pattern*. It has four possibilities for coloring a 3×3 block, reducing the search space to a half that for the checkerboard pattern.

11. Simulation

We added Gaussian random noise of mean 0 and standard deviation ϵ% of the average interval width to the coordinates of each grid point of the pattern shown in Figs. 8 and 9. The positions of all 3×3 blocks are computed from the observed cross ratios, and this process was repeated 100 times using different noise each time. Figure 10 plots the average error ratio for ϵ. Here, we compared the following four methods:

1. The simple method: $|\tau_x - \tau_{x(i)}|$ and $|\tau_y - \tau_{y(j)}|$ are minimized.
2. Maximum likelihood estimation: Eq. (33) is minimized.
3. Maximum likelihood estimation combined with checkerboard coloring information.
4. Maximum likelihood estimation combined with framework coloring information.

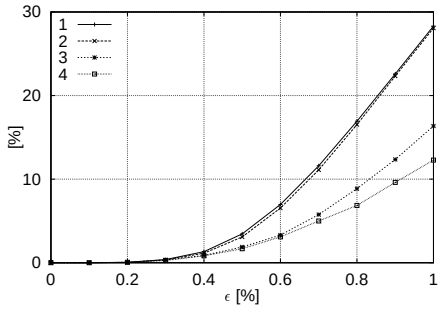


Fig. 10 The error ratio of matching. 1. The simple method. 2. Maximum likelihood estimation without using coloring information. 3. Maximum likelihood estimation using checkerboard coloring information. 4. Maximum likelihood estimation using framework coloring information.

We can see that maximum likelihood estimation slightly reduces the matching error as compared with the simple method. In contrast, the coloring information dramatically improves the accuracy and that the framework pattern is more effective than the checkerboard pattern.

12. Real Image Experiment

Figure 11 shows a real image of a part of the pattern in Fig. 9 viewed from an angle. First, we thresholded the gray levels to create a binary image. Then, we traced the boundaries of the quadrilateral regions, segmented each resulting digital loops into four linear segments,

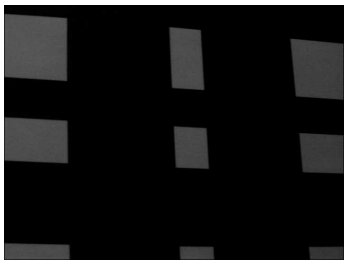


Fig. 11 A real image of a part of an optimal grid pattern.

and fitted lines to each by least squares. Using the original gray levels of the pixels near the fitted lines, we fitted new lines by least squares and determined the vertex positions by computing their intersections. Using a 3×3 block in the resulting grid pattern, we computed the cross ratios in both directions and determined the position of the block in Fig. 9. The correct position was obtained for all the four methods described in the preceding section.

Finally, we computed the 3-D position and the focal length of the camera by the method described in [6], [7]. The focal length is estimated to be 1718 pixels, and its reliability is evaluated to be ± 36.6 pixels (standard deviation). The reliability of the computed position and orientation of the camera turned out ± 3.4cm and $\pm 0.25°$, respectively.

13. Bootstrap

The above experiment is just for one instance of noise occurrence. In order to evaluate the reliability of this solution, we need to compare it with the values that would be obtained if the noise occurred differently. We can generate such potential noise by *bootstrap* [1].

The first step is to estimate the noise magnitude. To do this, we used the fact that the grid lines should ideally be *concurrent* (i.e., meeting at a single intersection) in each orientation. Applying the theory of statistical optimization of Kanatani [4], we optimally fitted a concurrent pattern to the observed vertices in such a way that the sum of the squares J of the deviations of the observed vertices from their corresponding grid points of the fitted concurrent pattern is minimized (we omit the details of the procedure).

According to the theory of statistical estimation (e.g. [4]), an unbiased estimator of the variance σ^2 of the image noise is obtained in the form

$$\hat{\sigma}^2 = \frac{\hat{J}}{32 - 12}, \tag{34}$$

where $\hat{J}$ is the residual (i.e., the minimum value of J for the optimally fitted concurrent pattern). The number 32 in the denominator is the number of data (the x and y coordinates of the 16 vertices), while the number 12 is the degree of freedom of the concurrent pattern.

The reason that a concurrent grid pattern has 12 degrees of freedom is as follows. Each line has two degrees of freedom, so the eight lines have sixteen degrees of freedom. However, the four lines in each orientation must have a common intersection, i.e., the third and fourth lines should pass through the intersection made by the first and the second lines, reducing two degrees of freedom in each orientation. Hence, the remaining degree of freedom is $16 - 2 - 2 = 12$. The validity of Eq. (34) was confirmed by numerical simulation.

According to Eq. (34), the standard deviation of the image noise in Fig. 11 was estimated to be 0.14 pixels. We generated random Gaussian noise of standard deviation 0.14 pixels and added it to each of the coordinates of the grid points of the optimally fitted concurrent pattern. We repeated the matching 10,000 times, using different noise each time. If the coloring information is not used, the ratio of correct matching was 96.31%; if the coloring information is used, it turned out 100.00%.

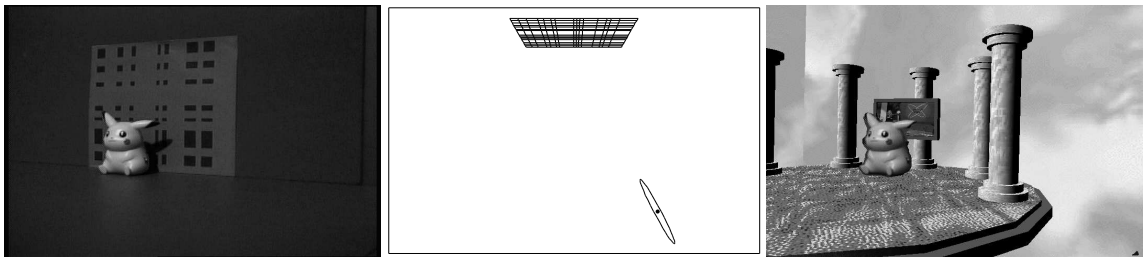
14. Virtual Studio Application

Figure 12(a) is a real image of a toy, behind which we placed our optimal grid pattern. After segmenting the toy image from the background by using a chromakey technique, we computed the 3-D position and focal length of the camera by observing an unoccluded portion of the grid pattern. The focal length is estimated to be 576 pixels. The standard deviations of the focal length, the translation, and the rotation are evaluated to be ± 38.3 pixels, ± 5.73cm, and $\pm 0.812°$, respectively.

Figure 12(b) is the top view of the estimated camera position and its uncertainty ellipsoid (three times the standard deviation in each orientation). Figure 12(c) is a composition of the toy image and a graphics scene generated by VRML.

15. Concluding Remarks

With a view to virtual studio applications, we have designed an optimal grid pattern such that an observed image of a small portion of it can be matched to its corresponding position in the pattern easily. The grid shape is so determined that the cross ratio of adjacent intervals is different everywhere. By statistical analysis of image noise, we have generated the cross ratios and grid intervals by an optimal Markov process that maximizes the accuracy of matching. We have tested our camera calibration system using the resulting grid pattern in a realistic setting and show that the performance is greatly improved by applying techniques derived from the designed properties of the pattern.



(a) (b) (c)

Fig. 12 (a) Original image. (b) Estimated camera position and its reliability. (c) A virtual scene generated from (a).

References

[1] B. Efron and R. J. Tibshirani, *An Introduction to Bootstrap*, Chapman-Hall, New York, 1993.
[2] S. Gibbs, C. Arapis, C. Breiteneder, V. Lalioti, S. Mostafawy and J. Speier, "Virtual studios: An overview," *IEEE Multimedia*, vol.5, no.1 pp.24–35, 1998.
[3] K. Kanatani, *Geometric Computation for Machine Vision*, Oxford University Press, Oxford, 1993.
[4] K. Kanatani, *Statistical Optimization for Geometric Computation: Theory and Practice*, Elsevier Science, Amsterdam, 1996.
[5] G. Lei, "Recognition of planar objects in 3-D space from single perspective views using cross ratio," *IEEE Trans. Robotics Automation*, vol.6, no.4 pp.432–437, 1990.
[6] C. Matsunaga and K. Kanatani, "Calibration of a moving camera using a planar pattern," *Proc. 5th Symp. Sensing via Image Information*, June 1999, Yokohama, Japan, pp. 255–260.
[7] C. Matsunaga and K. Kanatani, "Calibration of a moving camera using a planar pattern: Optimal computation, reliability evaluation and stabilization by model selection," *Proc.6th Euro. Conf. Computer Vision*, June 2000, Dublin, Ireland, vol.2, pp. 595–609.
[8] J. L. Mundy and A. Zisserman, *Geometric Invariance in Computer Vision*, MIT Press, Cambridge, 1992.
[9] M. Pollefeys, R. Koch and L. Van Gool, "Self-calibration and metric reconstruction in spite of varying and unknown internal camera parameters," *Int. J. Comput. Vision*, vol.32, no.1, pp.7–26, 1999.
[10] J. G. Semple and G. T. Kneebone, *Algebraic Projective Geometry*, Clarendon, Oxford, 1998.
[11] M. Tamir, "The Orad virtual set," *Int. Broadcast Eng.*, March, pp.16–18, 1996.
[12] R. Y. Tsai, "A versatile camera calibration technique for high-accuracy 3D machine vision methodology using off-the-shelf TV cameras and lenses," *IEEE J. Robotics Automation*, vol.3, no.4, pp.323–344, 1987.

Chikara Matsunaga received his B.E. and M.E. degrees in electronic engineering from Muroran Institute of Technology, Hokkaido, Japan, in 1988 and 1990, respectively. Currently, he is at the Broadcast Division of FOR-A Co. Ltd. He is developing digital video equipment for broadcasting. His research interests include image processing, computer vision and computer graphics.

Yasushi Kanazawa received his Ph.D. in information and computer science from Osaka University, Osaka, Japan, in 1997. He is currently Lecturer at the Department of Knowledge-Based Information Engineering at Toyohashi University of Technology. His research interests include image processing and computer vision.

Kenichi Kanatani received his Ph.D. in applied mathematics from the University of Tokyo in 1979. He is currently Professor of computer science at Gunma University. He is the author of *Group-Theoretical Methods in Image Understanding* (Springer, 1990), *Geometric Computation for Machine Vision* (Oxford, 1993) and *Statistical Optimization for Geometric Computation* (Elsevier, 1996).