Computer Vision and Image Understanding **79,** 324–329 (2000)
doi:10.1006/cviu.2000.0853, available online at http://www.idealibrary.com on IDEAL®

NOTE

Computing the Cost of Occlusion

Gabriel Fielding and Moshe Kam[1]

Data Fusion Laboratory, Department of Electrical and Computer Engineering, Drexel University, 3141 Chestnut St., Philadelphia, Pennsylvania 19104

Received November 6, 1998; accepted April 13, 2000

Recently, Cox *et al.* (1996, CVGIP: *Image Understanding* **63**, 542–567) presented a new dynamic programming-based stereo matching algorithm. The algorithm uses a parameter which represents the cost of occlusion. This cost is levied if the algorithm decides that two measurements, each from a different camera along corresponding epipolar lines, are not projections of the same point in space. The occlusion cost is dependent on the standard deviation of the (Gaussian) sensor noise, σ, and the probability of match detection, P_D. Under certain conditions such as low signal-to-noise ratio, the algorithm of Cox *et al.* will declare occlusions where they do not exist. We offer an alternative definition for the cost of occlusion, based on a decision-theoretic formulation for the matching process. This alternative improves the performance of the matching algorithm.

1. INTRODUCTION

In [1], Cox *et al.* present a fast matching algorithm for stereo images using dynamic programming. This novel algorithm has three main features: (i) it computes the maximum-likelihood estimate of the observed feature (e.g., intensity); (ii) the matching algorithm uses a dynamic programming formulation; and (iii) a new method is suggested to minimize changes between matchings on successive epipolar lines. The performance index in [1] includes a parameter which represents the *cost of occlusion*. This parameter determines when a feature or pixel should not be matched to any other candidate feature or pixel.[2] In [1], the cost of occlusion is a function of σ, the standard deviation of the sensor noise, and P_D, the probability of detection for matches.

[1] Corresponding author. Fax: (215) 895-1695; E-mail: kam@lazarus.ece.drexel.edu.

[2] In [1] and in this work the *features* to be matched can be pixels, lines, or other primitives. In [1] and in this paper, pixels are used as the features. We use the term "feature" exclusively hereafter.

1077-3142/00 $35.00

For the ranges of values for P_D and σ considered in [1], the resulting stereo matches are satisfactory. However, for higher values of σ (low signal-to-noise ratio (SNR)) and lower values of the probability of detection, the cost of occlusion can become negative, while the cost of matching is always nonnegative. Under these circumstances the algorithm of [1] always selects the *occlusion* option even when there are matchable regions in the images. One of our objectives here is to modify the algorithm in order to avoid this error.

We offer a new method to compute the cost of occlusion based on a decision theoretic argument. This modified cost is always nonnegative and can be extended to use windows of features [3]. Using synthetic images, we show that the modified cost allows successful use of [1] on noisy images.

2. PRELIMINARIES

We summarize the relevant notation in [1]. Each camera in the stereo configuration is identified by the subscript $s = \{1, 2\}$. $\mathbf{Z}_s$ is the vector of measurements from each camera, obtained along corresponding epipolar lines. Hence, $\mathbf{Z}_s = \{\mathbf{z}_{s,i_s}\}_{i_s=0}^{m_s}$ where m_s is the number of measurements from camera s. A candidate matching pair of features is denoted by Z_{i_1,i_2}. Each measurement vector $\mathbf{z}_{s,i_s}$ has a covariance matrix denoted by $\mathbf{S}_{s,i_s}$. We assume that there exists some ideal value of the measurement vector, $\mathbf{z}$, and that each epipolar measurement is normally distributed about its ideal value with zero mean and covariance $\mathbf{S}_s$. Throughout this work we have assumed that the variance of the measurement error for each pair of features is equal and given by σ^2. The objective function in [1, Eq. 6] is:

$$f_{i_1,i_2} = \delta_{i_1,i_2} \overbrace{\ln\left(\frac{P_D^2\phi}{(1-P_D)|(2\pi)^d S|^{\frac{1}{2}}}\right)}^{C_O,\ \text{occlusion cost}} + \left(1-\delta_{i_1,i_2}\right)\underbrace{\left(\frac{1}{4}\left(\mathbf{z}_{1,i_1}-\mathbf{z}_{2,i_2}\right)'\mathbf{S}^{-1}\left(\mathbf{z}_{1,i_1}-\mathbf{z}_{2,i_2}\right)\right)}_{C_M,\ \text{matching cost}}, \tag{1}$$

where δ_{i_1,i_2} is an indicator variable ($\delta_{i_1,i_2} = 1$ when a match is not declared, $\delta_{i_1,i_2} = 0$ otherwise), d is the dimension of the measurement vectors, and ϕ is the field of view of the camera (in radians). The objective function in (1) includes two terms: the first (which we denote C_O) describes the cost of occlusion and the second defines the cost of matching two candidate features. C_O is a function of the parameters σ and P_D. When comparing two measurements, each from a separate camera, we assume $d = 1$, $\mathbf{S} = \sigma^2$, and $\phi = \pi$. These substitutions reduce the matching cost in (1) to $((z_{1,i_1} - z_{2,i_2})/2\sigma)^2$.

The difficulty with the cost of occlusion in (1) is that for high values of σ and a certain range of P_D, C_O becomes *negative*. The cost of matching two features, C_M, is always nonnegative, so any matching strategy that minimizes the cost functional will *always declare occlusions* when C_O is negative. In fact, this happens whenever

$$\sigma > \frac{P_D^2}{1-P_D}\sqrt{\frac{\pi}{2}} \Rightarrow C_O < 0. \tag{2}$$

If an occlusion is taken to mean "no match present" or "uncertain match" then the occlusion cost should be a function of sensor noise. This must be the interpretation in [1] because the cost of occlusion in (1) *decreases* (makes occlusions more likely) for increasing

σ. We maintain that occlusions are fundamentally a property of the stereo camera geometry, not of the sensor noise. Using this interpretation, we seek a definition of the occlusion cost that does not depend on σ and examine the performance of the algorithm in [1] with this modified cost.

3. A DECISION THEORETIC OCCLUSION COST

Let z_{1,i_1} and z_{2,i_2} be measurements from corresponding epipolar lines in two cameras. If the measurements are matching (i.e., they are projections of the same point or surface in space), we assume that the difference between the measurements is an independent normal random variable $z_{1,i_1} - z_{2,i_2} \sim N(0, \sigma^2)$. We define $z_{i_1,i_2} = (z_{1,i_1} - z_{2,i_2})/2\sigma$ which is $N(0, \frac{1}{4})$. The random variable $z^2_{i_1,i_2}$ has a gamma distribution [3],

$$f_x(x) = \frac{c^b}{\Gamma(b)} x^{b-1} e^{-cx}, \tag{3}$$

where $b = \frac{1}{2}$ and $c = 2$. Knowing the underlying distribution (3) of the matching cost, the probability P_D of obtaining a matching cost between 0 and $\tilde{C}_O$ can be calculated through the integral

$$P_D = \int_0^{\tilde{C}_O} \sqrt{\frac{2}{\pi}} x^{-\frac{1}{2}} e^{-2x}\, dx = \gamma\left(\frac{1}{2}, 2\tilde{C}_O\right), \tag{4}$$

where γ is the incomplete gamma function defined by:

$$\gamma(b, x) = \frac{1}{\Gamma(b)} \int_0^x t^{b-1} e^{-t}\, dt.$$

Conversely, we may fix P_D in (4) and compute $\tilde{C}_O$ through tabulated percentile values of the cumulative gamma distribution or through numerical evaluation of the integral in (4). In this case, each matching instance may be considered a hypothesis test where we compare the test statistic (the matching cost) to the threshold $\tilde{C}_O$ which is determined by P_D.[3] The original objective function (1) is modified as follows:

$$\tilde{f}_{i_1,i_2} = \delta_{i_1,i_2} \tilde{C}_O + \left(1 - \delta_{i_1,i_2}\right) \left(\frac{\mathbf{z}_{1,i_1} - \mathbf{z}_{2,i_2}}{2\sigma}\right)^2. \tag{5}$$

We can compare our values of $\tilde{C}_O$ (4) with C_O (1).[4] Both are shown in Fig. 1 with $\sigma = 4$ as P_D varies from 0 to 1. C_O, which takes values from [1], takes values in $(-\infty, \infty)$ while $\tilde{C}_O \in [0, \infty)$.

4. EXAMPLES OF MATCHING WITH THE MODIFIED OCCLUSION COST

To illustrate the difference in algorithm performance for the two costs of occlusion, we tested the algorithm presented in [1] on the set of synthetic stereo images shown in Fig. 2. These images are 128 pixels square, 8-bit grayscale, and are corrupted by Gaussian noise. The true disparity map for the simulated images is shown in Fig. 2c where the black regions represent occlusions (i.e., where no matches exist).

[3] The significance of the hypothesis test would be $1 - P_D$.

[4] All examples in [1] use $P_D = 0.99$ and $\sigma = 4$ which yield $C_O = 4.12$. We obtain $\tilde{C}_O = 1.66$.

TABLE 1
Mean-Square Error for the Images in Fig. 3

	$\sigma = 4$	$\sigma = 8$	$\sigma = 10$
Original C_O	3.97	6.10	14.76
Modified $\tilde{C}_O$	4.62	4.55	4.60

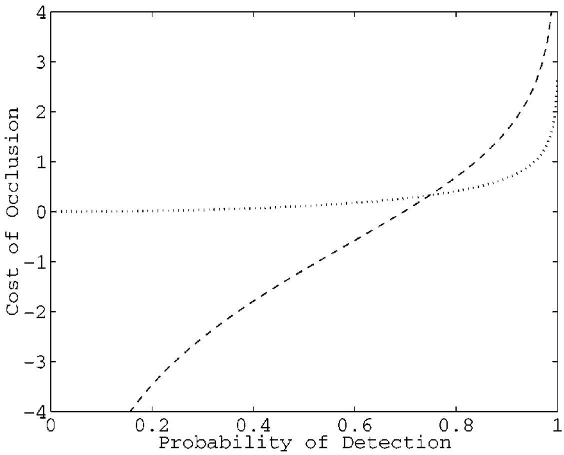

FIG. 1. The cost of occlusion from [1] in (1) (dashed) and the cost obtained numerically from (4) (dotted) with $\sigma = 4$.

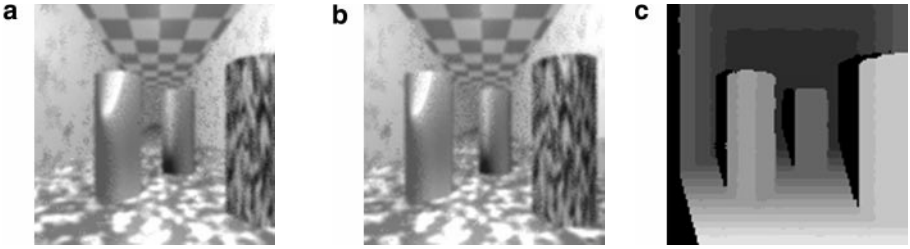

FIG. 2. A synthetic stereo pair of images: (a) left and (b) right and (c) the true disparity map with respect to the left image where larger disparities (closer objects) are represented by lighter colors and occluded regions are marked in black.

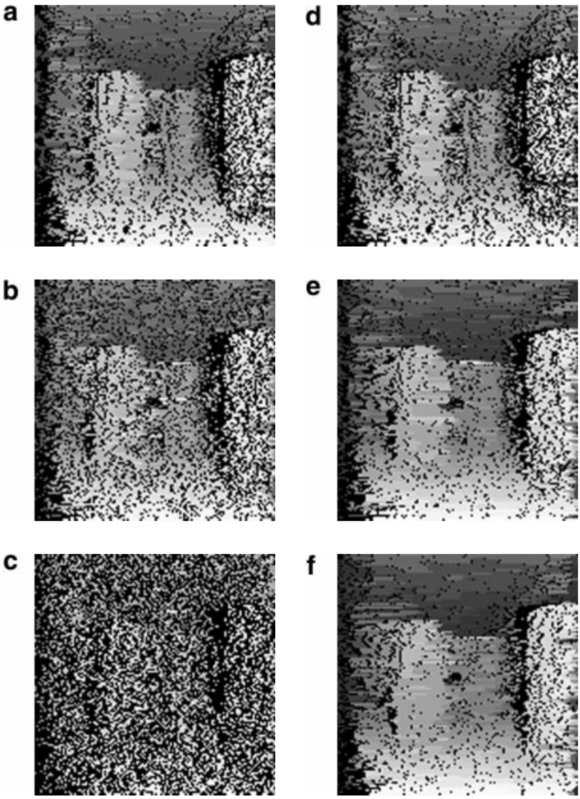

FIG. 3. Results of the dynamic programming algorithm in [1] with the original cost of occlusion C_O shown on the left and the modified cost $\tilde{C}_O$ shown on the right with three different values of σ. $P_D = 0.9$ for all examples. (a) $C_O = 0.93$, $\sigma = 4$; (b) $C_O = 0.24$, $\sigma = 8$; (c) $C_O = 0.015$, $\sigma = 10$; (d) $\tilde{C}_O = 0.676$, $\sigma = 4$; (e) $\tilde{C}_O = 0.676$, $\sigma = 8$; (f) $\tilde{C}_O = 0.676$, $\sigma = 10$.

In Fig. 3 we show the results of applying the dynamic programming matching scheme in [1] to the stereo images in Fig. 2 with both $\tilde{C}_O$ and C_O. To quantify the difference in performance, we computed the mean-square error (MSE) for each feature where disparity information exists and where the matching algorithm has declared a match.[5] The results are given in Table 1. The performance of the matching algorithm with the modified occlusion cost remains fairly constant for the range of σ considered. However, matching with the original occlusion cost in [1] shows deteriorating performance as the noise level increases. Not only does the frequency of occlusion declarations increase as the original occlusion cost approaches zero, but the accuracy of matching declarations decreases as well.

Our occlusion cost can also be extended to use local windows of features [2]. Experiments we have run with 3×3 windows of pixels reduced the mean-square disparity error by a factor of 2–3 over matching without windows.

5. CONCLUSION

We have presented an alternative cost of occlusion for the algorithm of [1]. Compared with the conditions considered in [1], our modification is appropriate for higher values of

[5] In other words, we do not include declared occlusions when computing the MSE. Moreover, we do not include those instances where a match is declared yet the true disparity map indicates occlusion.

sensor noise and a wider range of values for the probability of detection. The modified cost which we calculate can be integrated easily into the algorithm presented in [1], while maintaining the algorithm's other powerful features (such as minimizing changes in the matching between epipolar lines).

Reasonable values of σ and P_D in [1] (such as the values used there, $\sigma = 4$ and $P_D = 0.99$) will produce satisfactory disparity maps using the cost of occlusion presented there. Better performance can be expected if the cost C_O is made independent of σ (such as our proposed alternative $\tilde{C}_O$). Indeed, $\tilde{C}_O$, obtained through decision theoretic derivation, results in a more appropriate and intuitive relationship to the probability of detection.

APPENDIX: NOMENCLATURE

C_O	The original occlusion cost in [1]
$\tilde{C}_O$	The modified occlusion cost presented in this paper
$\hat{C}_O$	The modified occlusion cost using windows
$\mathbf{Z}_s$	Vector of measurements from camera s
$\mathbf{z}_{s,i}$	Measurement i from camera s
m_s	Number of measurements from camera s
Z_{i_1,i_2}	Candidate matching pair of features
σ^2	Variance of measurement error
δ_{i_1,i_2}	Matching indicator variable; 1 when a match is not declared
ϕ	Field of view of the camera
d	Dimension of measurment vectors

REFERENCES

1. I. Cox, S. Hingorani, S. Rao, and B. Maggs, A maximum likelihood stereo algorithm. *CVGIP: Image Understanding* **63**, 1996, 542–567.
2. G. Fielding, “Matching in Dynamic Stereo Image Sequences,” Ph.D. Thesis, Drexel University, 1999.
3. A. Papoulis, “Probability, Random Variables, and Stochastic Processes,” 3rd ed., McGraw-Hill, New York, 1991.