Disparity estimation window size

Tamás Frajka
Kenneth Zeger
University of California, San Diego
Department of Electrical and Computer Engineering
La Jolla, California 92093-0407
E-mail: frajka,zeger@code.ucsd.edu

Abstract. Disparity estimation plays a crucial role in many stereo image compression techniques. To reduce computational complexity most methods limit the estimation search area to a limited window. The performance of the disparity estimation depends on the choice of the limited search window. Most techniques use a predetermined value for the window size, which is not optimal over a wide range of images. We show how the choice of the window size affects the performance of the stereo image compression algorithm and propose a method to obtain a better search window size. Our simulation results indicate an improvement of up to 1.81 dB over rigid window size selection and with performance very close to the optimal selection. [DOI: 10.1117/1.1614808]

Subject terms: stereoscopic image; disparity estimation.

Paper 020528 received Dec. 9, 2002; revised manuscript received Apr. 9, 2003; accepted for publication Apr. 28, 2003.

1 Introduction

Stereoscopic image pairs represent a view of the same scene from two slightly different positions. When the two images are presented to the respective eye, the human observer perceives the depth in the scene as in three dimensions. One can obtain stereo pairs by taking pictures with two cameras that are placed in parallel 2 to 3 in. apart. A typical scenario is shown in Fig. 1.

Stereo images play an important part in applications such as remote sensing, surveillance, telemedicine, and computer vision. Stereo pairs present twice the amount of data to be stored or transmitted when compared with regular images, so efficient compression methods are required to reduce transmission delay and storage requirements.

The left and right views differ only in small areas of the scene, and independent coding of the two images does not take advantage of this inherent dependency. A wavelet-based method was presented in Ref. 1 that potentially describes all wavelet coefficients of one image, but only a subset of the coefficients of the other image. More sophisticated schemes use disparity estimation, a technique similar to motion estimation for video coding. Disparity estimation aims at finding the displacement of an object between the left and right images. Unlike in motion estimation, the displacement between the two images is restricted to a well-defined direction for all parts of the image. Block-matching-based methods were used in Refs. 2 and 3 and further reduction of complexity was achieved with hierarchical matching in Refs. 4, 5, and 6 and by matching using selective sample decimation in Ref. 7. To improve the matching performance of these techniques, others proposed more complex algorithms for disparity estimation, such as the use of overlapped blocks,[8] the combination of block matching with subspace projection,[9,10] rate-distortion optimization,[11] dynamic programming,[12] and generalized block matching.[13]

To find the best match for any given block one must search over all blocks of the corresponding image pair. To reduce the complexity of this operation the matching is generally limited to a smaller window. In previous works, this window size was some predetermined, fixed value. Because of the nature of the true disparity in images, such a predetermined value does not tend to work well across a wide range of images.

In this paper, we propose an efficient method for determining an estimate for the window size to be used with disparity estimation. This estimate is independent of the actual disparity estimation technique used and it reflects the underlying characteristics of the stereo image pair. We also show how the choice of window size affects the coding rate of the disparity vector field for both fixed rate and variable rate encoding. Our simulation results show that a proper search window size for disparity estimation can improve coding efficiency by up to 1.81 dB over using some predetermined value.

Section 2 gives a detailed description of disparity estimation. The effect of the search window size is analyzed in Sec. 3. Our method for window size estimation based on examining the correlation between the shifted image pairs is given in Sec. 4. Simulation results follow in Sec. 5 and we conclude with Sec. 6.

2 Disparity-Based Stereo Image Coding

Because of different perspectives in stereo imagery, a point in an object will be mapped to different coordinates in the left and right images. Let (x_l, y_l) and (x_r, y_r) denote the coordinates of an object point in the left and right images, respectively. The disparity is the difference between these vectors, $\mathbf{d}=(x_l-x_r, y_l-y_r)$. We assume the cameras are placed in parallel, so that $y_l-y_r=0$, and the disparity is limited to the horizontal direction. Let b denote the separation between the two cameras, f the focal length, and Z the depth of the object (or the distance of the object from the camera), as shown in Fig. 1. If all of these parameters are known, one can compute the disparity of each object as[2]

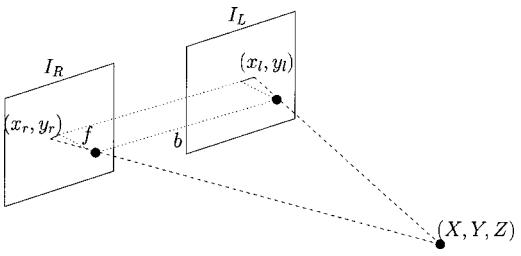


Fig. 1 Stereo camera system.

$$|d| = \frac{bf}{Z}. \tag{1}$$

This equation agrees with the intuition that objects closer to the camera will exhibit larger disparity than objects farther away.

To exploit the dependency between the left and right images most compression schemes use a conditional coder structure, as depicted in Fig. 2. One image of the pair serves as a reference image I_{ref} and the other, I_{pred}, is predicted from the reference image. The encoder describes to the decoder the reference image, the residual image (i.e., the difference of the predicted image and its estimate) I_{diff} and the disparity vectors used to obtain the estimate image. At the decoder, the reconstructed reference image $\hat{I}_{\text{ref}}$ and the decoded disparity vector field are used by the disparity compensation process to arrive at the estimate $\tilde{I}_{\text{pred}}$ of the predicted image. The reconstructed predicted image $\hat{I}_{\text{pred}}$ is obtained by adding the reconstructed difference image $\hat{I}_{\text{diff}}$ to $\tilde{I}_{\text{pred}}$. Most earlier methods used a discrete cosine transform (DCT)-based encoding for both the reference image and the residual image. Recently, progressive, wavelet-based techniques have been introduced[14–16] to replace the DCT-based encoding, and Perkins[17] showed that in general the conditional coder structure is suboptimal in the rate-distortion sense.

If the parameters in Eq. (1) are known ahead of time, then, in principle, for each pixel one could compute the disparity and use that in the prediction. Unfortunately these parameters are seldom available at the time of the encoding. Using disparity estimation, one could try to obtain approximate values for each pixel of the image. Since this process would be quite complex if performed for each pixel individually, it is usually carried out for groups of pixels instead. One such grouping is the use of $k \times l$ nonoverlapping blocks.

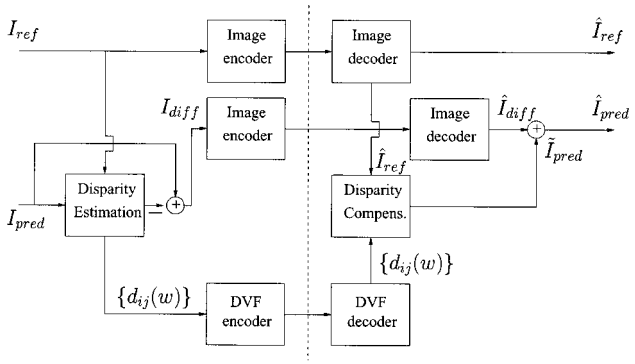


Fig. 2 Stereo image coding system based on a conditional coder structure. The left side of the dashed line is the encoder and the right side is the decoder.

Let D denote a distortion measure between image blocks and let $I[i,i',j,j']$ denote a $(i'-i)\times(j'-j)$ block with upper left coordinates (i,j) and lower right coordinates (i',j'). Then the disparity estimate for a block with upper left coordinates (i,j), assuming only horizontal displacement, is defined as

$$\tilde{d}_{ij}(w) = \underset{d \le w}{\operatorname{argmin}} D[I_{\text{pred}}(i,i+k,j,j+l), I_{\text{ref}}(i+d,i+k+d,j,j+l)], \tag{2}$$

where w is the disparity window size within which the search is performed.

The quantity $\tilde{d}_{ij}(w)$ is the horizontal amount by which a $k \times l$ block in one image must be shifted to most closely match in similarity a given $k \times l$ block in another image. The two most often used similarity measures in block matching are the maximum absolute difference and the mean-squared error. Using the mean-squared error, Eq. (2) becomes

$$\tilde{d}_{ij}(w) = \underset{d \le w}{\operatorname{argmin}} \sum_{m=i}^{i+k} \sum_{n=j}^{j+l} [I_{\text{pred}}(m,n) - I_{\text{ref}}(m+d,n)]^2, \tag{3}$$

where $I(m,n)$ is the pixel intensity value at coordinate (m,n).

For any positive number w, define the disparity vector field (DVF) of an $X \times Y$ image to be the matrix of integers

$$\{\tilde{d}_{ij}(w)\}$$

where $0 \le i \le X-1$, $0 \le j \le Y-1$, and i and j are divisible by k and l, respectively. The DVF is sent to the decoder, usually DPCM encoded, and followed by adaptive arithmetic coding.

The reconstructed predicted image depends on the quality of the reconstructed reference image $\hat{I}_{\text{ref}}$, the DVF, and the reconstructed difference image $\hat{I}_{\text{diff}}$. The image predicted from the reconstructed reference image using the DVF can be written as

$$\hat{I}_{\text{pred}} = \tilde{I}_{\text{pred}} + \hat{I}_{\text{diff}},$$

where $\tilde{I}_{\text{pred}}$ depends on $\hat{I}_{\text{ref}}$ and $\{\tilde{d}_{ij}(w)\}$.

Then given the compressed reference image $\hat{I}_{\text{ref}}$ of size $X \times Y$, the predicted image estimate is

$$\tilde{I}_{\text{pred}} = \{\hat{I}_{\text{ref}}[i+\tilde{d}_{ij}(w), i+\tilde{d}_{ij}(w)+k, j, j+l]: 0 \le i \le X-1, 0 \le j \le Y-1, k | i, l | j\}$$

and the disparity estimation distortion is defined as

$$\tilde{D}_{\text{pred}}(\{\tilde{d}_{ij}(w)\},\hat{I}_{\text{ref}}) = \|I_{\text{pred}} - \tilde{I}_{\text{pred}}\|^2 = \sum_{m,n} [I_{\text{pred}}(m,n) - \tilde{I}_{\text{pred}}(m,n)]^2.$$

The total rate is the sum of the rates of coding the reference image, the disparity vector field, and the residual image, namely,

$$R_T = R_{\text{ref}} + R_{\text{DVF}} + R_{\text{diff}}.$$

The distortion is defined as the average of the distortions of the two images:

$$D_T = (D_{\text{ref}} + D_{\text{pred}})/2 = (\|I_{\text{ref}} - \hat{I}_{\text{ref}}\|^2 + \|I_{\text{pred}} - \hat{I}_{\text{pred}}\|^2)/2.$$

The overall rate-distortion performance depends on the rate allocation between R_{ref}, R_{DVF}, and R_{diff}.

In practice, without the DVF no real stereo effect can be achieved. It is often assumed that the minimum coding rate is at least R_{DVF} and that the DVF is encoded first, leaving the rate allocation to the reference and difference images for the remaining available rate. The most widely used technique for the encoding of the DVF is differential pulse code modulation (DPCM) followed by entropy coding.[6,7,12,16,18,19] Entropy coding alone is employed in Refs. 14 and 20, and fixed length coding in Ref. 10. Our analysis focuses on these methods.

Using an embedded coding method, as proposed in Ref. 14, the rate allocation can be performed automatically without computationally complex optimization.

3 Disparity Window Size

Most disparity estimation algorithms use a predetermined, fixed maximum search window size w. For example, the following window sizes or ranges of window sizes were used: 15 in Ref. 21, 63 in Refs. 18 and 22, $\{-8,...,48\}$ in Ref. 10, $\{-30,...,30\}$ in Ref. 6, $\{-63,...,63\}$ in Refs. 7 and 23, and some fixed unspecified values were used in Refs. 11, 12, 14, and 20. Many of these choices reflect that disparity estimation was modeled after motion estimation. As noted, the disparity of each object is inversely proportional to its distance from the camera. A predetermined, fixed window size may not work well for any image.

The best disparity estimate is a function of the window size as given in Eq. (3). It is nondecreasing in w, since searching over a larger area can only improve the displacement estimate; i.e., if $w_1 < w_2$, then $\tilde{d}_{ij}(w_1) \leq \tilde{d}_{ij}(w_2)$ for all blocks in the image. Then, the disparity estimation distortion is nonincreasing in w, i.e.,

$$\tilde{D}_{\text{pred}}(\{\tilde{d}_{ij}(w_2)\},\hat{I}_{\text{ref}}) \leq \tilde{D}_{\text{pred}}(\{\tilde{d}_{ij}(w_1)\},\hat{I}_{\text{ref}}) \quad \text{if } w_1 \leq w_2,$$

since searching for the best match over a larger area can only improve the outcome.

The DVF must be transmitted to the decoder for the reconstruction of the predicted image. The transmission rate R_{DVF} is a nondecreasing function of the disparity search window size w. It is most obvious in the case of fixed-length encoding where $\log_2 w$ bits are used to transmit a disparity value.

If the window size is smaller than the true disparity of certain objects in an image, the disparity estimation process cannot provide the best prediction, and the rate-distortion performance can be improved by increasing w. On the other hand, if the window size is significantly larger than the disparity of the objects in the image, their encoding may not be the most efficient. Even with DPCM coding followed by arithmetic coding for the DVF, too large a window size and thus too large a potential maximal value would dilute the probability model and thus yield a suboptimal coding rate.

In Ref. 24 the authors showed that using adaptive arithmetic coding, the description length R of a source of alphabet size n is

$$R = \log_2 \left[\frac{\Pi_{i=0}^{t-1}(n+i)}{\Pi_{i=1}^{k} c_i!} \right], \tag{4}$$

where k is the number of alphabet symbols that occur in the stream to be compressed, c_i is the number of occurrences of symbol i, and t is the total length of the stream. In the case of DVF coding, t equals the number of disparity vectors and k is the number of distinct displacement values. If only arithmetic coding is used, then $n = w$, and if it is coupled with DPCM coding, then $n = 2w + 1$.

Let $d^*_{\max}$ denote the maximum true disparity of any object in the image and $R_{\text{DVF}}(d^*_{\max})$ the description length using an ideal disparity estimator. If one chooses a disparity window size larger than $d^*_{\max}$, then the change in rate from Eq. (4) is

$$\Delta R_{\text{DVF}} = R_{\text{DVF}}(w) - R_{\text{DVF}}(d^*_{\max}) = \log_2 \frac{\Pi_{i=0}^{t-1}(n_w+i)/\Pi_{i=1}^{k'} c'_i!}{\Pi_{i=0}^{t-1}(n^*+i)/\Pi_{i=1}^{k} c_i!}, \tag{5}$$

where n^* and n_w are the number of possible input symbols to the arithmetic coding using window size of d^* and w, respectively; $k' \geq k$, i.e., the new disparity values found by increasing the window size from $d^*_{\max}$ to w are added after the initial k found using a window size of $d^*_{\max}$; and c_i and c'_i represent the occurrence of the same symbol, obtained using a search window size of $d^*_{\max}$ and w, respectively. For $k < i \leq k'$, $c_i = 0$ and for some $i \leq k$, c'_i can become zero. For such cases, we adopt the usual convention that $0! = 1$ in Eq. (5).

Note, if $k' = k$, the increase in disparity window size does not affect the DVF and one can just transmit $d_{\max}$ to the decoder and incur only a small overhead penalty for choosing too large a disparity window size. Unfortunately, because of photometric variations and occlusion, the disparity estimation process often finds false matches for a block beyond the true disparity of the object.

Since $w \geq d^*_{\max}$, let $w = d^*_{\max} + \Delta_w$ and $n_w = n^* + \Delta_n$. To show that the rate does increase with increasing window size it suffices to show that the argument of the logarithm in Eq. (5) is greater than 1.

We have

Table 1 Effect of disparity window size choice on predicted image peak SNR (PSNR) alone, where *L* or *R* indicates whether the left or right image was predicted using the uncompressed version of the other image, and the rate is the sum of the rate of the DVF (coded using DPCM and arithmetic coding) and the difference image.

Image	*L* or *R*	Rate (bpp)	w_{opt}	PSNR(64) (dB)	PSNR(w_{opt}) (dB)
"Room"	*L*	0.1	10	30.51	31.53
"Room"	*R*	0.1	10	32.39	32.70
"Closeup"	*L*	0.2	171	24.03	25.49
"Closeup"	*R*	0.2	262	23.95	25.71
"Outdoors"	*L*	0.2	434	21.74	21.82
"Outdoors"	*R*	0.2	386	21.77	21.85

$$\frac{\Pi_{i=0}^{t-1}(n_w+i)}{\Pi_{i=0}^{t-1}(n^*+i)}=\prod_{i=0}^{t-1}\frac{n^*+\Delta_n+i}{n^*+i}>\left(1+\frac{\Delta_n}{n^*+t-1}\right)^t, \qquad (6)$$

and

$$\frac{\Pi_{i=1}^{k}c_i!}{\Pi_{i=1}^{k'}c_i'!}=\prod_{i=1}^{k'}\frac{c_i!}{c_i'!}$$

$$=\prod_{i:c_i>c_i'}[(c_i'+1)\ldots c_i]\prod_{i:c_i<c_i'}\frac{1}{(c_i+1)\ldots c_i'}$$

$$>\prod_{i:c_i>c_i'}(c_i'+1)^{c_i-c_i'}\prod_{i:c_i<c_i'}(1/c_i')^{c_i'-c_i}$$

$$>(c'_{\min,*}+1)^{\Sigma_{i:c_i>c_i'}(c_i-c_i')}(1/c'_{\max,w})^{\Sigma_{i:c_i<c_i'}(c_i'-c_i)}$$

$$=\left(\frac{c'_{\min,*}+1}{c'_{\max,w}}\right)^{\Sigma_{i:c_i>c_i'}(c_i-c_i')}, \qquad (7)$$

where $c'_{\min,*}=\min_{i:c_i>c_i'}c_i'$, $c'_{\max,w}=\max_{i:c_i<c_i'}c_i'$, and $\Sigma_{i:c_i>c_i'}(c_i-c_i')=\Sigma_{i:c_i<c_i'}(c_i'-c_i)$.

Thus Eqs. (6) and (7) imply that

Table 2 Effect of disparity window size choice on overall image quality, where *L* or *R* indicates whether the left or right image was predicted using the compressed version of the other image, and the rate is the sum of the rate of the DVF (coded using DPCM and arithmetic coding), the difference image, and the reference image.

Image	*L* or *R*	Rate (bpp)	w_{opt}	PSNR(64) (dB)	PSNR(w_{opt}) (dB)
"Room"	*L*	0.2	10	27.45	27.93
"Room"	*R*	0.2	10	27.93	28.02
"Closeup"	*L*	0.4	177	29.48	30.33
"Closeup"	*R*	0.4	273	29.27	30.22
"Outdoors"	*L*	0.4	91	23.18	23.20
"Outdoors"	*R*	0.4	97	23.16	23.17

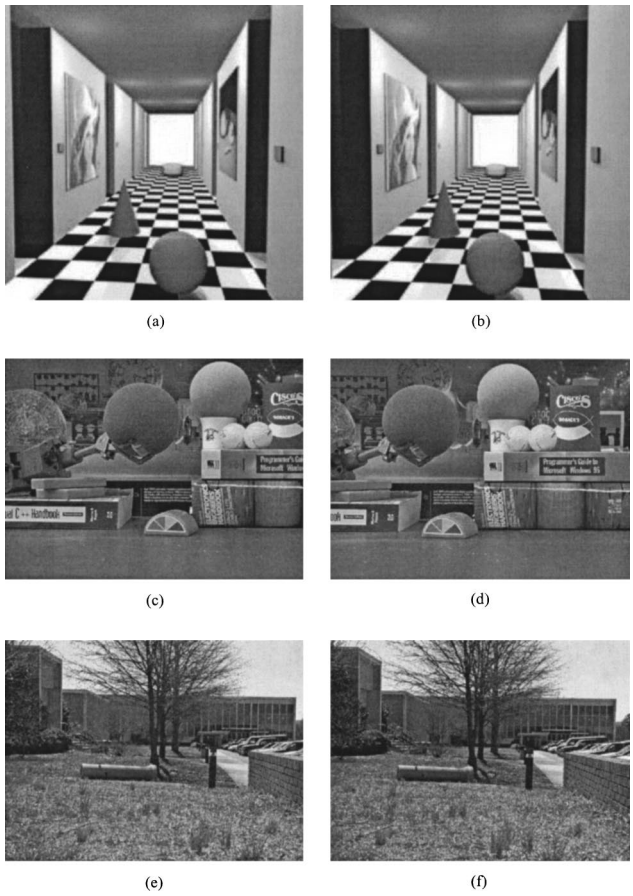


Fig. 3 Original images of the (a) and (b) "Room," (c) and (d) "Closeup," and (e) and (f) "Outdoors" stereo pairs.

$$\frac{\Pi_{i=0}^{t-1}(n_w+i)/\Pi_{i=1}^{k'}c_i'!}{\Pi_{i=0}^{t-1}(n^*+i)/\Pi_{i=1}^{k}c_i!}=\frac{\Pi_{i=0}^{t-1}(n_w+i)\Pi_{i=1}^{k}c_i!}{\Pi_{i=0}^{t-1}(n^*+i)\Pi_{i=1}^{k'}c_i'!} \qquad (8)$$

$$>\left(1+\frac{\Delta_n}{n^*+t-1}\right)^t\left(\frac{c'_{\min,*}+1}{c'_{\max,w}}\right)^{\Sigma_{i:c_i>c_i'}(c_i-c_i')}. \qquad (9)$$

In practice, the changes in the occurrence count are around one or two and the number of all changes is typically less than 1 or 2% of all disparity vectors.

The first term on the right-hand side of Eq. (9) is always greater than 1. The second term is approximately 1 if arithmetic coding is used without DPCM, or when DPCM and arithmetic coding are used, but all the displacement vectors found using the larger window size create new difference values.

For DPCM followed by arithmetic coding the ratio $(c'_{\min,*}+1)/c'_{\max,w}$ can be less than 1 if the new displacement values create differences that existed for window size d^*. In that case, the lower bound in Eq. (9) is not useful in bounding the ratio in Eq. (8).

For this case, we evaluated the actual value of this ratio for 23 test images in the following experiment. For each

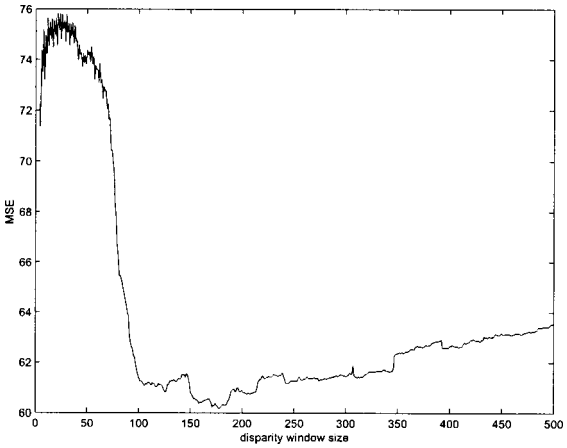


Fig. 4 Distortion D_T as a function of maximum window size w for the "Closeup" stereo image pair at an overall bit rate of 0.4 bpp.

stereo image pair, first the optimal window size d^* was estimated using exhaustive search disparity estimation with a window size equal to half the image size. Then d^* was chosen as the largest displacement value that was used for at least 1% of all blocks. Given the values of d^*, the ratio in Eq. (8) was computed using $w=d^*+1$, $w=d^*+5$, $w=d^*+10$, $w=d^*+50$, and $w=d^*+100$. For all images and for all window size choices the ratio was observed to always be larger than 1. While this observation is not always guaranteed, for practical purposes we assume

$$\frac{\Pi_{i=0}^{t-1}(n_w+i)/\Pi_{i=1}^{k'}c_i'!}{\Pi_{i=0}^{t-1}(n^*+i)/\Pi_{i=1}^{k}c_i!}\geqslant 1$$

for both DPCM/arithmetic coding and plain arithmetic coding of the disparity vector field.

The choice of the optimal window size is image dependent. Since $d^*_{\max}$ is generally unavailable at the time the images are compressed, it is necessary to be able to find a good maximum in real time without having to perform an exhaustive search. Using predetermined, fixed values does not work well over a wide range of images, as is shown in Tables 1 and 2. These results were obtained for the following three images: the 256×256 synthetic "Room" stereo image pair, and the 640×480 "Outdoors" and "Closeup" image pairs (Fig. 3). The disparity estimation and the encoding are the same as in Ref. 15, where the DVF was encoded using DPCM followed by adaptive arithmetic coding. The optimal window size was determined using an exhaustive search. It is optimal given the encoding mechanism and the target bit rate. The choice of a predetermined window size of 64 displacement values was motivated by previous results in the literature using that value. The PSNR for the predicted image alone (Table 1) is defined as

Table 3 Comparison of predetermined, optimal, and approximated disparity window size choices on the predicted right image quality at 0.2 bpp using the left image as a reference image, where the rate is the sum of the rate of the DVF (coded using DPCM and arithmetic coding) and the difference image.

Image	Size	w_{opt}	w_C	PSNR(w_{opt}) (dB)	PSNR(w_C) (dB)	PSNR(64) (dB)	PSNR(200) (dB)
"Toys"	212×134	12	32	30.14	**29.36**	28.98	28.79
"Fruit"	512×512	503	32	34.67	**34.26**	34.04	33.66
"Whgarden"	250×250	29	24	25.98	**25.89**	25.70	25.46
"Cart"	250×250	69	64	30.94	**30.88**	**30.88**	30.35
"Parts"	512×512	456	264	35.40	**33.91**	33.43	33.68
"Rubik"	512×512	4	40	42.44	**41.89**	41.86	41.77
"Arch"	512×512	4	24	42.55	**41.39**	40.35	40.53
"Room"	256×256	10	16	35.73	**35.68**	35.55	35.15
"Closeup"	640×480	262	128	25.71	**25.53**	23.95	25.49
"Outdoors"	640×480	386	120	21.85	21.80	21.77	**21.81**
"Oldbridge"	320×192	317	24	25.55	**25.48**	25.40	25.27
"Ball"	512×512	510	24	43.53	**41.99**	41.39	41.49
"Booklr"	250×250	6	40	34.46	**33.32**	32.96	32.55
"Plants"	512×400	112	120	30.29	**30.27**	25.78	30.14
"Cart-alt"	250×250	57	40	32.20	**32.10**	32.05	31.66
"Bottle"	320×240	17	48	27.53	**27.24**	27.09	26.65
"Apple"	512×512	504	192	26.09	25.39	**25.49**	25.39
"Manege"	720×288	31	72	26.77	26.50	**26.54**	26.07
"Book"	512×512	4	88	37.45	35.50	**35.67**	35.18
"Aqua"	360×288	251	72	26.04	25.52	25.47	**25.88**
"Sphere"	256×256	252	80	33.17	32.61	30.17	**32.97**
"Tunel"	720×288	509	48	27.31	25.58	25.64	**26.46**
"Fjord"	233×256	4	120	28.60	28.35	27.93	**28.54**

Table 4 Comparison of predetermined, optimal, and approximated disparity window size choices on overall image quality at 0.4 bpp using the left image as a reference image, where the rate is the sum of the rate of the DVF (coded using DPCM and arithmetic coding), the difference image, and the reference image.

Image	Size	w_{opt}	w_C	PSNR(w_{opt}) (dB)	PSNR(w_C) (dB)	PSNR(64) (dB)	PSNR(200) (dB)
"Toys"	212×134	12	32	33.76	**33.14**	32.77	32.77
"Fruit"	512×512	5	32	38.19	**37.92**	37.77	37.59
"Whgarden"	250×250	33	24	28.08	**28.02**	27.96	27.90
"Cart"	250×250	70	64	35.11	**35.10**	**35.10**	34.76
"Parts"	512×512	456	264	39.62	**39.01**	38.92	38.95
"Rubik"	512×512	5	40	46.89	**46.66**	46.64	46.54
"Arch"	512×512	4	24	45.16	**44.75**	44.71	44.70
"Room"	256×256	11	16	33.54	**33.52**	33.46	33.28
"Closeup"	640×480	273	128	30.22	**30.18**	29.27	30.13
"Outdoors"	640×480	97	120	23.17	**23.17**	23.16	23.16
"Oldbridge"	320×192	11	24	26.80	**26.76**	26.70	26.64
"Ball"	512×512	506	24	45.34	**44.99**	44.81	44.73
"Booklr"	250×250	5	40	37.85	**37.24**	37.07	36.82
"Plants"	512×400	109	120	33.87	**33.84**	32.03	33.77
"Cart-alt"	250×250	25	40	36.40	**36.32**	36.21	36.05
"Bottle"	320×240	18	48	26.26	**26.18**	26.12	25.95
"Apple"	512×512	504	192	27.54	27.22	**27.30**	27.23
"Manege"	720×288	31	72	28.18	28.09	**28.11**	27.97
"Book"	512×512	4	88	42.34	41.33	**41.44**	41.08
"Aqua"	360×288	266	72	26.02	25.80	25.78	**25.95**
"Sphere"	256×256	243	80	32.99	32.81	31.98	**32.93**
"Tunel"	720×288	510	48	28.85	28.13	28.13	**28.48**
"Fjord"	233×256	4	120	30.16	29.80	29.68	**29.86**

$$\mathrm{PSNR}=10\log_{10}\frac{255^2}{D_{\mathrm{pred}}}, \tag{10}$$

and for the stereo image pair (Table 2) as

$$\mathrm{PSNR}=10\log_{10}\frac{255^2}{(D_{\mathrm{ref}}+D_{\mathrm{pred}})/2}. \tag{11}$$

The notation PSNR(x) in the tables refers to the PSNR obtained using a maximum window size x in the disparity estimation process.

These results indicate that using the optimal window size is always better than using a predetermined one. In the case of the "Closeup" and "Room" images the improvement ranges between 0.31 and 1.76 dB for the predicted image alone, and 0.09 and 0.95 dB for the stereo image pair. Note that with the "Room" stereo pair, the optimal window size is smaller than the predetermined value, while with the "Closeup" image it is larger.

4 Determining Disparity Window Size

Conducting a full search over all possible disparity search window size values is a time-consuming process. It is also analytically difficult since the disparity and the resulting distortion are image dependent. Even quick search methods are difficult to implement because the distortion at a given target rate is not a monotone, concave, or convex function of the disparity window size, as revealed in Fig. 4.

Here we propose a heuristic approach that yields good results for many different stereo image pairs.

In Ref. 3, the authors use correlation measurements between the shifted left and right images of the stereo pair to determine a global disparity vector. The maximum of the correlation is assumed at the "average" disparity of the image. While using just a single value for the entire image is not the optimal strategy, it still gives an indication of the typical disparity values to be found in the image pair.

We use a similar strategy based on the correlation coefficient between two images:

$$C(I_1,I_2)=\frac{\mathrm{cov}(I_1,I_2)}{\sigma_{I_1}\sigma_{I_2}},$$

where I_1 and I_2 are images given by their intensity values, and cov(.) is the covariance (the expectations are taken over the sample distribution of the images). As one image is shifted with respect to the other, columns at the leading end of the shifting move out of the image and columns on the trailing end must be filled. We propose to fill those using a mirroring of the columns at the trailing end. This is a natural extension of the disparity estimation process. In disparity estimation, one tries to find matching blocks along the same horizontal direction for each block. However, at the far edge of the matching, that direction points outside the image at which point the estimation will look for matches in the reverse direction.

This is a computationally complex process since the correlation is determined for each disparity value. We perform the correlation computation on a subsampled version of the images instead. For computational simplicity, the subsampling is carried out as a succession of averaging. Using images subsampled by 8 in each direction reduces the computational cost by a factor of 64. Let $C_8(d)$ denote the correlation value for shift d when the images subsampled by 8 in each direction are used. Once these correlation values are obtained, the algorithm finds the maximum value, $C_{8,\max}=\max_d C_8(d)$. Let d denote the value after which the correlation drops below $C_{8,\max}/2$, that is, $C_8(d)\geq C_{8,\max}/2$, but $C_8(d+1)<C_{8,\max}/2$. The window size is chosen as $w_C=8d$, the displacement corresponding to d in the full resolution image. The particular choice of the factor of 1/2 as the cutoff value was motivated by experiments that indicated good approximation of the optimal disparity window size given the DVF encoding method and the target rate.

This results in a less accurate estimate of the window size as it is quantized to the subsampling factor, but it enables a faster computation.

5 Simulation Results

We tested our proposed method over a large set of test images. The PSNR is computed as defined in Eqs. (10) and (11). Tables 3 and 4 show the results using the correlation-based estimation for maximum window size. Two different scenarios of predetermined window sizes are shown as well. The correlation-based approximation works very well, especially when compared with predetermined window sizes. Neither of the predetermined values show a clear advantage over the other predetermined value, indicating that one fixed choice does not work for all images. Our proposed technique yields the best performance on most images and gets as close as 0.02 dB to the best achievable result for the given DVF coding technique.

When compared with a scheme with a predetermined window size, this method increases complexity due to the correlation computation. This increase is far less than performing a full-search disparity compensation and encoding of the DVF and the residual image to find the true optimum value.

6 Conclusion

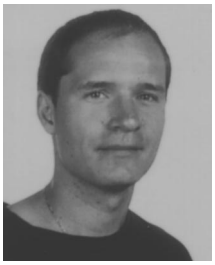
We presented a technique that estimates the optimal choice for the disparity estimation search window size. It can be combined with any particular disparity estimation algorithm. The performance improvement due to an adaptive choice of search window is up to 1.81 dB. While it leads to some increase in computation complexity, our proposed method is still computationally less complex than trying to find the optimal window size using an exhaustive search.

Acknowledgments

This research was supported in part by the National Science Foundation and the University of California, San Diego, Center for Wireless Communications.

References

1. W. D. Reynolds and R. V. Kenyon, ''The wavelet transform and the suppression theory of binocular vision for stereo image compression,'' in *Proc. IEEE Int. Conf. on Image Processing*, Vol. 2, pp. 557–560 (1996).
2. P. An, Z. Zhang, and L. Shi, ''Theory and experiment analysis of disparity for stereoscopic image pair,'' in *Proc. Int. Symp. on Intelligent Multimedia, Video and Speech Processing*, Vol. 1, pp. 68–71, IEEE, Hong Kong (2001).
3. H. Yamaguchi, Y. Tatehira, K. Akiyama, and Y. Kobayashi, ''Stereoscopic images disparity for predictive coding,'' in *Proc. Int. Conf. on Acoustics, Speech and Signal Processing*, Vol. 3, pp. 1976–1979 (1989).
4. S. Sethuraman, M. W. Siegel, and A. G. Jordan, ''A multiresolution region based segmentation scheme for stereoscopic image compression,'' *Proc. SPIE* **2419**, 265–273 (1995).
5. Y. Zhang and G. Li, ''An efficient hierarchical disparity estimation algorithm for stereoscopic video coding,'' in *Proc. IEEE Asia-Pacific Conf. on Circuits and Systems*, Vol. 1, pp. 744–747 (2000).
6. W. Zheng, Y. Shishikui, Y. Tanaka, and I. Yuyama, ''Disparity estimation with hierarchical block correlation for stereoscopic image coding,'' *Syst. Comput. Jpn.* **28**(6), 1–9 (June 1997).
7. W.-H. Kim and S.-W. Ra, ''Fast disparity estimation using geometric properties and selective sample decimation for stereoscopic image coding,'' *IEEE Trans. Consum. Electron.* **45**(1), 203–209 (1999).
8. W. Woo and A. Ortega, ''Overlapped block disparity compensation with adaptive windows for stereo image coding,'' *IEEE Trans. Circuits Syst. Video Technol.* **10**(2), 194–200 (2000).
9. H. Aydinoğlu and M. H. Hayes III, ''Stereo image coding: a projection approach,'' *IEEE Trans. Image Process.* **7**(4), 506–516 (1998).
10. S.-H. Seo and M. R. Azimi-Sadjadi, ''A 2-D filtering scheme for stereo image compression using sequential orthogonal subspace updating,'' *IEEE Trans. Circuits Syst. Video Technol.* **11**(1), 52–66 (2001).
11. D. Tzovaras and M. G. Strintzis, ''Disparity estimation using rate-distortion theory for stereo image sequence coding,'' in *Proc. Int. Conf. on Digital Signal Processing*, Vol. 1, pp. 413–416, IEEE (1997).
12. H.-H. Jeon, J.-H. Kim, and D.-G. Oh, ''Stereo image coding with disparity compensation using dynamic programming,'' in *Proc. Int. Symp. on Consumer Electronics*, pp. 214–217, IEEE (1997).
13. V. E. Seferidis and D. V. Papadimitriou, ''Improved disparity estimation in stereoscopic television,'' *Electron. Lett.* **29**(9), 782–783 (1993).
14. N. V. Boulgouris and M. G. Strintzis, ''Embedded coding of stereo images,'' in *Proc. IEEE Int. Conf. on Image Processing*, Vol. 3, pp. 640–643 (2000).
15. T. Frajka and K. Zeger, ''Residual image coding for stereo image compression,'' *Opt. Eng.* **42**(1), 182–188 (2003).
16. T. Palfner, A. Mali, and E. Müller, ''Progressive coding of stereo images using wavelets and overlapping blocks,'' in *Proc. IEEE Int. Conf. on Image Processing*, Vol. 2, pp. 213–216 (2002).
17. M. G. Perkins, ''Data compression of stereopairs,'' *IEEE Trans. Commun.* **40**(4), 684–696 (1992).
18. H. Aydinoğlu and M. H. Hayes III, ''Stereo image coding,'' in *Proc. Int. Symp. on Circuits and Systems*, Vol. 1, pp. 247–250, IEEE (1995).
19. M. S. Moellenhoff and M. W. Maier, ''Transform coding of stereo image residuals,'' *IEEE Trans. Image Process.* **7**(6), 804–812 (1998).
20. Q. Jiang, J. J. Lee, and M. H. Hayes III, ''A wavelet based stereo image coding algorithm,'' in *Proc. Int. Conf. on Acoustics, Speech and Signal Processing*, Vol. 6, pp. 3157–3160, IEEE (1999).
21. W. Woo and A. Ortega, ''Optimal blockwise dependent quantization for stereo image coding,'' *IEEE Trans. Circuits Syst. Video Technol.* **9**(6), 861–867 (1999).
22. H. Aydinoğlu, F. Kossentini, Q. Jiang, and M. H. Hayes III, ''Region-based stereo image coding,'' in *Proc. IEEE Int. Conf. Image Processing*, Vol. 2, pp. 57–60 (1995).
23. W.-H. Kim, J.-Y. Ahn, and S.-W. Ra, ''An efficient disparity estimation algorithm for stereoscopic image compression,'' *IEEE Trans. Consum. Electron.* **43**(2), 165–172 (1997).
24. P. G. Howard and J. S. Vitter, ''Analysis of arithmetic coding for data compression,'' in *Proc. Data Compression Conf.*, Vol. 6, pp. 3–12, IEEE (1991).



Tamás Frajka received his MS degree in computer science from the Technical University of Budapest in 1995 and his PhD degree in electrical and computer engineering at the University of California at San Diego, La Jolla, in 2003.

Kenneth Zeger received both his SB and SM degrees in electrical engineering and computer science from the Massachusetts Institute of Technology in 1984 and both his MA degree in mathematics and his PhD degree in electrical and computer engineering from the University of California, Santa Barbara, in 1989 and 1990, respectively. He was an assistant professor of electrical engineering with the University of Hawaii from 1990 to 1992. He was an assistant professor from 1992 to 1995 and an associate professor from 1995 to 1996 with the Department of Electrical and Computer Engineering and the Coordinated Science Laboratory, the University of Illinois, Urbana-Champaign. He was an associate professor from 1996 to 1998 and has been a professor since 1998 with the Department of Electrical and Computer Engineering, at University of California, San Diego. He received a National Science Foundation (NSF) Presidential Young Investigator Award in 1991. He was associate editor at-large for the *IEEE Transactions on Information Theory* from 1995 to 1998, a member of the Board of Governors of the IEEE Information Theory Society from 1998 to 2000, and is an IEEE fellow.