Block-based Illumination Compensation and Search Techniques for Multiview Video Coding

Joaquin López, Jae Hoon Kim, Antonio Ortega
{jlopez,jaekim,ortega}@sipi.usc.edu

Signal and Image Processing Institute
Department of Electrical Engineering
University of Southern California
Los Angeles, CA 90089-2564

George Chen[†]
george-qian.chen@st.com

[†]Advanced System Technology
STMicroelectronics
4690 Executive Dr, Suite 200
San Diego, CA 92121

***Abstract*— We consider block-based interview coding techniques for multiview video sequences that are robust to illumination variations across views. We consider both global illumination differences which could be due to lack of camera calibration or heterogeneity, as well as local ilumination changes caused by lack of camera aligment. We propose a two-parameter model for illumination compensation (IC) that is used at the block matching search step and can be adaptively applied by taking into account the rate-distortion characteristics of each block. We also present a modified search (MS) method for block-based interview disparity estimation that makes use of feature points in each frame in order to provide a good prediction of likely disparity vectors. We present coding experiments on different multiview sequences based on the H.264/AVC reference codec. The proposed techniques show significant coding gains for interview coded frames, as compared to methods that do not employ IC and MS.**

I. INTRODUCTION

Multiview video systems are used for capturing a scene from different viewpoints, so that 3-D information of a scene is contained in the resulting set of sequences. Such videos are emerging as a potential tool for new multimedia services, e.g., 3-D TV and free viewpoint video. Recently, MPEG has started to explore 3D Audio-Visual (3DAV) technology. Because a multiview video contains several video sequences from different cameras, widespread use of multiview video requires the design of efficient compression techniques [3], [4], [7], [9], [11], [13].

A straightforward approach for multiview video encoding would be to use standard video coding techniques independently on the video sequences corresponding to each camera. In this paper we consider interview coding approaches that aim at exploiting the correlation across views. We use block-based predictive techniques similar to the motion compensated predictive coding techniques used in MPEG-2, H.264/AVC, etc. After block-based disparity compensation we compress the interview prediction residuals using standard approaches (in our case coding tools in the H.264/AVC standard.) While interview and interframe coding appear to be similar, interview coding poses specific problems that may render inefficient a direct application of motion search and estimation techniques [3]. In this paper, we focus on (i) global and local illumination mismatches across views and (ii) efficient techniques to predict disparity vectors, aimed at reducing the disparity estimation complexity and improving overall coding efficiency.

Note that block-based approaches for disparity estimation (DE) and compensation have been proposed for multiview coding, e.g., [2], but we believe that we are the first to consider block based illumination compensation along with disparity prediction. Note also that block-based predictive techniques have also been frequently used in the particular case of stereo coding but the basic premise of our work is that multiview video sequences are inherently more difficult to encode than stereo sequences because video capture conditions are more likely to be heterogeneous across views. Thus, we assume that some multiview video systems of interest will be built by using heterogeneous cameras, or cameras that have not been perfectly calibrated; this could lead to differences in luminance and chrominance when the same parts of a scene are viewed with different cameras, affecting negatively the DE algorithm. Moreover, camera distance and positioning also affects illumination, in the sense that the same object surface can appear to be different from different angles. This is because in general illumination sources are not situated infinitely far away from the scene, e.g., in the case of indoor scenes. In stereo video it is easier to adjust more accurately the parameters involved, such as camera distance and alignment, in order to make coding easier. Given that in multiview systems there are too many variables to adjust, calibration of cameras becomes more difficult: the higher the number of cameras, the higher the complexity.

Illumination compensation techniques have also been used in the context of motion compensated video coding. For example, in [8], global IC and local refinement are used but the local IC parameters are only selected *after* the best matching block has been found. Instead, in our approach we use IC as part of the search process, which is important when illumination mismatches can be significant. A similar approach is followed in [10] but this is based on multiplicative illumination model, while we consider both multiplicative and additive terms. Similarly, [6] also proposes a modified motion estimation (ME) but a global IC model is used. In summary, in multiview video environments the impact of illumination mismatches can be very significant and thus we propose that global models are not sufficient and the IC step has to be fully integrated with DE process.

In standard ME, an efficient technique to improve the quality of the motion vectors while reducing the search complexity consists of using the motion vectors for neighboring blocks to predict the motion of the current block e.g., a median predictor in H.264/AVC. Approaches like these are efficient if multiple blocks share similar motion *and* the motion vectors estimated are accurate. We have observed that these assumptions tend not to be valid in the context of DE for multiview video. First, the disparity of each vector depends primarily on its depth in the scene, which can lead to lack of smoothness in the disparity field. Second, the illumination mismatches discussed above can lead to disparity vectors being chosen that provide good prediction but are far from the "true" disparity. These problems can be more serious if the cameras are sparsely located. In this paper, in order to achieve disparity vector prediction, and thus improve the quality of the disparity field and the coding efficiency, we do not use neighboring disparity vectors. Instead we compute an initial correspondence between frames by locating corner points (which are usually found along object edges or region boundaries) and identifying matches between the corner points in each image. This information allows us to improve the prediction and correct potential prediction errors due to poor block matching.

In this paper techniques for both IC and improved disparity prediction are proposed and tested within the state of the art H.264/AVC codec. Our results show up to 1dB coding gains for the anchor frames (i.e., frames for which interview coding is used). In Section II, we describe our proposed IC technique, while in Section III our MS approach. In Section IV, we provide experimental results based on H.264/AVC, and Section V summarizes and concludes the paper.

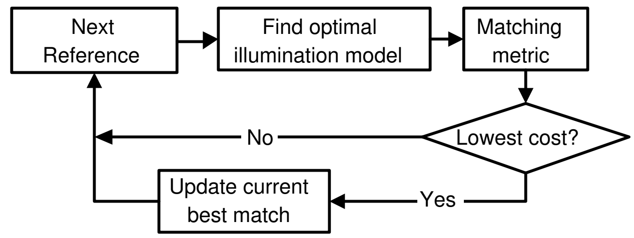


Fig. 1. Modified Search Loop for the given current block

II. ILLUMINATION COMPENSATION

Illumination mismatch due to imperfectly calibrated cameras, different perspective projection directions, and different reflection effects varies from object to object. We consider both the global mismatch typically due to lack of camera calibration, as well as local mismatches that arise due differences in camera positioning. We first consider the influence of DE matching metrics on the illumination mismatch problem.

A. Need for Modified Matching Metrics

Video encoders based on block-matching methods use error metrics, such as Sum of Absolute Differences (SAD) or Sum of Squared Differences (SSD), between reference and current blocks to evaluate the quality of the match. These metrics take into account the total energy of the residue block but do not provide direct information about the relative coding efficiency achieved by coding different residue blocks. Clearly two residue blocks with same energy will be coded with two different rates, depending on their exact energy contents. Thus, it is possible that the best match in terms of these metrics may contain frequency patterns that lead to lower coding efficiency than another block that had higher residual energy. Illumination mismatches are typical situations where this effect can be observed. While other techniques mentioned in the introduction introduce IC only after a best match block has been found, in our work we use modified matching metric which considers an optimal illumination compensation.

B. Multiview One-Step Affine Illumination Compensation (MOSAIC)

For the given current block, we look for the best matching block within a predetermined search range in the reference image using a distance measure that incorporates an illumination adjustment between reference and current block before deciding the best match.

As shown in Fig. 1, when considering the match between the i-th block in the current image and a candidate block j in the search window of reference image an optimal IC model is computed for the pair (i, j). Using this IC model, Sum of Absolute Difference After Compensation (SADAC) between two blocks is found and then among all candidates within search range, the block with the minimum SADAC is selected as a match. Obviously, after the best match is identified, the corresponding optimal IC model is encoded and sent to the decoder.

The first order affine IC model for current block i is composed of an offset parameter C^i and a scale parameter S^i:

$$\Psi^i = \{S^i, C^i\} \quad (1)$$

A candidate reference block signal $B_R(x,y)$ for the current block i can be decomposed into the sum of its mean μ_R^i and a zero mean signal, ω_R^i: $B_R^i(x,y) = \mu_R^i + \omega_R^i(x,y)$. Then the illumination compensated reference block signal $\widetilde{B}_R^i(x,y)$ with IC model Ψ^i is:

$$\widetilde{B}_R^i(x,y) = [\mu_R^i + C^i] + S^i \cdot \omega_R^i(x,y) \quad (2)$$

In (2), C^i modifies the mean and S^i modifies the variation of reference block. In what follows to simplify the notation we omit the block index i on the variables.

Our goal is to find the IC parameters that minimize SSDAC defined as

$$SSDAC \equiv \Sigma_{\forall(x,y)} |B_C(x,y) - B_R(x,y)|^2, \quad (3)$$

where B_R is a candidate matching block for B_C and (x,y) represent coordinates in each block. The optimal IC parameters in terms of minimizing (3) are:

$$S = \frac{\sigma_{CR}^2}{\sigma_{RR}^2}, \quad C = \mu_C - \mu_R \quad (4)$$

with

$$\sigma_{AB}{}^2 = \frac{1}{N} \sum_{\forall(x,y)} [(B_A(x,y) - \mu_A) \cdot (B_B(x,y) - \mu_B)] \quad (5)$$

with $A, B \in \{C, R\}$ and with N being the number of pixels in the block.

While the additive parameter, C, directly removes constant offset mismatches, the multiplicative parameter compensates zero-mean variations (AC coefficients) according to the block statistics, i.e., the lower the cross-correlation between current and reference blocks, the smaller the scale parameter. Therefore if patterns of current and reference blocks are completely different, no scale parameter is applied to the modified reference block.

Among all candidates within search range, the reference block $\widetilde{B}_R$ minimizing SADAC with IC parameters is selected as the best match and the minimum SSDAC is given as follows,

$$\widetilde{SSDAC} = N \cdot (\sigma_{CC}^2 - \frac{\sigma_{CR}^4}{\sigma_{RR}^2}) = N \cdot \sigma_{CC}^2 \cdot (1 - \rho^2), \quad (6)$$

where ρ is the correlation coefficient between B_R and B_C. As can be seen in the above equation, the proposed technique finds the best reference block in the sense of maximum correlation with the current block so that the patterns of the two blocks are well-matched, and adjusts parameters to minimize the residual energy.

C. Illumination Compensation Model Coding

As stated earlier, once a model is used for compensating a block, it must be sent to the decoder and thus we need to consider approaches to encode these model parameters. We have observed experimentally that the two parameters tend to be statistically independent. On the other hand, there exists some correlation between IC parameters in neighboring blocks. This redundancy appears more clearly in the offset parameter case. Intuitively this is reasonable, as we would expect the means of successive blocks to exhibit more correlation than their distribution of energy across frequencies. Based on these observations, we encode separately the two parameters using the same encoding technique which involves three steps (i) prediction, (ii) uniform quantization, and (iii) adaptive binary arithmetic coding.

In the prediction step spatial redundancy between the models for neighboring blocks is removed. We use a simple difference between parameters of consecutive blocks following a left to right and top to bottom scan order. The differences between parameters can be modeled by a symmetric probability density function (pdf) with a peak at zero. Then, these differential parameters are uniformly quantized and entropy-encoded with a binary arithmetic coder after performing binarization of the quantization indices using a unary code.

Clearly, different blocks suffer from different levels of illumination mismatch. Thus we allow the encoder to decide whether or not the MOSAIC technique is used on a block by block basis. This is achieved by computing the R-D values associated to coding each block with and without IC, and then letting the standard R-D optimization tools in the H.264/AVC codec make an R-D optimal decision. There is an added overhead needed to indicate for each block whether IC is used but this is more efficient overall than sending IC parameters for all blocks, including those blocks for which IC does not provide significant gains.

III. MODIFIED SEARCH

Recall that median predictors were popular in the context of motion prediction. Assume the same type of predictor is used for DE and refer to the example of Fig. 2. In this example we can see that block 2 will use a predictor based on the disparity estimated for the background (i.e., block 1), which, due to the lack of texture in the background, could indicate the location far from the true one (i.e., block a). In this simple example, the best match for block 1 provides a very poor prediction for block 2. This prediction error may also be propagated to the following blocks (3 and 4). Our proposed IC finds the best match independently for each block and thus cannot take into account whether

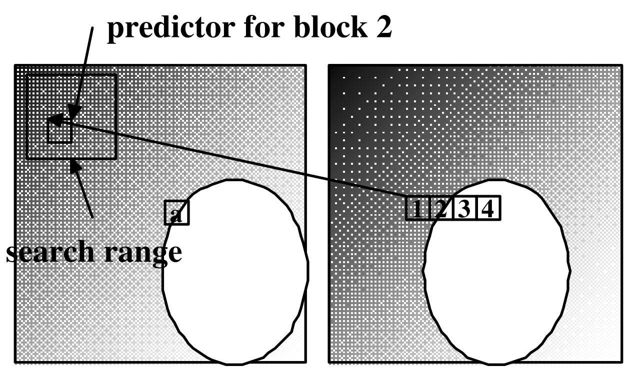


Fig. 2. Example of incorrect predictor

TABLE I
H.264 ENCODING CONDITION

Feature / Tool / Setting	AVC Parameters
Rate control	No
RD optimization	Yes
Specific settings	Loop filter, CABAC
Search range	±32
# Reference picture	1
GOP Structure	IPPP
Block size for inter coding	8*8 only
Resolution	Half-Pixel

the vectors obtained lead to true disparity. Conversely, our proposed MS approach uses geometric information between the two views in order to estimate a more exact disparity predictor to be used in DE. By combining IC and MS we achieve significant gains as shown in the results. In what follows, a point correspondence finding procedure and a modified search algorithm are proposed.

A. Point Correspondence Finding Procedure

We start by finding corner points using functions from the opencv library [5]. With the corner points found in the reference and current images, feature matching by singular value decomposition (SVD) [12] is used to find point correspondences between the two images. Because incorrect matching points can have a significant negative effect, they should be removed before disparity search. In this paper we adopted following two filters for the matching point set $S = \{((x,y),(x',y')) : (x,y) \in R, (x',y') \in C\}$. Here, R is a reference image and C is a current image.

1) Removal of outliers
   a) Calculate distance d of all matching point pairs (x,y) and (x',y') in the set S.
   b) Calculate mean μ and variance σ^2 of the distance.
   c) Remove points with distance greater than $\mu + a \cdot \sigma$ or smaller than $\mu - a \cdot \sigma$.
   d) Iterate step-c) until there are no outliers.
2) Correlation matching
   Let two $(2w+1) \cdot (2w+1)$ areas centered on corresponding points in reference and current image be $R(x,y)$ and $C(x,y)$.
   a) Calculate normalized correlation which is given by

$$NC = \frac{SS_{RC}^2}{SS_{RR} \cdot SS_{CC}}, \tag{7}$$

where

$$SS_{AB}^2 = \sum_{\forall(x,y)} (A(x,y) - \mu_A) \cdot (B(x,y) - \mu_B) \tag{8}$$

with μ_A and μ_B, the averages of $A(x,y)$ and $B(x,y)$ for the given areas and $A, B \in \{R, C\}$.
   b) If $NC < th$, remove those matching point.

In this paper, we used $a = 3$, $th = 0.7$ and $w = 10$. That is, 6-σ accuracy is used to remove outliers, and threshold and correlation window size is decided empirically. Order of filtering is 1), 2), and 1) again.

B. Modified Search Algorithm

1) Find matching point set according to point correspondence finding procedure.
2) Start block disparity vector search.
3) In the given block or neighboring blocks‡,
   → If matching points exist, use the averaged difference of those points as a predictor.
   → If a matching point does not exist, use the predictor given as a median of disparity vectors from neighboring blocks.

Neighboring blocks‡ used in this paper are 8 (8*8) blocks surrounding current block.

IV. RESULTS

A. Illumination Compensation

For the interview coding of the test sequence, KDDI object 3, we used H.264/AVC reference codec, JM7.6 [1]. Encoding conditions are shown in Table I. As shown in Fig. 3, about 0.5-1.0 dB gains in a medium bit-rates range have been achieved by MOSAIC with ON/OFF mode (MOSAIC ON/OFF) compared to simple interview coding (SIC). QP in Fig. 3, 4, and 5 (QP_{mod}) is the parameter used in IC parameter quantization. Large QP_{mod} means the finer quantizer has been used.

As can be seen in Fig. 4, in low bit rate, gains from MOSAIC are overwhelmed by the overhead for IC parameters. By selecting blocks which need IC, this overhead

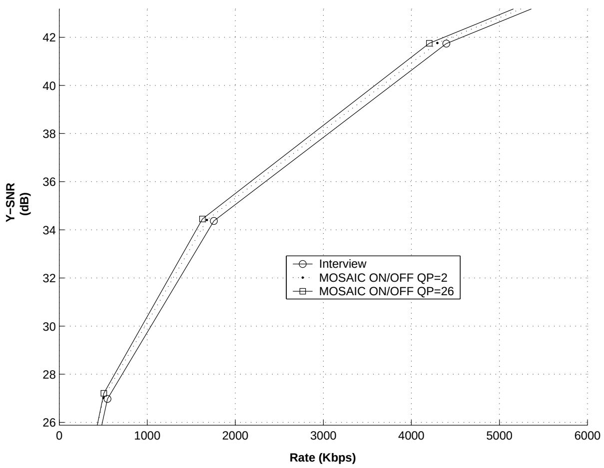


Fig. 3. Interview vs. MOSAIC ON/OFF

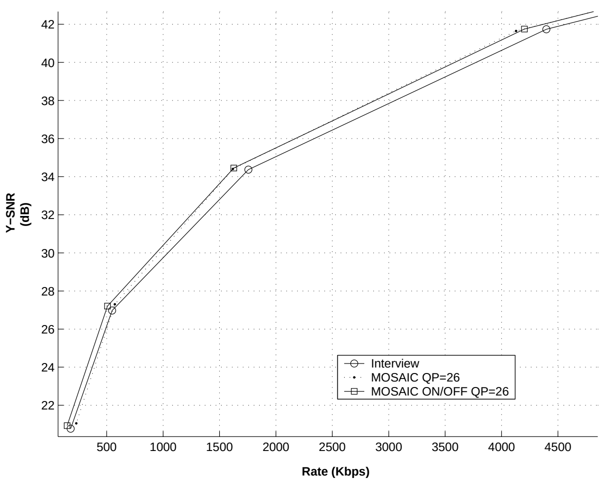


Fig. 4. Interview vs. MOSAIC vs. MOSAIC ON/OFF

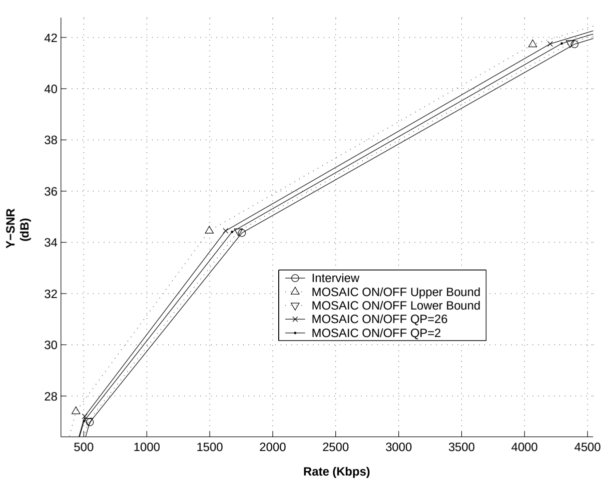


Fig. 5. Interview vs. MOSAIC with Bounds

TABLE II
THREE TEST SEQUENCES

Sequence	Size	View interval	QP used
ST	640* 480	Sparse	44, 40, 34
KDDI (object 1)	320* 240	Dense	38, 36, 33
AQ (aqua)	320* 240	Dense	40, 36, 32

has been reduced and 0.2-0.5dB gains over SIC have been achieved even in the low bit rate.

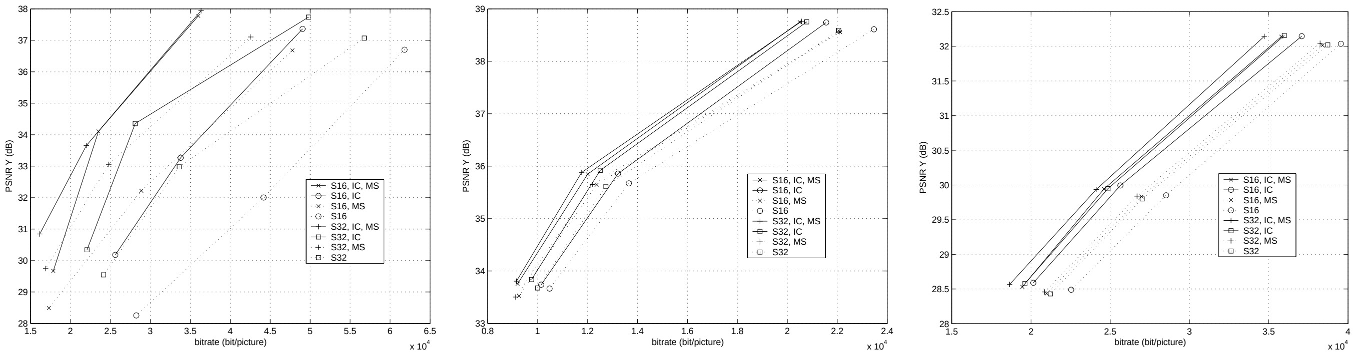
In Fig. 5, an upper and a lower bound are shown for MOSAIC ON/OFF. The upper bound is ideally achieved when zero bits are used to encode the IC models and no quantization error is introduced. It is observed that best performance by MOSAIC ON/OFF is in the middle of the baseline (SIC) and the upper bound. The lower bound has been found by using IC models for the search but not using them for the compensation and the reconstruction, not to send IC parameters to the decoder.

The virtues of MOSAIC can be summarized as follows: First, searching has been improved by considering illumination mismatch. In Fig. 5, MOSAIC ON/OFF lower bound corresponds to this situation. Second, IC parameters find the best match in the sense that two patterns of blocks are similar and SSDAC is the minimum. The similarity of two blocks makes high frequency components of residue small and results in the coding efficiency or the better R-D performance.

B. Modified Search

In the test of MS, the same conditions given in Table I have been used except that additional search range ±16 also has been tested. Table II explains three test sequences used in simulation. Interview coding results for first 4 views have been averaged in Fig 6. In ST, because of smooth background, large disparity (about 30 pixel), and illumination mismatch between views, MS with intra mode disabled achieves significant gain. With intra mode disabled, every block should be encoded by disparity search and therefore, propagated predictor error affects the gain significantly than the case of intra mode enabled. Therefore, with intra mode disabled, it can be observed very clearly that MS corrects wrong predictors and helps search to find the optimal match. As can be seen in Fig. 6(a), the gain achieved by "MS and IC with search range (SR) 16" is up to 4dB compared to "IC with SR 16". As can be seen in Fig. 6(b) these gains are reduced to 0.5dB-1dB when intra mode is enabled because for the blocks with large errors by inter prediction, intra prediction is used selectively.

As shown in Fig. 6(c), for KDDI where the disparity is smaller and background is more complex than ST, about



(a) ST: Intra OFF (b) ST: Intra ON (c) KDDI: Intra ON

Fig. 6. MS results

0.3 dB gain has been achieved in both SR ± 16 and SR ± 32 for MS and IC, compared to IC only. In AQ sequence, due to very small disparity (less than 5 pixel) and texture in background, the result with MS and intra enabled is similar to the one without MS.

Note that with intra enabled, "MS and IC with SR 16" (option A) gives better (ST and KDDI) or similar (AQ) result than "IC only with SR 32" (option B) in PSNR. Based on the running time measure provided in H.264/AVC reference codec, in KDDI and AQ, option A is faster than option B by 50%. Also in ST, option A is faster by up to 30% depending on QP.

Even though MS shows efficiency in both PSNR and run time when IC is used, there exists an additional complexity in feature point finding and SVD-based point matching algorithm. For example, if there are n candidate feature points in the reference and the current images, basic complexity of SVD is $O(n^3)$. Also the complexity for feature detection and filters should be considered. This increased complexity by MS, which depends on the size of image, the number of feature points, etc, is about 10-20% for ST with SR ± 32 and IC not applied.

V. CONCLUSION

New block-based disparity compensation techniques that are robust to illumination mismatches between views have been proposed. Our simulation results with H.264/AVC show that MOSAIC with ON/OFF mode provides substantial gains. The potential gain depends on the level of illumination mismatch present in the sequence. MS uses geometric information between views in finding the best matching block by introducing a modified predictor. This predictor can also correct the propagation of DE estimation errors, which can be severe under illumination variations. MS performance is not very sensitive to the use of reduced search ranges, but comes at the cost of requiring a preliminary step to identify key image points and match them to points in other images. Further research includes investigating adaptively changing the size of the search region in order to enable lower overall complexity encoding.

REFERENCES

[1] H.264/AVC software coordination. http://bs.hhi.de/~suehring/tml/.

[2] M. Accame, F.G.B. De Natale, and D.D. Giusto. Hierarchical block matching for disparity estimation in stereo sequences. In *Proc. of IEEE International Conference on Image Processing (ICIP)*, volume 2, pages 23–26, Washington, DC, Oct. 1995.

[3] H. Aydinoglu and M. H. Hayes. Compression of multi-view images. In *Proc. of IEEE International Conference on Image Processing (ICIP)*, volume 2, pages 385–389, Austin, TX, Nov. 1994.

[4] D.R. Clewer, L.J. Luo, C.N. Canagarajah, D.R. Bull, and M.H. Barton. Efficient multiview image compression using quadtree disparity estimation. In *Proc. of ISCAS '01*, 2001.

[5] Intel corporation, Santa Clara, CA. *Open source computer vision library - reference manual*, 2001. http://www.intel.com/research/mrl/research/opencv/.

[6] Kazuto Kamikura, Hiroshi Watanabe, Hirohisa Jozawa, Hiroshi Kotera, and Susumu Ichinose. Global brightness-variation compensation for video coding. *IEEE Trans. Circuits and Systems for Video Technology*, 8(8), Dec. 1998.

[7] Chaitanya Kamisetty and C. V. Jawahar. Multiview image compression using algebraic constraints. In *Proc. of Conf. on Convergent Technologies for Asia-Pacific Region, TENCON 2003*, volume 3, pages 927–931, Oct. 2003.

[8] Sang Hyun Kim and Rae-Hong Park. Fast local motion-compensation algorithm for video sequences with brightness variations. *IEEE Trans. Circuits and Systems for Video Technology*, 13(4), Apr. 2003.

[9] Michael E. Lukacs. Predictive coding of multi-viewpoint image sets. In *Proc. ICASSP'86*, pages 521–524, Oct. 1986.

[10] Wolfgang Niehsen and Stephan F. Simon. Block motion estimation using orthogonal projection. In *Proc. of IEEE International Conference on Image Processing (ICIP)*, Lausanne, Switzerland, Sept. 1996.

[11] J.-R. Ohm. Encoding and reconstruction of multiview video objects. *IEEE Signal Processing Magazine*, 16(3):47–54, May 1999.

[12] Maurizio Pilu. Uncalibrated stereo correspondence by singular value decomposition. Technical Report Tech. Rep. HPL-97-96, HP Labs, Palo Alto, CA, 1997.

[13] Rahul Shukla, Hayder Radha, and Martin Vetterli. Disparity dependent segmentation based stereo image coding. In *Proc. of IEEE International Conference on Image Processing (ICIP)*, Barcelona, Spain, Sept. 2003.