

Enhancing the Error Detection Capabilities of the Standard Video Decoder using Pixel Domain Dissimilarity Metrics

Reuben A. Farrugia and Carl J. Debono

Department of Communications and Computer Engineering, University of Malta, Msida MSD 2080, Malta

Email: rrfarr@eng.um.edu.mt; cjdebo@eng.um.edu.mt

Abstract—The video compression standards commonly adopted in wireless multimedia services utilize variable length codes (VLC) in order to attain high compression ratios. While providing the high data rates required, this technique makes the system more susceptible to transmission errors. Thus the end-to-end quality of the video stream transmitted over an error-prone channel depends on the detection, and concealment of the corrupted macroblocks. The error detection capability of standard decoders is quite limited, for example, in the case of the H.263+ codec around 40.54% of the corrupted macroblocks are undetected, placing a bound on the perceived quality of the reconstructed video sequence.

This paper presents a novel solution using eight pixel domain dissimilarity metrics computed in the CIE LUV color space which can be used at decode time to improve the error detection rate of the standard decoder. The spatial dissimilarity metric has been found to perform the best with an average increase in error detection rate of 60.38% when compared to the standard decoder (about 20% more than other published results) with 0% of false detection and a gain in peak signal-to-noise ratio (PSNR) of 3.94 dB.

Keywords— Error detection coding, Quality assurance, Video codecs, Video signal processing.

I. INTRODUCTION

Wireless multimedia services, such as Digital TV broadcasting, video-telephony, and videoconferencing applications, have become increasingly popular in recent years. Most international standards for video compression, such as H.261 [1], H.263+ [2], MPEG-1 [3], MPEG-2 [4] and MPEG-4 [5] rely on inter-frame predictive coding and block based discrete cosine transform (DCT) in order to exploit both temporal and spatial redundancy present in the video sequence. Variable length codes (VLC) are then used to provide the required compression rates.

The resulting compressed video sequence is more vulnerable to transmission errors, where a single corrupted bit will cause loss in synchronization between the encoder and decoder until the next synchronization marker, which is placed in every group of block (GOB) and Picture header. This results in a number of corrupted macroblocks (MBs) causing annoying visual artifacts which propagate both in the temporal and spatial domains.

In order to improve the video quality transmitted over a noisy wireless channel, several concealment algorithms were proposed in literature [6] – [8]. For these algorithms to operate appropriately, accurate detection and localiza-

tion of the transmission errors is required. Most error concealment algorithms employ syntax and semantic based error detection capabilities provided by the standard decoder. In [9] it was revealed that the error detection capabilities of the standard decoder are quite limited. Taking the case of the H.263+ decoder as an example, this decoder does not manage to detect, on average, 40.54% of the corrupted macroblocks.

Another bit-level error detection approach was suggested in [10], where a reversible VLC (RVLC) was proposed. This allows the compressed stream to be decoded in both forward and backward directions to maximize the error detection capabilities of the decoder. The performance of the RVLC is limited by the fact that not all errors provide syntax and semantic violations. Moreover, it does not manage to detect the exact location of the corrupted macroblocks. This method has also the disadvantage of wasting resources as some errors have no impact on the quality of the video, and therefore it is useless trying to correct them [11]. For the latter reason, error detection at picture level was considered a better solution and used in this work.

Throughout the implementation of this work it was observed that error detection can be enhanced by considering the following two properties:

- (1) Spatial Consistency – The boundaries across a corrupted macroblock will be irregular;
- (2) Temporal Consistency – A corrupted macroblock should be temporally significantly different than the motion compensated macroblock in the previous frame.

The spatial consistency property was investigated in [11] – [16]. The dissimilarity metrics defined in these publications only manage to detect 20% - 40% of the corrupted macroblocks. Moreover, these approaches result in a number of false detections which will reduce the quality of the received video even when operating under an error free condition.

An iterative error detection and concealment approach was presented in [11]. Although this algorithm achieves a substantial gain in peak signal-to-noise ratio (PSNR), this is achieved with a significant increase in computational time required by the decoder which makes it unsuitable for real-time applications.

The error detection algorithms used in [11 – 16] are performed in the YUV color space, which is not a uniformly perceptual color model. Moreover, these features require the whole frame to be decoded resulting in additional delays in the system.

The temporal consistency property was only exploited in [17]. This algorithm achieves an average gain in PSNR of 3.32 dB at a bit error rate (BER) of $1.30E-3$. This dissimilarity metric was only tested on videoconferencing like sequences, where the performance of the algorithm is relatively good. However, the proposed algorithm is significantly dependent on movement and the false detection rate will increase significantly when fast moving objects or camera movements are present in the video sequence.

In order to reduce computational complexity, an error detection algorithm for DCT based codecs in the compressed domain was proposed in [18]. However, the authors have tested the algorithm on just one video sequence and therefore the algorithm may be sequence dependent.

In this paper, a novel approach based on eight dissimilarity metrics, which exploits both the temporal and spatial consistency properties, is proposed. These dissimilarity metrics were computed in the CIE LUV color space, since it was found to perform best when compared to other color models. The CIE LUV color space model is perceptually uniform and is very popular in the television and video displays industry [19].

The major advantage of the proposed dissimilarity metrics is that error detection can be performed at decode time, and therefore no additional delay is introduced in the system. It was found that the spatial dissimilarity metric performs best with an increase in average error detection of 60.38% when compared to the standard decoder, with a gain in PSNR of 3.94 dB, and a false detection rate of 0%. These results outperform other results found in literature.

This paper is organized as follows. An overview of the distortion caused by channel bit errors is presented in Section II, followed by a discussion on the importance of the color space model in solving error detection problems. In Section IV the dissimilarity metrics are described in some detail. The simulation results are delivered in Section V while the final comments and conclusion are provided in Section VI.

II. DISTORTION CAUSED BY TRANSMISSION ERRORS

Most video coding standards support two coding modes, INTRA and INTER modes. INTRA coded macroblocks are spatially encoded using the DCT transform, quantized, zigzag pattern encoded and VLC coded. When operating in INTER mode, the motion estimation process is used to derive the motion vector. The difference between the prediction and current macroblock is spatially encoded, quantized, zigzag-pattern encoded and VLC encoded.

While providing good compression, VLC data suffers from the fact that, when transmission errors occur, the incoming bitstream can still form valid entries in the VLC table, and the decoder may not detect the occurrence of an error [15]. In such situations, the decoder will continue decoding the unsynchronized sequence until a syntax or semantic violation is encountered.

Fig. 1 shows a decoded frame of the Akiyo sequence with distortions caused by channel errors. Two main types of distortions, which are typically observed in corrupted video sequences, are shown in this figure. These are:

1. Type I – A corrupted 16 x 16 pixel macroblock whose color and texture are significantly

irregular in both the spatial and the temporal domain.

2. Type II – A single corrupted 8 x 8 pixel block whose color and texture is irregular in both the spatial and the temporal domain.

Type I errors are usually caused by either a corrupted DC coefficient of the chrominance components or else a corrupted VLC codeword which is mapped to a valid but incorrect run-level pair. Type II errors are usually caused by a corrupted luminance DC coefficient.

Both Type I and Type II corrupted macroblocks will be present in the decoded sequence if the error detection capabilities of the decoder are not enhanced. From Fig. 1, it can be observed that these corrupted blocks are highly perceivable and contribute significantly to the reduction in quality perceived by the end-user. Eliminating these distortions will significantly improve the quality of the decoded video sequence. To successfully conceal these impairments the corrupted macroblock must be accurately detected and localized by the error detection algorithm.



Fig. 1 A decoded frame of the Akiyo sequence representing the two main distortions caused by channel errors.

III. COLOR SPACE SELECTION

It is evident from the previous section that most of the transmission errors result in corrupted blocks whose color does not fit in both space and time. Thus, color plays an important role in error detection algorithms and therefore, selecting the appropriate color space model (where the dissimilarity metrics will be computed) becomes of fundamental importance.

A number of color spaces were considered in this work including the YUV, RGB, HSV, CIE LAB and CIE LUV. The simulation results presented later in this work show that the CIE LAB and CIE LUV color spaces outperform the other color spaces. This is because, YUV, RGB, and HSV color spaces are perceptually non-linear, i.e. equal distances in the space do not in general correspond to perceptually equal color sensations [19]. Moreover, it was found that the CIE LUV color model performs slightly better than the CIE LAB color space, and therefore the proposed dissimilarity metrics were computed in the CIE LUV color space.

IV. DISSIMILARITY METRICS

A. Spatial Feature Extraction

1) Average Intersample Difference across Boundaries (AIDB)

In natural video sequences, the pixels generally vary smoothly across boundaries, even in the presence of edges [11]. In order to exploit this property, the AIDB dissimilarity metric was used. The technique we employed is illustrated in Fig. 2, where only the upper ($X_{mb,gob-1}$) and left ($X_{mb-1,gob}$) macroblocks were considered, thus utilizing half the information used in [11], [12]. This allows this dissimilarity metric to be computed at decoding time. The AIDB is capable of detecting most of the Type I errors but Type II errors usually remain undetected.

The AIDB of a $K \times K$ dimensional macroblock $X_{mb,gob}$ is given by:

$$d_{AIDB} = \frac{1}{2K} \sum_{i=0}^{K-1} \left\{ \|x_{i,0}^{mb,gob} - x_{i,K-1}^{mb-1,gob}\| + \|x_{0,i}^{mb,gob} - x_{K-1,i}^{mb,gob-1}\| \right\} \quad (1)$$

where $\|\bullet\|$ represents the Euclidean distance in CIE LUV color space.

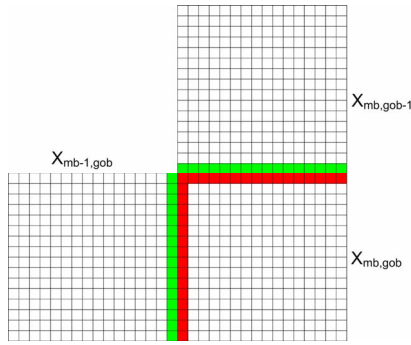


Fig. 2 Boundary pixels used to derive the AIDB dissimilarity metric

2) Internal Average Intersample Difference across Boundaries (IAIDB)

Similar to the AIDB, the IAIDB is based on the fact that for an uncorrupted macroblock the transition between pixels at the boundary of the blocks is smooth. On the other hand, if a block is corrupted there will be a significant difference between adjacent blocks. As illustrated in Fig. 3, the IAIDB computes the sum of Euclidean distances between boundary block pixels in CIE LUV color space.

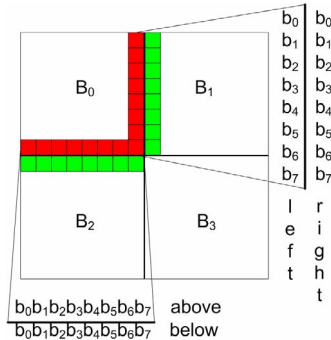


Fig. 3 Block boundary pixels used to derive the IAIDB dissimilarity metric for block B0

The IAIDB metric for each block b is computed using:

$$d_{IAIDB}^b = \frac{1}{K} \sum_{i=0}^{K/2-1} \left\{ \|b_{i,b}^{above} - b_{i,b}^{below}\| + \|b_{i,b}^{left} - b_{i,b}^{right}\| \right\} \quad (2)$$

Therefore the array d_{IAIDB}^b contains the IAIDB dissimilarity metric of each block. The resulting metric is derived using the equation:

$$d_{IAIDB} = \max(d_{IAIDB}^b) - \max_2(d_{IAIDB}^b) \quad (3)$$

where $\max(\bullet)$ provides the maximum value and $\max_2(\bullet)$ represents the second maximum value from the array. This metric can be used to detect Type II errors.

3) Spatial Dissimilarity Metric

The spatial dissimilarity metric combines the AIDB and IAIDB in order to detect both Type I and Type II errors. This is achieved through:

$$d_s = d_{AIDB} + d_{IAIDB} \quad (4)$$

B. Temporal Feature Extraction

1) Sum of Euclidean Pixel Difference (SED)

The SED dissimilarity metric is based on the fact that for an uncorrupted macroblock the pixel transition of a macroblock and the corresponding macroblock in the previous frame varies smoothly. This dissimilarity metric is adopted to detect Type I errors and is computed by deriving the total Euclidean difference between the pixels contained in a macroblock X_t and the motion compensated $K \times K$ dimensional block X_{t-1} . This is formally given by:

$$d_{SED} = \frac{1}{K \times K} \sum_{i=0}^{K-1} \sum_{j=0}^{K-1} \|x_t(i, j) - x_{t-1}(i, j)\| \quad (5)$$

where i and j are the pixel coordinates of a $K \times K$ dimensional block.

2) Chi-Square Histogram Difference Test (χ^2 HDT)

Another temporal feature which can be used to detect Type I errors is the χ^2 HDT which is based on the same concept of the SED dissimilarity metric. In order to detect whether a macroblock is corrupted or not, a 3-D color histogram in CIE LUV color space was set for each macroblock X_t and the corresponding $K \times K$ dimensional motion compensated macroblock X_{t-1} in the previous frame. Since the CIE LUV color model is perceptually uniform, the bins can be selected uniformly. Through simulation it was found that 25 bins provide the best compromise between error detection and false detection rate.

From the simulation results, it was found that the modified version of the Chi-Square Test [20] performs better than the traditional Chi-Square Test. The χ^2 HDT dissimilarity metric is evaluated as follows:

$$d_{\chi^2HDT} = \sum_{b=1}^B \left\| \frac{(x_{t,b} - x_{t-1,b})^2}{\max(x_{t,b}, x_{t-1,b})} \right\| \quad (6)$$

where B is the total number of bins.

3) Internal Sum of Euclidean Difference (ISED)

In order to enhance the error detection capabilities of Type II errors, the ISED algorithm, which is based on the

same concept as the IAIDB, can be used. However, instead of evaluating the difference of each block across boundaries, the total Euclidean difference between pixels of each block within macroblock X_t and the corresponding block within macroblock X_{t-1} is evaluated. This can be done using the following equation for each block b :

$$d_{ISED}^b = \frac{4}{K \times K} \sum_{i=0}^{K/2-1} \sum_{j=0}^{K/2-1} \|x_{t,b}(i,j) - x_{t-1,b}(i,j)\| \quad (7)$$

The ISED dissimilarity metric of each block is contained within d_{ISED}^b . The resulting ISED dissimilarity metric, which will be used for error detection of Type II errors, is computed as follows:

$$d_{ISED} = \max(d_{ISED}^b) - \max_2(d_{ISED}^b) \quad (8)$$

where $\max(\bullet)$ and $\max_2(\bullet)$ represent the maximum and the second maximum value from the array respectively.

4) Temporal Dissimilarity Metric

The temporal dissimilarity metric combines the effect of both the SED and the ISED dissimilarity metric together, and is given by:

$$d_T = d_{SED} + d_{ISED} \quad (9)$$

C. Spatio-Temporal Feature Extraction

The combined effect of the spatial and temporal dissimilarity metric is termed as being the spatio-temporal dissimilarity metric and is derived as:

$$d_{S-T} = d_{spatial} + d_{temporal} \quad (10)$$

V. SIMULATION RESULTS

A. Color Space Selection

The aim of the dissimilarity metrics described in the previous section is to provide small metrics for uncorrupted macroblocks and large metrics for corrupted macroblocks. However, as illustrated in Fig. 4, these classes usually overlap and therefore a macroblock cannot be easily classified as being corrupted or not.

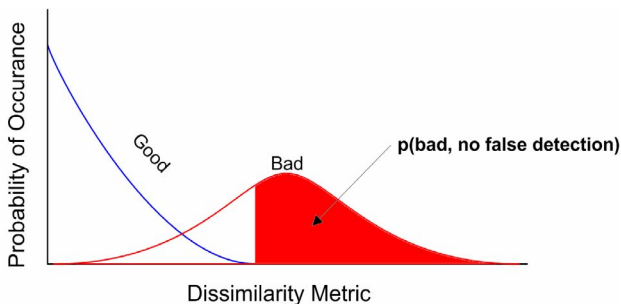


Fig. 4 Probability density function of Good (uncorrupted MBs) and Bad (corrupted MBs)

To identify the color model where the above mentioned metrics will be computed, the joint probability of detecting a corrupted macroblock without having false detections, $P(\text{bad, no false detection})$, should be maximized. The SED dissimilarity metric was computed for different

test video sequences (“Erik”, “Silent”, “Akiyo”, “News” and “Football”) at different color spaces and the results obtained are summarized in Table I. Similar values were achieved when computing the other dissimilarity metrics.

TABLE I. COLOR SPACE SEPARATION TEST

Color Space	$p(\text{bad, no false detection})$
CIE LAB	70.37%
CIE LUV	71.54%
RGB	15.79%
YUV	22.52%
HSV	67.75%

These results clearly show that the dissimilarity metrics perform better when operating in a perceptually uniform color space, such as the CIE LAB and CIE LUV color models. Moreover, performance in the CIE LUV color space was best and therefore the proposed dissimilarity metrics were computed in this color space.

B. Error Detection Algorithm Testing

The performance of the proposed dissimilarity metrics was evaluated by calculating the probability of successfully detecting a corrupted macroblock without having false detections. The video test sequences were analyzed using this algorithm and the results obtained are summarized in Table II. From these results it can be immediately seen that the combined spatial and temporal dissimilarity metrics perform better than the individual metrics. Moreover, the spatial, temporal and spatio-temporal dissimilarity metrics manage to separate the classes quite effectively.

TABLE II. DISSIMILARITY METRIC CLASS SEPARATION TEST

Diss. Metric	$p(\text{bad, no false detection})$
d_{AIDB}	68.98%
d_{IAIDB}	32.65%
d_{SED}	60.50%
d_{ISED}	15.15%
d_{2HDT}	25.84%
d_S	73.01%
d_T	63.53%
d_{S-T}	74.12%

The video test sequences stored at CIF resolution were compressed by an H.263+ Decoder. The resulting compressed bitstream was corrupted with a BER of $1.30E-3$. If a syntax or semantic error is detected by the H.263+ Decoder, all the macroblocks which follow the detected error up to the next synchronization marker are concealed. Otherwise, the spatial, temporal and spatio-temporal dissimilarity metrics are computed. If the dissimilarity metric is superior to a defined global threshold, an error is flagged to the decoder and the following macroblocks are concealed.

The global thresholds (T) employed in the simulation were derived heuristically in order to maximize the probability of good error detection ($P(ED)$) with zero false detection rate ($P(FD)$). The error concealment adopted by the system replaces the macroblocks with the motion compensated macroblock from the previous frame whose motion vector is the median of the neighboring macroblocks.

Table III summarizes the performance of the spatial, temporal and spatio-temporal dissimilarity metrics. In order to derive more accurately the false detection rate, a larger collection of video sequences consisting of a total of 2500 frames all encoded in CIF resolution was considered.

TABLE III. ERROR DETECTION CAPABILITIES

Diss. Metric	T	P(ED)	p(FD)	PSNR Gain
d_s	118.6123	60.38%	0.00%	3.94dB
d_T	238.5880	18.99%	0.00%	1.58dB
$d_{s,T}$	302.0270	49.29%	0.00%	3.41dB

From these results it was found that the spatial and spatio-temporal dissimilarity metrics outperform the performance of other error-detection algorithms present in literature. Applying these dissimilarity metrics in the CIE LUV color space has significantly increased the error detection capabilities of the H.263+ decoder.

The temporal and spatio-temporal dissimilarity metrics have performed well on videoconferencing like sequences such as “Erik”, “Silent” and “Akiyo”. However, their performance dropped in the presence of shots such as in the “News” sequence and fast moving objects such as in the “Football” sequence.

The spatial dissimilarity metric utilizes only the spatial information and therefore its performance does not degrade in the presence of shots or fast moving objects. Therefore, the spatial dissimilarity metric is the most robust and manages to detect on average 60.38% of the corrupted macroblocks which were not detected by the standard decoder, with a gain in PSNR of 3.94 dB. The gain in quality can be also evaluated subjectively from Fig. 5. Moreover, the false detection rate is 0% and therefore the quality of the uncorrupted video sequences will not drop.



Fig. 5 Frame 8 of a corrupted Akiyo Sequence using (a) the standard H.263+ Decoder (b) the enhanced H.263+ Decoder with the spatial dissimilarity metric.

VI. COMMENTS AND CONCLUSION

This paper presents a novel algorithm based on eight different dissimilarity metrics which can be employed to enhance the error detection capabilities of the standard video decoder. These metrics were tested on a wide range of video sequences including videoconferencing like sequences and sports sequences. From these tests it emerged that computing these metrics in a perceptually uniform color space results in a significant gain in performance.

The spatial dissimilarity metric has performed the best overall, and it outperforms other metrics published in literature by providing around 20% increase in the error detection rate. On the other hand, the spatio-temporal dissimilarity metric performs well when no shots or fast moving objects are present in the sequence. The results presented in this paper refer to the H.263+ codec, however, similar results are expected when other standard video codecs are adopted.

REFERENCES

- [1] ITU-T Rec. H.261, “Video Codec for audiovisual services at p x 64 bit/s,” 1993.
- [2] ITU-T Rec. H.263, “Video Coding for Low Bit-Rate Communication,” 2005.
- [3] ISO/IEC 11172, “Information technology – Coding of moving pictures and associated audio for digital storage media at up to 1.5 Mbit/s,” 1993.
- [4] ISO/IEC 13818, “Information technology – Generic coding of moving pictures and associated audio information,” 2000.
- [5] ISO/IEC 14496, “Information technology – Coding of audiovisual objects,” 2001.
- [6] Y. Wang and Q. Zhu, “Error Control and Concealment for Video Communication: A Review,” *Proc. IEEE*, vol. 86, no. 5, pp. 974-997, May 1998.
- [7] J. Suh and Y. Ho, “Error Concealment Techniques for Digital TV,” *IEEE Trans. Broadcast.*, vol. 48, no. 4, pp.299-306, Dec. 2002.
- [8] S. Tsekeridou and I. Pitas, “MPEG-2 Error Concealment Based on Block-Matching Principles,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 10, no. 4, pp. 646-658, Jun. 2000.
- [9] W. Park, and B. Jeon, “Error Detection and recovery by Hiding Information into Video Bitstream using Fragile Watermarking,” *Proc. SPIE Visual Communications and Image Processing 2002*, vol. 4671, pp. 1-10, Jan. 2002.
- [10] J. Wen and J.D. Villasenor, “Reversible Variable Length Codes for Robust Image and Video Transmission,” *Proc. 31st Asilomar Conf. on Signals, Systems, and Computers*, vol. 2, pp. 973-979, Nov. 1997.
- [11] E. Khan, S. Lehmann, H. Gargi, and M. Ghanbari, “Iterative Error Detection and Correction of H.263 Coded Video for Wireless Networks,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 12, pp. 1294-1307, Dec. 2004.
- [12] W. Chu and J. Leou, “Detection and Concealment of Transmission Errors in H.261 Images,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 8, no. 1, pp. 74-84, Feb. 1998.
- [13] H. Shyu, and J. Leou, “Detection and Concealment of Transmission Errors in MPEG-2 Images – A Genetic Algorithm Approach,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 9, no. 6, pp. 937-948, Sept. 1999.
- [14] Y. Han and J. Leou, “Detection and Correction of Transmission Errors in JPEG Images,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 8, no. 2, pp. 221-231, Apr. 1998.
- [15] M.R. Pickering, M.R. Frater and J.F. Arnold, “A Statistical Error Detection Technique for Low Bit-Rate Video,” *IEEE Proc. of TENCON Conf.*, vol. 2, pp. 773-776, Dec. 1997.
- [16] S. Ye, X. Lin and Q. Sun, “Content Based Error Detection and Concealment for image Transmission over Wireless Channel,” *IEEE Proc. of ISCAS Conf.*, vol. 2, pp. 368-371, May 2003.

- [17] [O. Lehtoranta, T.D. Hamalainen, and V. Lappalainen, "Detecting Corrupted Intra Macroblocks in H.263 Video," *IEEE Workshop on Multimedia Signal Processing*, pp. 33-36, Dec. 2002.](#)
- [18] [K. Bhattacharyya and H.S. Jamadagni, "DCT Coefficient-Based Error Detection Technique for Compressed Video," *IEEE Proc. of ICME Conf.*, vol. 3, pp. 1483-1486, Aug. 2000.](#)
- [19] [L.V. Tran, "Efficient Image Retrieval with Statistical Color Descriptors," Ph.D. dissertation, Linköping University, Dept. of Science and Tech., Sweden, 2003.](#)
- [20] [A. Dailianas, R.B. Allen, and P. England, "Comparison of Automatic video Segmentation algorithms," *Proc. of SPIE*, vol. 2615, pp. 2-16, Jun. 1995](#)