TOWARDS AUTOMATIC STEREO-VIDEO QUALITY ASSESSMENT AND DETECTION OF COLOR AND SHARPNESS MISMATCH

Alexander Voronov, Dmitriy Vatolin, Denis Sumin, Vyacheslav Napadovskiy, and Alexey Borisov

Lomonosov Moscow State University, Graphics & Media Lab
{avoronov, dmitriy, dsumin, vnapadovsky, aborisov}@graphics.cs.msu.ru

ABSTRACT

In this paper, we address the problem of stereo-video quality assessment. We introduce objective no-reference metrics for automatic color- and sharpness-mismatch detection in video captured using stereo cameras. The algorithms are based on view matching and reconstruction. A fast block-based color-independent algorithm for stereo matching is proposed. We verify the applicability of the proposed metrics by assessing the quality of full-length films. This quality-assessment procedure reveals scenes distorted during film production or postproduction and enables film comparison in terms of stereoscopic quality.

Index Terms — stereoscopic video, stereo video, 3D video, color mismatch, sharpness mismatch, focus mismatch, video quality

1. INTRODUCTION

Creation of 3D video has become a major trend. Most films today are released in a 3D format, and 3D TV is set to enter consumers' homes. The audience for 3D films is growing essentially. The technology for creating a 3D film is complicated, and the quality requirements are high since the presence of artifacts affects a viewer's perception of the film. The other difficulty is that although artifacts may not be clearly visible while watching a film in 3D, the presence of these artifacts essentially tires the viewer's brain. Thus, quality assessment of 3D films is in high demand today. Our goal is to construct a system for quality assessment of stereoscopic video (when we speak about 3D video in this paper, we mean stereoscopic video).

There are two common approaches to creating a stereoscopic video (besides rendering): video can be captured with a stereo camera system, or it can be made by converting a 2D video. The likely artifacts in captured and converted videos will differ. Thus, we must develop different objective metrics to assess the quality of different videos. In this work we address captured video and introduce our metrics to evaluate color mismatch and sharpness mismatch between views. Our proposed metrics are no-reference: they take a stereo image as input and measure mismatch between the left and right views of the image. The metric result is a number that quantifies the noticeability of distortion in the image; the greater the value, the more noticeable the detected distortion. Applying a metric to a long video sequence (e.g., a full-length film) enables discovery of scenes that require better processing or equipment setup.

2. RELATED WORK

The move from 2D video to 3D video begets new challenges in video-quality assessment. Some of these challenges are described in [1]. New issues arise throughout the pipeline from the video creator to the viewer: during content creation, format conversion, encoding and decoding, and rendering for a certain display device. More-detailed classification of issues with 3D video is presented in [2] with attention paid specifically to mobile-3DTV concerns.

There are several approaches to assessing 3D video quality. We divide them into three groups:

- approaches based on the characteristics of the human visual system (HVS) and subjective perceptual test results,
- no-reference methods based on 2D-video quality-assessment that are modified and applied to 3D video (possibly with disparity- or depth-map estimation), and
- algorithms for discovering or correcting specific stereo artifacts (like color mismatch or geometric misalignment) and combination of such algorithms.

The first approach includes empirical metrics based on the characteristics of the human visual system (HVS) and subjective perceptual results. In [3] and [4], the authors introduce metrics that are based on the luminance components of image pixels and compare the results with subjective evaluations, minimizing the difference between the metric values and the responses of test subjects. In [5], the authors also build their metric on the basis of subjective tests. They use disparity-map segmentation to improve the quality of the assessment.

The second approach is to improve objective metrics for 2D-video quality assessment in order to apply them to 3D-video views or the estimated disparity field (depth map). But here we do not use any "ground truth" information to define the metric. In [6], the authors propose a blur measure based on computing the sharpness of each pixel in one view using the estimated disparity map. In [7], an ideal depth map for a video frame is used to calculate three distortion measures: temporal outliers, temporal inconsistencies and spatial outliers. Thee overall measure is then calculated.

The last approach is a group of algorithms for stereoscopic-video correction or detection of specified artifacts; these algorithms can be effectively modified for 3D-video quality estimation. In [8], the authors propose an effective color-correction algorithm based on feature-point estimation and matching between the views. In [9], a method for correcting zoom mismatch is presented. Cropping and scaling are applied to one of the views after several matching points are found.

3. SYSTEM FOR AUTOMATIC STEREO VIDEO QUALITY ASSESSMENT

In [10], the system for automatic stereo video quality assessment was presented, and the results were demonstrated for the well-known film trailers. In this work we are focusing on the captured stereo and full-length film processing. The result for every film is metric values and the set of the scenes with the most noticeable artifacts. The algorithms for color mismatch and sharpness mismatch estimation are described below. The horizontal parallax and geometry distortion were estimated the same way as in [10]. The results of contemporary films comparison are presented in Section 6.

Presented algorithms can be useful for capturing devices comparison and calibration. For example, analysis of the same scenes captured with different cameras makes possible to understand which is better in terms of color mismatch. Algorithms of sharpness mismatch and geometry distortion detection are applicable for on-fly camera system setup.

4. COLOR MISMATCH ESTIMATION

In [10], the color-mismatch metric assumes only global color distortion, which appears when cameras have different brightness or color balance. A typical feature of such distortions is they can be fixed using color adjustment, such as with the “Curves” function in graphics editors.

We have improved our algorithm for handling not only these types of scenes but also scenes with local color distortions like glaring, nonuniform lighting and nonuniform color distribution owing to different light filters. The idea here is to compensate one view to the other and then estimate the difference, assuming possible errors on the boundaries and the occlusion areas.

Estimation for this metric is performed using the following steps:

1. Global color correction of the right view;
2. Matching of the original left view to the corrected right view;
3. Reconstruction of the right view from the left view on the basis of stereo-matching results;
4. Analysis of differences between the original right view and the views reconstructed from the left, and computation of the metric value.

4.1. Global Color Correction

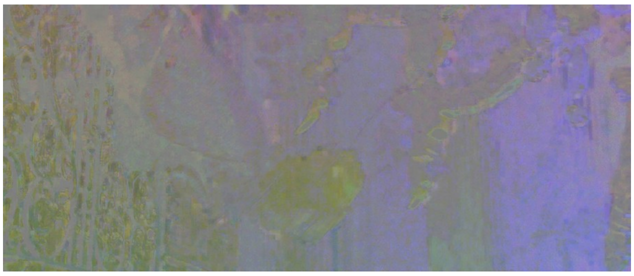
Global color correction is a preprocessing step before views are matched. This step eliminates differences in lighting and color balance thereby improving the robustness of compensation. The idea is to estimate a color transform that minimizes the difference between the integral histograms of images from the left and the right views.

The results of color correction are used only in the stereo-matching procedure. For the final estimation of the metric value, we use original images.

4.2. Color-Independent Stereo Matching

For matching the right view to the left view we use a block-based stereo-matching algorithm that is color independent. We chose the block-based approach because of its lower computational complexity relative to optical flow and feature point matching. The accuracy of this type of approach is suitable for evaluating distortions. The stereo-matching algorithm is applied to the original left view, and the right view processed is using global color correction.

(a) Left view of the source frame.



(b) Visualization of the difference between the original left view and the left view reconstructed from the right view.

(c) Checkerboard visualization of color mismatch between the original left view and the left view reconstructed from the right view.

Figure 1: Example of scenes with color mismatch – frame #21738 from the film *Pirates of the Caribbean On Stranger Tides*.

As the basis of our approach, we employed the motion-estimation algorithm described in [11] and modified it to make it more robust for images with color mismatch. In the original algorithm, the sum of absolute differences (SAD) metric is used to compare blocks. SAD is computed in the following way:

$$SAD_{L,R}(\Omega) = \sum_{x\in\Omega}\Big|L(x) - R(x+s)\Big|, \qquad (1)$$

where L and R are the left and right views, respectively; x is a point in block Ω; and s is a supposed displacement vector.

This metric was replaced with the modified sum of normalized Differences (MSND):

$$MSND_{L,R}(\Omega) = SND_{L,R}(\Omega) + \\ + \frac{k}{|\Omega|}\sum_{x\in\Omega}\Big(L(x) - R(x+s)\Big)^2 \qquad (2)$$

The first term is the SND metric value between blocks, which is given in [3]. The second term eliminates the need to consider the case where a solid white block could be matched to a solid black block with zero penalty, a case that is possible with original SND metric. The parameter k quantifies the magnitude of the possible color difference between blocks. The greater the color difference between views, the lesser the value needed to provide a good result.

$$SND_{L,R}(\Omega) = \sum_{x\in\Omega}\Big|\Big(L(x) - R(x+s)\Big) + \\ + \frac{1}{|\Omega|}\sum_{x\in\Omega}\Big(L(x) - R(x+s)\Big)\Big| \quad (3)$$

Comparing the value of the MSND metric to the SAD relies more on the texture similarity than on the color similarity of the blocks. Thus, the MSND metric enables matching of the views even if they have local color differences. Relative to the original SND, the proposed metric yields more robust and uniform results.

The number of operations for the stereo-matching step increases approximately three times, because we must now calculate the sum of the pixel values for every block, in addition to the pixelwise difference between blocks.

4.3. Metric-Value Calculation

To compare views we reconstruct the right view from the left one using the results of the stereo- matching algorithm. For reconstruction and metric estimation, we use original images, not the results of the color correction. The color-mismatch metric value is calculated as the mean square error (MSE) between the reconstructed left view and the original right view. To eliminate the influence of poorly compensated areas and occlusions, we ignore the 5% of pixels with the highest compensation error when calculating the MSE. Higher metric values indicate greater color distortion. Figure 1 shows an example of a scene detected by the metric. The greater the variation from gray in the visualization, the greater the color distortion present in that region.

5. SHARPNESS-MISMATCH ESTIMATION

The goal of this metric is detection of regions in the frame with a noticeable difference in sharpness between the left and right views. Such artifacts appear when cameras have different focal settings; for example, the cameras are focused on different objects or have different depth-of-field values. The difference in sharpness implies that some area in one view is blurred more than in the other view, which means that the amount of high frequencies in this area differs between views. So the task can be formulated as detection of areas with different amount of high frequencies between views.

The algorithm for solving this task comprises the following steps:

1. Extraction of high frequencies from both views;
2. Matching of the high-frequency maps;
3. Penalty-map computation and refinement;
4. Extraction of dense regions (noticeable regions with high-frequency mismatch) from the penalty map;
5. Metric-value estimation on the basis of the properties in the worst region.

To get the high frequencies, we subtract the result of bilateral smoothing from the reference view. We prefer a bilateral filter to a Gaussian filter, because it avoids extraction of unnecessary information on the edges. An example of the extracted high-frequency maps for the left and right views is presented in Figures 2c and 2d.

Matching of high-frequency maps is based on the results of block-based stereo matching [11]. To increase the reliability of the matching, the high-frequency maps are smoothed by a Gaussian filter. The penalty map is computed as a SAD (sum of absolute differences) of the left-view high- frequency map and the compensated right-view high-frequency map.

The goal of the penalty-map refinement step is to discard occlusion areas from the analysis. The refinement algorithm is based on the left-right consistency (LRC) check algorithm, which is applied to the results of the bidirectional (left-to-right and right-to-left) stereo-matching algorithm. The LRC is computed by projecting a pixel back and forth, using the pixels disparity values and its projection. If the difference between the initial and final positions is greater than a given threshold, the pixel is assumed to be occluded.

We apply a thresholding technique, omitting the worst 15% of the penalty map. An example penalty map is presented in Figure 2b. The connected-components labeling algorithm is applied to extract from the penalty map any dense regions with possible mismatch. For every region we compute a penalty value, considering the average value of the region on the penalty map and the size of that region. Figure 2 shows an example of a scene with sharpness mismatch, along with visualizations of the intermediate steps in the algorithm.

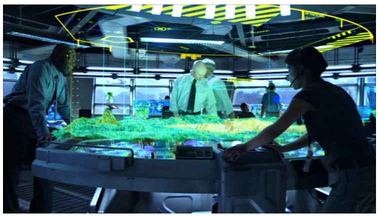
6. EVALUATION

We evaluated our presented algorithms using full-length films. Twelve films were analyzed with the sharpness- and color-mismatch detection algorithms described above and with the geometry-distortion detection and horizontal-parallax estimation algorithms presented in our previous work [10]. Video was taken from Blu-ray 3D discs, so it is compressed, but this format offers the best publicly available quality. Table 1 presents the average metric values for some of the tested films. Horizontal parallax is measured in percent of screen width. This measure is selected because it is invariant with respect to image scale if the aspect ratio is fixed. Vertical parallax is measured in per mil, because the order of magnitude for a typical value is less than for horizontal parallax. Color mismatch and sharpness mismatch are presented as dimensionless quantities.

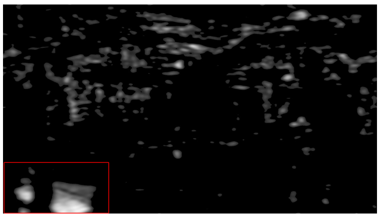
In terms of stereo quality, *Pirates of the Caribbean: on Stranger Tides* is the leader, but it also is one of the “flattest” films. And even the leader has numerous scenes with color or sharpness mismatch, which our metrics detected.

Figures 3 and 4 contain integral histograms demonstrating the distribution of color- and focus- mismatch values for the examined films. A point with coordinates (x, y) on the graph means that y percent of frames in the film have a metric value not greater than x. Curves for films with better quality are “closer” to the top-left corner of the graph.

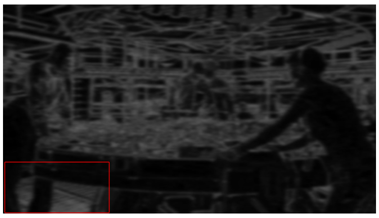
The evaluation system is implemented as a console application that can be extended with plug-ins. Code is written in C++ and uses OpenCV and Boost libraries. The processing speed for the full set of metrics is 0.4 fps at Full HD resolution using an Intel Xeon E5520 2.26GHz processor with 12GB of RAM.



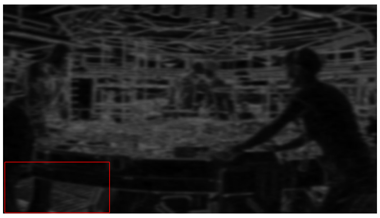
(a) Left view of the source frame.



(b) Sharpness mismatch penalty map estimated for the left view.



(c) Extracted high frequency map for the left view.



(d) Extracted high frequency map for the right view.

(e) The worst region extracted from the left and right views.

Figure 2: Example of scene with sharpness mismatch – *Avatar Pandora Discovered trailer*, frame #3810. The brightest region in the penalty map (b) indicates an area with sharpness mismatch. The red rectangle marks this region in the penalty and extracted high-frequency maps. In (e), we show this region cropped from both views.

Film title	Pos. parallax (% of screen width)	Neg. parallax (% of screen width)	Vertical parallax (‰ of screen width)	Color mismatch	Sharpness mismatch
Avatar	**1.10956**	0.53390	0.05520	8.68422	0.84841
Step Up 3D	0.87002	0.72828	**0.69251**	**23.50915**	**1.30704**
Resident Evil: Afterlife	0.67100	0.47235	0.25704	12.03261	1.14776
Sanctum	**1.07146**	0.81021	0.40791	7.39567	1.14795
Pirates of the Caribbean: On Stranger Tides	0.65012	0.31193	0.09160	2.00761	0.67996
Dolphin Tale	0.94476	**1.36344**	**0.58167**	**37.23440**	**1.17072**
The Three Musketeers	0.82839	0.03457	0.26934	6.92880	0.60300
Hugo	0.57896	**0.96668**	0.17126	3.49118	0.91383

Table 1: Results of the performed full-length film evaluation. The largest values of every metric are marked in bold. Positive and negative parallax metric values describe how much is the "3D" perception in a film. High metric values of vertical parallax, color and sharpness mismatch mean that a film is not of high quality.

7. CONCLUSION

We have presented our algorithms for automatic detection of scenes with color and sharpness mismatch in stereoscopic video. The proposed methods are suitable for stereo-video quality assessment. The proposed metrics may also increase quality of setup and calibration of capturing devices. To demonstrate the applicability of our system, we processed several Blu-ray 3D releases of well-known films, revealing many scenes with strong artifacts. Average metric values can also be used to compare different films in terms of quality and strength of the 3D effect.

8. FURTHER WORK

Our next goal is to develop new metrics to perform quality assessment of converted video. These metrics will be more complicated than those for captured stereo video because numerous issues arise in the 2D-to-3D conversion process. We plan to estimate some depth-map features and their correspondence with the source views. It will then be possible to estimate the quality of edge processing and occlusion filling. Such a set of metrics will enable complex quality assessment of converted films.

Another direction of future work is optimization of the algorithm's processing speed. The main opportunity here is porting the most time-consuming parts to the GPU.

We plan to share our results with the community. For one, we intend to involve specialists in 3D- video creation to get feedback on our results. This may help improve the quality of future stereo films.

Owing to the complexity of the problem studied in this work, we did not perform any complicated subjective testing or any analysis of correlation between human perception of artifacts and our results. The results of human-perception tests can be dependent on the type and quality of a display device; also, the results are susceptible to viewers' visual acuity and their ability to perceive 3D. This topic should therefore be addressed separately. One of

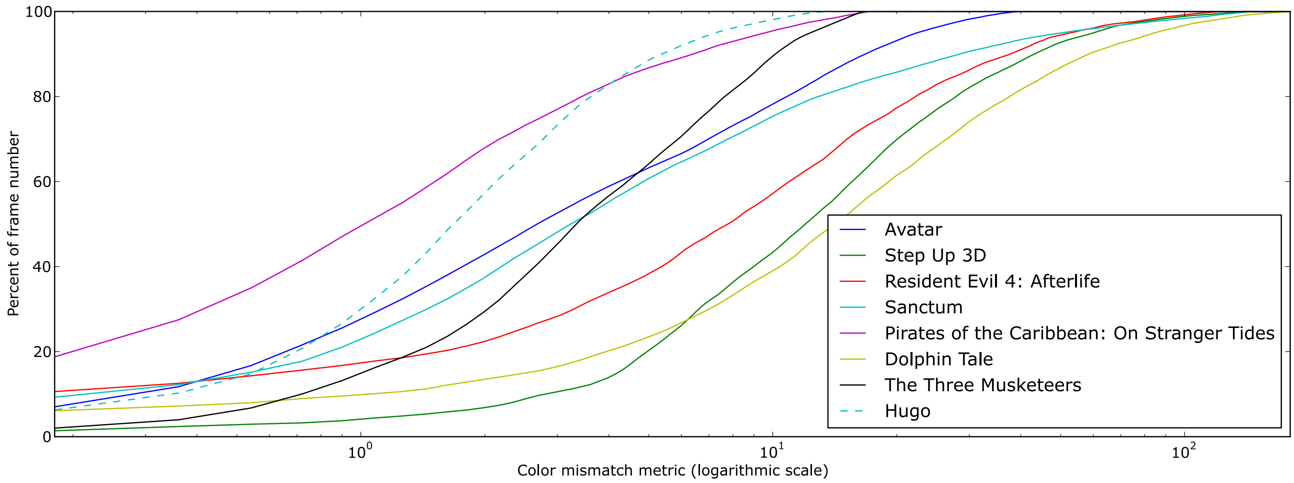


Figure 3: Integral histogram of the color mismatch metric results for full-length films. The best films in this graphs are *Pirates of the Caribbean* and *Hugo*. The average level of *Pirates of the Caribbean* is better, but the worst scenes of *Hugo* are better than the worst ones in *Pirates of the Caribbean*.

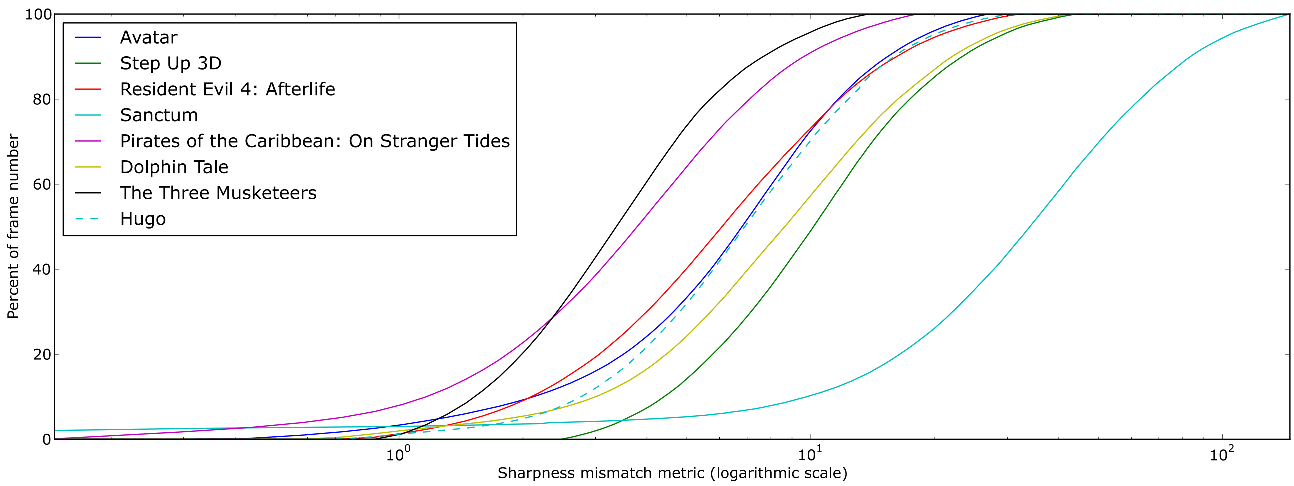


Figure 4: Integral histogram of the sharpness mismatch metric results for full-length films. *Pirates of the Caribbean* and *The Three Musketeers* are the best. The results for *Sanctum* film are significantly worse than the results of the others.

the first things to determine in this case is the threshold of artifact noticeability for each metric.

9. ACKNOWLEDGEMENT

This work is partially supported by the Intel/Cisco Video Aware Wireless Network (VAWN) Program and by grant 10-01-00697a from the Russian Foundation for Basic Research.

10. REFERENCES

[1] Quan Huynh-Thu, P. Le Callet, and M. Barkowsky, “Video quality assessment: From 2d to 3d – challenges and future trends,” in *17th IEEE International Conference on Image Processing (ICIP)*, Sept. 2010, pp. 4025–4028.

[2] A. Boev, D. Hollosi, A. Gotchev, and K. Egiazarian, “Classification and simulation of stereoscopic artifacts in mobile 3dtv content,” in *Proc. SPIE Stereoscopic Displays and Applications XX*, 2009.

[3] P. Joveluro, H. Malekmohamadi, W.A.C. Fernando, and A.M. Kondoz, “Perceptual video quality metric for 3d video quality assessment,” in *3DTV-Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON)*, June 2010, pp. 1–4.

[4] G. Nur, H.K. Arachchi, S. Dogan, and A.M. Kondoz, “Extended vqm model for predicting 3d video quality considering ambient illumination context,” in *3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON)*, May 2011, pp. 1–4.

[5] S.S. Sarikan, R.F. Olgun, and G.B. Akar, “Quality evaluation of stereoscopic videos using depth map segmentation,” in *Third International Workshop on Quality of Multimedia Experience (QoMEX)*, Sept. 2011, pp. 67–71.

[6] Frédéric Devernay, Sergi Pujades, and Vijay Ch.A.V., "Focus mismatch detection in stereoscopic content," in *SPIE Proceedings on Stereoscopic Displays and Applications*, San Francisco, États-Unis, Jan. 2012, vol. 8288.

[7] M. Solh and G. AlRegib, "A no-reference quality measure for dibr-based 3d videos," in *IEEE International Conference on Multimedia and Expo (ICME)*, July 2011, pp. 1–6.

[8] Jung-Jae Yu, Hae-Dong Kim, Ho-Wook Jang, and Seung-Woo Nam, "A hybrid color matching between stereo image sequences," in *3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON)*, May 2011, pp. 1–4.

[9] C. Doutre, M.T. Pourazad, A. Tourapis, P. Nasiopoulos, and R.K. Ward, "Correcting unsynchronized zoom in 3d video," in *IEEE International Symposium on Circuits and Systems (ISCAS)*, June 2010, pp. 3244–3247.

[10] A. Voronov, A. Borisov, and D. Vatolin, "System for automatic detection of distorted scenes in stereo video," in *Proceedings of Sixth International Workshop on Video Processing and Quality Metrics (VPQM)*, Jan. 2012.

[11] K. Simonyan, S. Grishin, D. Vatolin, and D. Popov, "Fast video super-resolution via classification," in *Proceedings of IEEE International Conference on Image Processing*, 2008, pp. 349–352.