DETECTING ANAGLYPH IMAGES WITH CHANNEL ALIGNMENT FEATURES

Andrew C. Gallagher

Kodak Research Laboratories, Eastman Kodak Company

ABSTRACT

An anaglyph image is typically created by combining the color channels from a pair of stereo images to create a new three-channel image. Presently, when an imaging system encounters a three-channel image (such as a JPEG image), it is treated in a similar fashion as is any other three-color image. There currently is no method for determining whether the image is an anaglyph image or a single-view image.

In this paper, we propose a set of features for detecting anaglyph images. The algorithm extracts features based on stereo matching, edge coincidence between channels, channel correlation, and distributions of chroma values. On an independent test set with an even number of anaglyph and non-anaglyph images, the algorithm achieves a recognition accuracy of 95.2%.

Index Terms— anaglyph images, stereo matching

1. INTRODUCTION

The history of photography is a rich history of innovation and invention. From the earliest days of “sun drawing” in 1826, imaging systems have allowed photographers to record moments with ever-increasing levels of realism [1]. Innovators have had the goal of providing viewers with the most realistic experience possible. Not long after the first single-view cameras (roughly the 1850’s), stereoscopes and anaglyph images were introduced to allow a viewer to view two pictures of the same scene from two slightly different viewpoints, one view for each eye. As a result, the viewer experiences the scene with an impression of realism and depth that cannot be achieved when viewing a single-view 2D image.

An anaglyph image is one image that contains multiple views, such that each view is a specific color in the image. Rather than relying on optics for directing the proper view for each eye as is the case with a stereoscope, an anaglyph image is properly viewed with special anaglyph glasses having a different color for each lens (often red and blue). Each lens allows one view to pass through the lens to the eye and blocks the other. The human visual system then merges these stereo views to create the impression of stereo vision.

Although there has not been work on detecting anaglyph images, there have been efforts to identify the source of images in various domains. For example, by analyzing low-level image statistics, computer-generated images can be distinguished from camera-captured images [2], manipulated images can be identified [3, 4], and the camera model can even be identified in some cases [5]. In our work, we rely both on low-level and mid-level features based on stereo matching to determine whether or not an image is an anaglyph.

Determining whether an image is an anaglyph image (or, more broadly, a multi-view image) versus a single-view image enables several system applications. In a viewing system, a viewer can be notified to either put on or to remove anaglyph glasses for viewing a specific image. Furthermore, specialized algorithms can be applied to images depending on their type. For example, an algorithm such as stereo matching can be applied to an anaglyph image to produce a depth map.

Our contributions are the following: We present an algorithm for distinguishing between anaglyph and non-anaglyph images. Our approach is based on the observation that objects are misaligned between the color channels for anaglyph images. Based on this observation, the algorithm uses features based on stereo matching and edge coincidence measures. This work shows that the origin of an image can be recovered and provokes interesting questions related to deploying 3D systems in the consumer market.

2. FEATURES

To understand what features might be useful for detecting whether an image is an anaglyph image or not, it is useful to review the steps used to create an anaglyph image. The approach is to extract a set of features that capture the relationship between the alignment of image channels and whether those channels are from the same image, or are from two different views of the same scene. A classifier is then used to learn the relationship between feature values and classification (anaglyph or non-anaglyph). Then, when a new image is encountered, the features can be extracted and the image classified as the most likely class for those observed feature values.

2.1. Producing Anaglyphs

The production of an anaglyph image begins with the capture of a stereo pair of images. To produce a grayscale anaglyph image, one color channel (usually red) of the anaglyph image
Fig. 1. Anaglyph images are more likely to contain saturated colors than non-anaglyph images. Left: The average histogram of chroma values for anaglyph images (left) shows a higher concentration of saturated pixels (on the edges of the histogram) than for non-anaglyph images (right). The center of each histogram represents a neutral or unsaturated pixel.

is set to the left image of a grayscale stereo pair, and the other channels (green and blue) of the anaglyph are set equal to the right image of a grayscale stereo pair. A color anaglyph is produced by combining the red channel of the left image with the green and blue channels of the right image, or some permutation of this procedure. To improve the visual experience, stereo registration can be performed on the stereo pair to maximize the amount of overlap between the common structures of the stereo pair [6]. This step has the additional benefit of reducing “ghosting” of one or both stereo views to the unintended eye caused by mismatch between the spectral sensitivities of the filters in the anaglyph glasses and the display. Software is available to perform the steps to produce anaglyph images with these and other options [7].

In many color anaglyphs, the red channel is from the left stereo view, and the green and blue channels are from the right view. Because different channels of an anaglyph are from different views, there is a misalignment of the objects and edges between those channels. Measures of this misalignment are found with stereo matching algorithms, and by directly examining the coincidence of edges from different channels. In addition, the correlation of the pixel values between a pair of channels is less when each channel is from a different view than when the channels are from the same view (i.e. are two channels of the same image). Further, the distribution of the colors of pixels between anaglyph and non-anaglyph images are different. All features are computed on a low resolution version of the image having no more than 384 pixels across the long dimension.

2.2. Chroma Distribution Feature

In an anaglyph image, a particular pixel’s value is related to the incident light from one of two scene objects, depending on the channel. Consequently, the values across a channel for a pixel are not necessarily related and can result in the appearance of colors that were never present in the scene. In an extreme example, an anaglyph image of a completely gray scene will often contain highly saturated pixel values where object boundaries are misaligned.

To produce a measure of these falsely colored pixels, the chroma histogram of the image is computed as follows: First, each pixel is rotated to the Ohta color space [8], where two chroma channels are computed as follows:

\[
\begin{align*}
    c_1 &= -\frac{1}{4}R + \frac{1}{2}G - \frac{1}{4}B \\
    c_2 &= -\frac{1}{2}R + \frac{1}{4}B
\end{align*}
\]  

Next, the two-dimensional histogram for each image is found using a histogram of 13 × 13 bins, and each pixel increments a single bin by one. The mean histogram for anaglyph and non-anaglyph images is shown in Fig. 1, where it is apparent that anaglyph images contain a higher proportion of saturated pixel values. This chroma histogram is an effective feature for distinguishing between a single-view three color image and an anaglyph (a multi-view three color image) because anaglyph images tend to have a greater number of pixels with a red or cyan/blue hue than a typical single-view three color image would.

2.3. Channel Correlation

The correlation coefficients between each pair of image channels are found, for a total of three features: \[^{\rho_{12}}\rho_{13}\rho_{23}\]. As shown in Fig. 2, the correlation between the red and the green channels is usually smaller when the image is an anaglyph. However, because the green and blue channels are usually from the same single-view image even for anaglyphs, there is very little difference between the respective distributions of the correlation coefficient \[^{\rho_{23}}\].

2.4. Edge Coincidence

For a single-view three color image, the edges found in one channel tend to coincide in position with the edges in another channel because edges tend to occur at object boundaries.
However, in anaglyph images the channels originate from disparate perspectives of the same scene, and the edges from one channel are less likely to coincide with the edges from another. Therefore, measuring the edge overlap between the edges from multiple channels provides information relevant to the decision of whether an image is an anaglyph or a non-anaglyph. For these features, two channels are selected (red and blue) and the candidate edges for each are found as those pixels with a gradient magnitude (Prewitt operator) greater than the remaining 90% of the other pixels from the channel. In addition, edge pixels also must have a greater gradient magnitude than any neighbor in a local neighborhood ($3 \times 3$). Three feature values are computed: the number of locations that are edge pixels in both channels, the number of locations that are edge pixels in at least one channel, and the ratio of the two numbers. Fig. 3 illustrates the edge coincidence features by showing the edge pixel locations for one anaglyph image and one non-anaglyph.

### 2.5. Inter-channel Stereo Matching

An anaglyph image contains views of a scene from multiple perspectives in different color channels. As such, a stereo correspondence algorithm [9] used to determine the disparities between feature points between image channels.

Intuitively, when the two channels are from a single-view image and correspond only to two different colors, the alignment between a patch of pixels from one channel with the second channel is often best without shifting or offsetting the patch with respect to the second channel 122. However, when the two channels are each from different view of a multi-view image (as is the case with an anaglyph image), then the best local alignments between a patch of pixels from one channel with the second image channel is often a non-zero offset.

Note that in some ways, this correspondence problem is actually more difficult that the one typically addressed in stereo matching literature. The typical stereo matching problem addresses finding correspondence between image features across stereo views in the same color channels. Usually, two grayscale channels are matched. However, in our case, the stereo views differ in perspective and also in color (since the red channel is from one view and the green and blue are from the other). This difficulty has implications on the selection of the quality measure that is used to measure local alignment.

All stereo alignment algorithms require a measure of the quality of a local alignment, also referred to as matching cost, (i.e. an indication of the quality of the alignment of a patch of pixel values ($7 \times 7$) from the first channel at a particular offset with respect to the second image channel). Typically, a measure of pixel value difference (e.g. mean absolute difference, mean square difference) is used as the quality measure. However, because the channels undergoing matching represent different colors, a preferred quality measure is correlation between the channels rather than pixel value differences. Quality measures based purely on pixel value difference tend to fail in areas corresponding to saturated colors (e.g. the sky) that naturally have a large difference between the pixel values for each color. We have good results performing stereo matching on the gradient magnitude for the pairs of color channels, using a quality measure of the sum of squared differences. The stereo matching is computed over a sparse grid (every $16^{th}$ pixel in each row and column. Assuming that the image is a stereo image captured with horizontally displaced cameras, the stereo alignment need only search for matches along the horizontal direction. Three features are computed from the disparity maps: the number of pixels with a non-zero displacement, the average pixel displacement and the median displacement at all pixel locations.

Fig. 3 shows the stereo matching results for an anaglyph and a non-anaglyph image. Despite the difficulty of performing matching across color channels, the features that are computed have been found to be insensitive to a small number of mismatches.

### 3. IMAGES AND EXPERIMENTS

To test our ideas, we collected a set of 2000 images from Flickr using the search terms: “anaglyph” and “3D”. Further, a second set of 2000 general images (non-anaglyph) were collected. A training set is generated using the first 1000 images from each of the two classes, and the remaining 2000 images comprise the test set.

Adaboost is used to train a classifier (using 200 iterations with decision stumps on single features as the weak classifiers), although other classifiers yield nearly similar results. Table 1 shows the classification accuracy (equal error rate) for each of the four feature categories, and for the combined Adaboost model. Overall, the test set accuracy is 95.2%. Performance is best when all features are considered by Adaboost. Fig. 4 discusses results of the classifier on the test set. Support vector machines yield a similar result on this dataset.

### 4. CONCLUSIONS

As multi-view image capture becomes more prevalent, images systems will need methods for distinguishing between the different types of content such as multi-view from single-view images. This ability will allow a display system to op-

<table>
<thead>
<tr>
<th>Feature</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chroma Distribution</td>
<td>93.8%</td>
</tr>
<tr>
<td>Channel Correlation</td>
<td>82.7%</td>
</tr>
<tr>
<td>Edge Coincidence</td>
<td>90.7%</td>
</tr>
<tr>
<td>Stereo Matching</td>
<td>87.2%</td>
</tr>
<tr>
<td>All Features</td>
<td>95.2%</td>
</tr>
</tbody>
</table>

Table 1. Performance of features for anaglyph detection.
Fig. 3. An illustration of feature extraction for an anaglyph image (top), and non-anaglyph image (bottom). Each row shows the red and blue channels. In an anaglyph, these channels correspond to different perspectives. Images (d) and (i) show the results of stereo matching, where middle gray indicates that the block is best aligned with no offset. This occurs most often for the non-anaglyph image. Images (e) and (j) show the edges for the red and green channels. Coincident edges, which occur most often for the non-anaglyph image, appear white.

Fig. 4. Classification results, showing (a) correctly classified anaglyph images, (b) images misclassified as anaglyphs, (c) correctly classified non-anaglyph images, and (d) images misclassified as non-anaglyphs. False positive anaglyphs tend to have saturated colors (middle of (b)), or subtle texture (left and right of (b)). Missed anaglyphs tend to have either very little parallax (left of (d)) or are poorly aligned (middle of (d)). Best viewed by alternately wearing and removing anaglyph glasses.

The algorithm extracts features related to the misalignment between color channels that characterizes in anaglyph images. The algorithm extracts features based on stereo matching, edge coincidence, channel correlation, and chroma distribution. On a test set of anaglyph and non-anaglyph images, the algorithm achieves an accuracy of 95.2%.

5. REFERENCES