ISSN 0278-6419, Moscow University Computational Mathematics and Cybernetics, 2016, Vol. 40, No. 4, pp. 185–191. © Allerton Press, Inc., 2016.
Original Russian Text © D.S. Vatolin, S.V. Lavrushkin, 2016, published in Vestnik Moskovskogo Universiteta, Seriya 15: Vychislitel'naya Matematika i Kibernetika, 2016, No. 4, pp. 40–48.

Investigating and Predicting the Perceptibility protect of Channel Mismatch in Stereoscopic Video#

D. S. Vatolin* **and S. V. Lavrushkin****

Faculty of Computational Mathematics and Cybernetics,
Moscow State University, Moscow, 119991 Russia

Received February 11, 2016

Abstract—The degree of visual discomfort caused by watching stereoscopic scenes with channel mismatch is investigated and predicted. A scene with channel order mismatch is one in which the right and left views are swapped. A way of finding channel mismatch is used to analyze 105 3D films; the scenes found in this analysis are used for experimental study of the visual discomfort caused by channel mismatch. The experimental results are used to construct sampling with a reference pattern. This sampling is used to learn different regression analysis algorithms, and the best way of predicting visual discomfort caused by channel mismatch is chosen.

Keywords: stereoscopic video, channel mismatch, channel mismatch perceptibility, regression analysis.

DOI: 10.3103/S0278641916040075

1. INTRODUCTION

A large number of 3D films are produced every year. In 2013–2014, more than 100 stereoscopic films with an average budget of US$130 million were released. The interest of the audience in 3D films, however, is waning. Some people suffer from headaches, fatigue, discomfort, thus losing the desire to watch more 3D films. This is due to a number of problems of stereoscopic film production, of which the quality of the produced content is one. There are many stereoscopic video artifacts capable of causing discomfort. Channel mismatch is one such artifact: the discomfort caused by scenes with channel mismatch is quite great, but fixing the problem (if it is detected in time) is very simple. Even so, the above problem was found in 23 out of the 105 examined films. In the correct channel order, the points of objects in front of the screen plane in the left channel are more to the right than the corresponding points in the right channel, while the opposite is true for the points of objects behind the screen plane. The shift between the corresponding points is called disparity. Objects in front of the screen plane thus have negative disparity, while objects behind the screen plane have positive disparity. It is disparity that largely determines the distance to the object or the depth perceived by the viewer. If the channels are swapped, positive disparity becomes negative; i.e., the closest spatial points are transformed into the farthest, and vice versa.

It is quite difficult to recognize an artifact of this type by just watching it. When we watch a scene with channel mismatch, we observe an image that is impossible in reality. Because of disparity inversion, the relief is turned inside out: convex objects seem concave, and vice versa. A scene with channel mismatch watched by an unprepared viewer negatively influences the way he feels and causes some discomfort, even headaches. The presence of such artifacts in a film is thus extremely undesirable.

*E-mail: dmitriy@graphics.cs.msu.ru.

**E-mail: slavrushkin@graphics.cs.msu.ru.

#This work was supported by the Russian Foundation for Basic Research, project no. 15–01–08632–a.

An algorithm for finding scenes with assumed channel mismatch was proposed in [1]. The authors, however, did not study the dependence of the degree of visual discomfort on the characteristics of a scene with channel mismatch. The experiment performed in this work showed that the degree of visual discomfort can vary strongly, depending on a particular scene. The aim of this work was improve this algorithm by adding automatic estimates of the discomfort caused by scenes with channel mismatch. This work is devoted to creating a data base of scenes with channel mismatch that contains a reference pattern of the degree of discomfort, and to constructing an algorithm for automatically predicting the degree of visual discomfort.

2. SURVEY OF THE SUBJECT AREA

A great many works in the field of stereoscopic vision have been devoted to investigating the discomfort caused by 3D films. One defining factor of such discomfort is excessive disparity. Excessive disparity results in conflict between accommodation and convergence, which increases the load on the human visual system. A great many ways of measuring the discomfort caused by stereoscopic films is therefore based on the statistical characteristics of disparity [2–4]: its average value, variance, maximum value, and total range.

In [2], the average disparity and its variance for the whole image were calculated to estimate visual discomfort from a frame. In [3], the average disparity and a range of the disparity map for an entire frame were analyzed. The disparity map range was calculated as the difference between the 95 and 5 percentiles of a disparity histogram. In [4], the maximum disparity and range of a disparity map were calculated. The maximum disparity was calculated as sum $\delta\%$ of the greatest disparities (where δ is the heuristic parameter of the algorithm). In [4], a weighted disparity map that considered the image texture was used; the values of the disparity map were multiplied by the absolute value of an image gradient calculated using the Sobel operator.

In [5], two additional disparity characteristics were considered along with the standard ones: the relative disparity (average difference between the disparities of neighboring objects) and object thickness (ratio of the average object width and average absolute object disparity). To calculate these characteristics, an image was first segmented with respect to disparity. These characteristics were then calculated for each object in the segmented image, and the maximum for the first characteristic and the minimum for the second one were chosen. It was experimentally established in [5] that the combined use of these characteristics and the characteristics used in [2–4] considerably increased the accuracy of predicting discomfort based on the statistical characteristics of disparity. A test set consisting of 120 stereoscopic images was used in each experiment. The test set was shown to 20 respondents who rated discomfort from each image from 1 to 5, where 1 was strong discomfort and 5 no discomfort. We used the decision making tree in [6] to predict discomfort.

In [7], it was proposed that we consider the visual model of human attention to estimate the discomfort caused by a stereoscopic video. The saliency map in [8] is used, and the statistical characteristics of disparity were calculated using this map. In [7], an experiment similar to the one described in [5] was performed to estimate discomfort. The reference vector method in [9] was used to predict discomfort.

Along with excessive disparity, discomfort from stereoscopic films can be caused by numerous stereography artifacts. These artifacts are color mismatch, geometric distortion (shift, rotation, Object scale discrepancy, and so on), time mismatch, sharpness mismatch, and swapped views in a scene [1, 10–13]. Even if disparity distribution in the frame meets certain norms, such artifacts can cause considerable discomfort, up to headaches. We are not familiar with any studies devoted to investigating and predicting the degree of discomfort caused by scenes with channel mismatch.

3. INVESTIGATING THE DEGREE OF DISCOMFORT CAUSED BY SCENES WITH CHANNEL MISMATCH

The procedure for finding channel mismatch proposed in [1] was used to analyze 105 various stereoscopic films. In 23 of these, we detected 65 scenes with channel mismatch with a total duration of 189 s. We performed the following experiment to investigate the degree of visual discomfort caused by stereoscopic video with channel mismatch (below, channel mismatch perceptibility):

The subjects watched a video sequence that included scenes with swapped views and had to rate the value of this artifact perceptibility from 1 to 5, where 1 meant that the artifact was imperceptible and

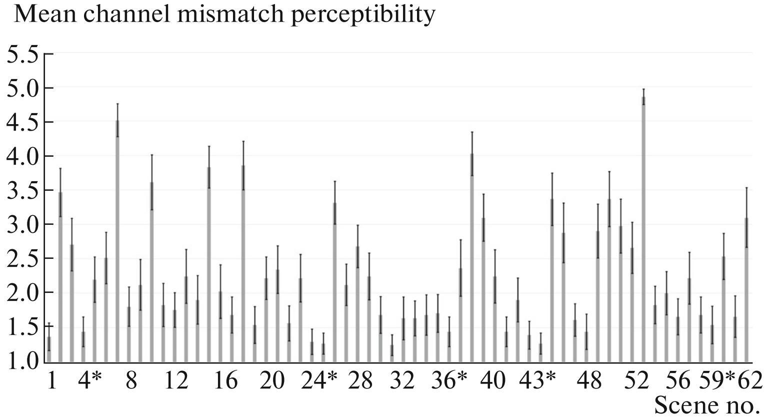

Fig. 1. Subjective evaluation of channel mismatch perceptibility in scenes (reference scenes are marked by asterisks).

the scene caused no discomfort, and 5 meant that the scene caused strong discomfort. The sequence was comprised of 56 scenes detected in analyzing our 105 films (some scenes were not included in the sampling due to high similitude). The sequence also included scenes preceding and following those with channel mismatch, so that the subjects watched scenes in the correct viewing order, along with those with swapped views, as is the case in real films. Each scene with channel mismatch and the neighboring scenes were shown three times, and the respondents then had time to rest and rate the artifact perceptibility in a test form.

To control the objectiveness of recipients' answers, a sequence included scenes without swapped views, and each sequence was shown both in direct and inverse order, since the value of channel mismatch perceptibility can depend on the perceptibility of the preceding scene. A total of 59 subjects took part in the experiment. Answers from 10 people that differed strongly from the average and had high perceptibility values of reference scenes were eliminated from consideration. The experimental results are shown in Fig. 1.

It follows from these data that there were scenes in which channel mismatch was imperceptible and could not be distinguished from normal scenes, and there were also scenes that caused serious discomfort. In most cases, swapped views were less noticeable in dark scenes, scenes with narrow ranges of disparity, and short scenes. It was found that scenes with channel mismatch can make a person feel much worse. After watching the experimental video sequence, most subjects thus complained of fatigue, sleepiness, and some of headache.

4. PREDICTING CHANNEL MISMATCH PERCEPTIBILITY

Since quantitative estimates of the visual discomfort caused by scenes with channel mismatch using subjective studies requires much human input, an automatic method that included machine learning, i.e., regression analysis, was proposed for this purpose in [14]. In our case, it was necessary to predict the value of channel mismatch perceptibility in a scene. The criterial (predicted) variable was thus the channel mismatch perceptibility, a real number varying from 1 to 5. The following scene characteristics were used as predictors (features):

1) disparity variance;

2) average brightness;

3) average motion intensity;

4) scene duration;

5) feature vector calculated using the procedure for identifying channel mismatch proposed in [1]. This vector includes five features representing the results from the operation of channel mismatch search components that analyze perspective; out-of-order objects (objects whose depth is less than that of surrounding objects); disparity distributions; visibility domains (domains visible in one view and invisible in the other) in the stereo pair; and motion visibility domains.

Table 1. Estimated accuracy of prediction for perceptibility using linear regression methods

Algorithm	Error for training sampling	Cross validation error
Linear regression	0.3407	0.5691
Linear regression with predictor product	0	5572.3430
Linear regression with quadratic function	0.2764	4.4392

Table 2. Estimated accuracy of prediction for channel mismatch perceptibility using linear regression methods with regularization

Algorithm	Error for training sampling	Cross validation error	Regularization parameter
L_2 regularization for linear function	0.3880	0.5219	0.7300074
L_1 regularization for linear function	0.3534	0.5083	0.04803
L_2 regularization for quadratic function	0.3460	0.5290	0.8792964
L_1 regularization for quadratic function	0.4170	0.5676	0.1609762

The algorithm of frame block matching [15] is used to calculate disparity and motion vector maps for determining features 1 and 3. For each block in one image, this algorithm finds the corresponding block in the other image with quarter-pixel accuracy. The found block shifts are thus used as disparity values in matching views, as motion vectors for estimating motion, and in forming the corresponding maps.

Our experiment resulted in a sampling consisting of 56 examples (6 reference scenes used in the experiment were considered), each including 10 features and the value of channel mismatch perceptibility. We therefore had to choose the regression method that most accurately predicted the channel mismatch perceptibility in new scenes after learning with the given 56 examples. We used cross validation over separate objects to estimate quality and compare different regression models. The root mean square error was used.

4.1. Linear Regression

The simplest method of parametric regression is linear regression [2, 16]. Standard linear regression (using the initial 10 features), linear regression with the predictor product (in which the products of original features are used as additional features) over 55 features, and regression with quadratic function (in which squares of the original features are used as additional features) over 20 features was tested. The results from these algorithms for predicting perceptibility are given in Table 1.

It can be seen from Table 1 that complex models of linear regression result in strong overfitting. If linear regression with the predictor product is used, the number of actual variables approaches the number of examples in the sampling. The data from reference sampling were therefore approximated ideally, while there were great overshoots beyond the range of the criterial variable in predicting cross validation.

Two ways of compensating for overfitting were tested: L_1 regularization [17] and L_2 regularization [18]. The regularization coefficients were chosen by minimizing the cross-validation error. The results from these algorithms for the predicting perceptibility are given in Table 2.

Regularization thus helped to overcome overfitting and reduce the cross validation error. Overfitting occurred even with linear regression over original 10 features.

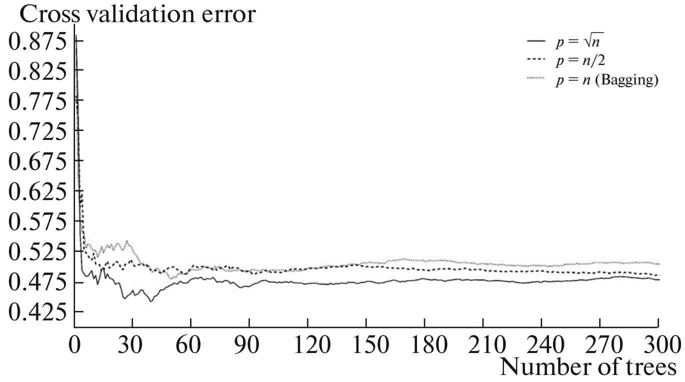

Fig. 2. Estimation of prediction accuracy for channel mismatch perceptibility using the Random Forest method.

4.2. Regression over k Nearest Neighbors

The simplest method of nonparametric regression is regression over k nearest neighbors [2]. When applied to our problem, the result was worse than simple linear regression. The best result for regression over k nearest neighbors was obtained for $k = 14$. The error for the training sampling was 0.5010; the cross validation error, 0.5805. When the feature space dimensionality was increased, the accuracy of regression over k nearest neighbors normally fell. The number of features was 10 in our problem of predicting channel mismatch perceptibility, which probably meant that we had a problem with the high dimensionality of the feature space for the sampling elements.

4.3. Decision Making Trees

We also used various regression algorithms with decision making trees [2, 19] for predicting channel mismatch. We first tried to apply learning to a regression tree only. A greedy algorithm for recursive binary division [2] was used to construct the regression tree. The tree node was divided if it contained 10 or more examples of sampling. The resulting tree had 27 vertices. The error for the training sampling was 0.0897; for the cross validation error, 0.8411. We thus obtained an overfitted model.

In order to overcome overfitting, the learned regression tree was cut, and the sub-tree with the lowest cross validation error was chosen. The resulting regression tree had just 3 leaf vertices. The error for the training sampling was 0.3190; for the cross validation error, 0.6694.

The accuracy of the method based on a regression tree for the problem of channel mismatch perceptibility was thus worse than the accuracy of regression over k nearest neighbors. Regression trees, however, produce good results when used in committee methods of regression analysis in which a set of models used for solving one and the same problem is learned. One of the algorithms that uses such methods is the Random Forest algorithm [20]. Let p be the number of random features used to for division in trees, and n be the number of features. The algorithm was used with the following values of p: $p = n$ (the algorithm was reduced to bagging [21]), $p = n/2$, and $p = \sqrt{n}$. The corresponding plot is shown in Fig. 2. For $p = \sqrt{n}$ and a number of trees equal to 38, we found that the error for the training sampling was 0.1975; for the cross validation error, 0.4429.

Another committee method of regression analysis was used to predict channel mismatch perceptibility: gradient boosted regression trees [2]. This method was used with a regularization parameter equal to 0.1; the maximum number of divisions for a tree d was 1, 2, 3; and the number of trees in the model varied from 1 to 100. The plot of the cross validation error as a function of the number of trees in a model is shown in Fig. 3. The best result was obtained for a number of trees equal to 48 and a number of divisions for a tree equal to 1. The error for the training sampling was 0.1752; for the cross validation error, 0.4311.

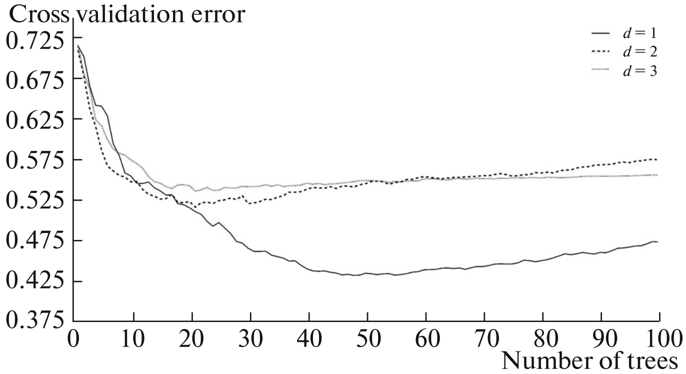

Fig. 3. Estimated accuracy of prediction for channel mismatch perceptibility using gradient boosted regression trees.

5. CONCLUSIONS

The method for finding channel mismatch described in [1] was used to analyze 105 stereoscopic films containing scenes with channel mismatch. Such scenes were used in an experimental study of channel mismatch perceptibility, and sampling for learning methods of regression analysis was performed. A number of regression algorithms were used to predict the perceptibility of channel mismatch in the scenes. The best results were obtained with gradient boosted regression trees. The average deviation from the true perceptibility values in cross validation was 0.5088, while the root mean square deviation was 0.4311. These predictions thus provide quantitative estimates of the degree of visual discomfort caused by scenes with channel mismatch and allowed us to separate scenes in which this artifact is imperceptible from those in which it causes moderate to strong discomfort.

REFERENCES

1. V. A. Lyudvichenko, S. V. Lavrushkin, V. A. Yanushkovskii, and D. S. Vatolin, “Detection of the time shift between views and channel mismatch in stereoscopic films,” *in 6th International Science and Technology Conference “Recording and Reproduction of 3D Images in Film Making and Other Fields,” Moscow, Russia, 2014* (IPP KUNA, Moscow, 2014).
2. J. Choi, D. Kim, B. Ham, S. Choi, and K. Sohn, “Visual fatigue evaluation and enhancement for 2D-plus-depth video,” in *17th IEEE International Conference on Image Processing (ICIP 2010), Hong Kong, Hong Kong, 2010 (IEEE, 2010)*, pp. 2981–2984.
3. M. Lambooij, W. A. Ijsselsteijn, and I. Heynderickx, “Visual discomfort of 3D TV: Assessment methods and modeling, Displays” **32** (4), 209–218 (2011).
4. D. Kim and K. Sohn, *Visual fatigue prediction for stereoscopic image*, IEEE Transact. on Circuits and Syst. for Video Technol. **21** (2), 231–236 (2011).
5. H. Sohn, Y. J. Jung, S. I. Lee, and Y. M. Ro, “Predicting visual discomfort using object size and disparity information in stereoscopic images,” IEEE Trans. on Broadcast. **59** (1), 28–37 (2013).
6. J. R. Quinlan, “Learning with continuous classes,” in *Proceedings of the 5th Australian Joint Conference on Artificial Intelligence (AI’92), Hobart, Tasmania, Australia, 1992* (World Sci., Singapore, 1992), pp. 343–348.
7. Y. J. Jung, H. Sohn, S. I. Lee, H. W. Park, and Y. M. Ro, “Predicting visual discomfort of stereoscopic images using human attention model,” IEEE Trans. on Circuits and Syst. for Video Technol. **23** (12), 2077–2082 (2013).
8. C. Yang, L. Zhang, H. Lu, X. Ruan, and M. H. Yang, “Saliency detection via graph-based manifold ranking,” in *26th IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2013), Portland, USA, 2013 (IEEE, 2013), pp. 3166–3173.*
9. C. C. Chang and C. J. Lin, *LIBSVM: A Library for Support Vector Machines.* http://www.csie.ntu.edu.tw/~cjlin/libsvm. Released December 14, 2015.
10. D. S. Vatolin, A. A. Voronov, V. V. Napadovskii, and A. V. Borisov, “Study of artifacts in stereoscopic films and examples of film analysis,” *Recording and Reproduction of 3D Images in Film Making and Other Fields, Moscow, Russia, 2013* (Moscow Design Bureau of Film Equipment, Moscow, 2013), pp. 190–203.
11. A. Voronov, D. Vatolin, D. Sumin, V. Napadovsky, and A. Borisov, “Methodology for stereoscopic motion-picture quality assessment,” *in Proceedings SPIE,* Vol. 8648: *Stereoscopic Displays and Applications XXIV, Burlingame, USA, 2013* (SPIE, Bellingham, 2013), pp. 864810-1–864810-14.

12. A. J. Woods, T. Docherty, and R. Koch, “Image distortions in stereoscopic video systems,” in *Proceedings of the SPIE*, Vol. 1915: *Stereoscopic Displays and Applications IV, San Jose, USA, 1993* (SPIE, Bellingham, 1993), pp. 36–48.
13. F. L. Kooi and A. Toet, “Visual comfort of binocular and 3D displays,” Displays **25** (2), 99–108 (2004).
14. T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. (Springer, New York, LLC, 2013).
15. K. Simonyan, S. Grishin, D. Vatolin, and D. Popov, “Fast video super-resolution via classification,” in *15th IEEE International Conference on Image Processing (ICIP 2008), San Diego, USA, 2008* (IEEE, 2008), pp. 349–352.
16. M. H. Kutner, *Applied Linear Statistical Models* (Chicago, 1996), Vol. 4.
17. R. Tibshirani, “Regression shrinkage and selection via the lasso,” J. R. Statist. Soc. Ser. B **58** (1), 267–288 (1996).
18. A. E. Hoerl and R. W. Kennard, “Ridge regression: biased estimation for nonorthogonal problems,” Technometrics **12** (1), 55–67 (1970).
19. L. Breiman, J. Friedman, R. Olshen, and C. Stone, *Classification and Regression Trees* (CRC Press, Boca Raton, FL, 1984).
20. L. Breiman, “Random forests,” Mach. Learn. **45**, 5–32 (2001).
21. L. Breiman, “Bagging predictors,” Mach. Learn. **26**, 123–140 (1996).

Translated by E. Baldina