WHAT'S ON TV: A LARGE SCALE QUANTITATIVE CHARACTERISATION OF MODERN BROADCAST VIDEO CONTENT

Felix Mercer Moss, Fan Zhang, Roland Baddeley and David R. Bull

Bristol Vision Institute,
University of Bristol, Senate House, Tyndall Ave, Bristol, BS8 1TH

ABSTRACT

Video databases, used for benchmarking and evaluating the performance of new video technologies, should represent the full breadth of consumer video content. The parameterisation of video databases using low-level features has proven to be an effective way of quantifying the diversity within a database. However, without a comprehensive understanding of the importance and relative frequency and of these features in the content people actually consume, the utility of such information is limited. Here, we present a large-scale analysis of programming on BBC One and CBeebies, the most popular television channels in the United Kingdom for adults and children, respectively. Twenty video features are extracted from almost three thousand television programmes shown throughout 2015 before principal components analysis is used to identify just five factors representing the most variation. The meaning and relative significance of these five factors together with the shape of their frequency distributions represent highly valuable information for researchers wanting to model the diversity of modern consumer content in representative video databases.

Index Terms— Multimedia databases, image databases, parameter estimation, feature extraction, image sequence analysis

1. INTRODUCTION

Digital video technologies are rapidly advancing. New immersive formats such as high dynamic range and high frame rate are changing the way we experience content, while the latest compression standards are making this possible at practical bit rates [1]. Before adoption on a wider scale, all new video technologies must be tested and evaluated using a relatively small pool of sequences containing representative samples of video content.

Video databases, such as VQEG-HD [2], LIVE [3], IVP [4], IVC [5] and BVI Texture [6], are used in many different contexts. Whether it be for the collection of subjective ground truth data for the validation of objective metrics or for benchmarking the performance of competing codecs, they represent an invaluable resource. However, not all video databases are equally useful. To maximise utility, they must contain a diverse and representative sample of consumer video. The process of selecting video sequences that make up these databases is, therefore, a very important one [7].

The parameterisation of video databases with respect to low-level structural information has become a convenient approach to quantifying the amount of diversity they contain. Currently, the most popular logic proposes that by maximising the range and uniformity of video features, such as edge density and motion, a database becomes more representative of real content [8]. While this approach is intuitively sensible, its validity is contingent upon two properties of the 'consumer video world' being true: (1) that it contains significant variation that the relevant features are sensitive to; and (2) that the frequency distributions of these features are uniform.

In order to build a video database that effectively models the content people consume, unprecedented analysis of a large volume of commercial video is needed. Here, we sample from recently archived content from the British Broadcasting Corporation (BBC), to identify the distribution of video features that typically appear across different programming catalogues. We then use dimensionality reduction techniques to identify a new efficient feature set that is sensitive to dimensions of most variance in the archived content.

This research was funded by the Engineering and Physical Sciences Research Council (EPSRC) and the British Broadcasting Corporation (BBC) Platform Grant EP/M000885/1

Table 1: Analysis was performed on a database of programmes sampled from BBC1 and CBeebies. Different categories contained different numbers of available programmes but were sampled equally. The Sport category contains 415 unique programmes - 65 less than the other categories due to their being fewer sports shows broadcast in the second half of 2015 on BBC One.

Genre	Channel	Available	Sampled
Drama/Comedy	BBC1	1211	480
Entertainment	BBC1	986	480
Factual	BBC1	1990	480
Sport	BBC1	445	415
News/Weather	BBC1	3157	480
Children's	CBeebies	20357	480

2. METHODOLOGY

This section details the procedure employed to sample and analyse the latent structure of archived BBC video content.

2.1. Sampling

All video content used for the original analysis was acquired from Redux, the BBC video archiving platform [9]. Redux contains all content broadcast on BBC channels since mid-2007. All videos are stored in standard definition (704×576 pixels), progressively scanned at 25 frames per second and compressed using H.264.

Practical constraints dictated that analysis was limited to television programmes broadcast in 2015, on two channels only. BBC One and CBeebies were selected as they are the most popular United Kingdom (UK) television channels for adult and child viewers, accounting for and 21.5% and 1.4% of the UK television audience, respectively [10].

Six popular genre categories, listed in Table 1, were identified that intuitively contain distinct visual qualities: five of these contained BBC One content and the sixth contained all of CBeebies (whose programmes contain no genre information). However, programming content on BBC One is not evenly distributed across genres. Forty percent of the programmes broadcast in 2015 were the *news* or *weather* while 26% were *Factual*, 15% were *Drama/Comedy* and 13% *Entertainment* shows. *Sport* programmes represented only 6 % of the programming on BBC One in 2015. Frequency of broadcasting on BBC One is not necessarily representative of the importance of a genre category worldwide, so to ensure sufficient coverage of all six content categories, each was sampled equally. To ensure equal representation across the year, programmes were also grouped by month, producing 72 uniformly sampled bins. Details of the six genre categories were obtained from Snippets, an accompanying platform to Redux that provides programming metadata.

Forty programmes were selected randomly from each of the bins and downloaded fully. This was successful for all bins apart from the *Sport* categories from June until December when less than 40 programmes were available. For each programme, five non-overlapping, two-second clips were then selected randomly and extracted. The length of the clips was kept short to reduce the likelihood of a shot transition being captured. The clips were resized to fit the display aspect ratio of 768×432 and saved in YUV 4:4:4 format. Five evenly-spaced frames were then selected from each two-second clip for analysis. In total, 2815 unique programmes were analysed, each represented by five, two-second clips.

Twenty sequences from the Bristol Vision Institute (BVI) Texture database [6] were also analysed using the chosen feature set and projected onto the first four components identified using the Redux data. Sequences from BVI are each 10 seconds long, high definition (1920×1080) and progressively scanned at 60 frames per second. Five frames were analysed for each video and the final feature value was taken as the mean over these 5 frames.

2.2. Features

In total, 20 features were calculated for each frame as listed in Table 2. Spatial information, motion vectors and colourfulness (defined in [8]) are the most common features used to quantify diversity in video databases and were used here also. Two further features were created by altering spatial information and motion vectors to produce the standard deviation of edge density and movement, respectively. Further colour information was also recorded in the mean and standard deviation for the red, green and blue channels of a frame. Temporal information features were calculated as the mean and standard deviation of absolute pixel difference between two adjacent frames. The static texture parameter is a feature that assumes static texture resides in areas dominated by high spatial frequency components. It is described formally in [11]. Here, both the mean texture and the standard deviation of the feature distribution are recorded. The dynamic texture parameter provides an estimation of complex and irregular motion between two adjacent frames and is also described formally in [11]. Here the original feature and the

Table 2: Descriptions of the original 20 features used for analysing each frame together with the rotated factor loadings for the first five components. Absolute loadings over .4 are in bold.

Feature	Description	Component				
		1	2	3	4	5
Spatial Information 1	Mean edge density calculated with Sobel operator [8]	**.90**	-.03	.22	.32	.06
Spatial Information 2	Standard deviation edge density calculated with Sobel operator	**.85**	-.04	.24	.36	.13
Colourfulness	Perceptual indicator of variety and intensity of colour [8]	.25	-.01	.29	.06	**.74**
Red 1	Mean red channel	.13	.04	.14	**.86**	-.03
Red 2	Standard deviation red channel	.26	.00	**.85**	.03	.12
Green 1	Mean green channel	.18	.07	.07	**.93**	-.08
Green 2	Stand deviation green channel	.28	.03	**.84**	.26	-.11
Blue 1	Mean blue channel	.16	-.01	.27	**.78**	.22
Blue 2	Standard deviation blue channel	.20	-.01	**.80**	.28	.16
Temporal Information 1	Mean frame difference	.02	**.95**	.00	.04	-.15
Temporal Information 2	Standard deviation frame difference	.11	**.94**	.02	.08	-.04
Motion Vectors 1	Mean motion vector [8]	-.30	**.88**	-.04	-.05	.04
Motion Vectors 2	Standard deviation motion vectors	-.32	**.84**	.02	-.07	.12
Texture Parameter 1	Static texture density [11]	**.94**	-.10	.07	.08	-.06
Texture Parameter 2	Standard deviation static texture density	**.92**	-.10	.07	.03	.21
Dynamic Texture Parameter 1	Estimates complex and irregular motion [11]	.08	**.96**	.01	.04	-.06
Dynamic Texture Parameter 2	Standard deviation dynamic texture	.09	**.91**	.01	.07	.02
Saliency	Entropy of saliency map calculated as the spectral residual [12]	**.45**	.08	.39	.00	**-.53**
Contrast 1	Mean contrast [13]	**.88**	.05	.28	.06	-.21
Contrast 2	Standard deviation	**.86**	.00	.36	.09	.08

pixel-wise standard deviation of the feature are employed. A saliency map is also computed for each analysed frame using the efficient spectral residual approach [12]. To represent the spread of the saliency map, entropy is also computed. Finally, both the mean and standard deviation of the contrast distribution is calculated using first-order directional Gaussian derivative filters, as described in [13].

2.3. Analysis

Each two-second clip was represented by a single value for each feature, which was calculated as the median over the five analysed frames. The median was chosen over the mean to prevent the combined feature output of two scenes, in the event of a clip covering a shot transition.

Before principal components analysis [14], the feature distributions were normalised using the Box-Cox power transform [15] followed by a z-score transformation. The Box-Cox transform searches for λ that maximises the log-likelihood function in:

$$data(\lambda) = \frac{data^{\lambda} - 1}{\lambda}$$

unless $\lambda = 0$, in which case the natural logarithm is used instead. The Z-score transformation was performed by subtracting the mean then dividing by the standard deviation for each feature distribution.

After principal components analysis, varimax rotation [16] was used to make the factor loadings more readily interpretable.

3. RESULTS & DISCUSSION

Of the twenty orthogonal factors identified using principal components analysis, only the associated eigenvalues of the first five exceeded a value of 1. Following the Kaiser criterion [17] the remaining 15 factors were discarded. The rotated loadings of these factors can be seen in Table 2.

While the most effective way to convey the meaning of these five factors is to present example frames from each, unfortunately legal constraints prevent this here. However, Table 3 provides a brief qualitative summary.

Factor 1 explained 28% of the variance found by the original 20 features. Frames that score low on Factor 1 are dark, and usually contain close-up or mid-range shots of people. Conversely, frames scoring high on Factor 1 are much brighter and usually contain strong graphic elements, animations or text. The Factor 1 loadings mirror these observations with strong contributions from the spatial information, contrast and, especially, static texture features. This group of features are especially sensitive to high spatial frequencies and non-natural visual elements typified in children's television, the news, weather and sports analysis. These features are insensitive to low spatial frequencies that often are more prevalent in cinematic footage shot using a shallow depth of field.

Factor 2 explained 25% of the variation found by the original 20 features. Frames with high scores for factor 2 usually contain significant motion blur indicating the second factor is most sensitive to movement. Indeed, frames scoring low for Factor 2 are news or quiz show graphics that are likely to be static, or freeze frames from sports analysis shows. Confirming this interpretation, the largest Factor 2 loadings are those sensitive to movement (TI, MV and DTP).

Factor 3 explained 14% of the variance found by the original 20 features. The most noticeable difference between frames scoring high and low in Factor 3 is overall brightness. Typically high scoring Factor 3 frames are bright and feature significant areas with homogenous colours. Low scoring Factor 3 frames are more dark with more varied content. The factor loadings for Factor 3 show strong weightings to the mean red, green and blue features which supports the idea of brightness being the principal source of variation.

Fig. 1: Two-dimensional plot of the BBC redux scores for Factors 1 and 2 (left) and Factors 3 and 4 (right). Features from the 20 videos in the BVI texture video database are overlaid to provide an indication of the database's coverage. Axes are scaled so that one unit equals one standard deviation.

Factor 4 explained 13% of the variance identified by the original 20 features. Frames scoring highly in Factor 4 are dominated by regions of both light and dark, producing images with high dynamic range, contrast and saturated colours. These frames almost always involve people, usually placed in a studio under artificial light. Frames with low scores for Factor 4 are low in contrast and saturation with the colour palette shifted noticeably toward greens and blues. Low scoring frames in Factor 4 are predominantly of children's animations, football matches or golf courses. The strongest loadings for Factor 4 are the standard deviation values for the red, green and blue features. Perhaps surprisingly, the contrast features do not make large contributions to this factor. This is likely to be due to the contrast features identifying the average contrast across the frame, as opposed to the maximum. The contrast in these frames is infrequent but strong, such as an illuminated figure in a studio in front of a dark background. A maximum but not average contrast feature would be sensitive to such variation.

Finally, Factor 5 explained 5% of the variance identified by the original 20 features. Frames producing high scores in Factor 5 are all highly saturated. In contrast, low scoring Factor 5 frames contain dull colours or are even completely desaturated. As expected, the strongest positive loading for Factor 5 is colourfulness. Unexpectedly, however, saliency entropy produced a strong negative loading on Factor 5. This may be an indication that Factor 5 is associated more with 'studio television' than simply saturated colour. Programmes shot in the studio are not only very colourful but also usually contain a single centrally-positioned object, such as a presenter's head, in front of plain background. Such a scene would produce a narrow saliency map with low entropy, explaining its negative loading on Factor 5.

The scores for Factors 1 and 2 are plotted in the left plot of Figure 1 while those of Factors 3 and 4 are plotted in the right plot of Figure 1. Scores from the BVI texture database overlaid over the Redux data provide an indication of how well they cover the space occupied by Redux. It is important to remember that while the plots are an accurate representation of the *range* of the new features, the shape of the distributions are not truly representative of BBC One and CBeebies. For that to be the case the content categories should be weighted by their frequencies on BBC channels. The extent to which the model should be influenced by the frequency of programmes is currently not clear. However, by using a uniform genre prior, the model presented here ensures that it captures more variation in consumer television content in general, and is influenced less by programming schedules.

Table 3: Qualitative description of the first five factors representing the most variation in Redux.

Factor	Positive Frames	Negative Frames
1. Naturalness	Graphics, text, animations, bright	Close-up faces, dark, dramatic
2. Movement	Motion blur, sports, animations	Single objects, text graphics,
3. Brightness	Bright, animations, uncluttered	Dark, text graphics, cluttered
4. Contrast	High contrast, saturated colours	Bright, pastels, low contrast, animations
5. Saturation	Saturated colours, studio, uncluttered	Low saturation, dramatic, cluttered

4. CONCLUSIONS

To create video databases that capture the structural variation contained in modern consumer video, it is important to identify features that are sensitive to that variation. Here, we extract 20 video features from over 2500 hours of television programming from two of the most popular channels in the United Kingdom. We reduce the dimensionality of our feature set to just five, accounting for 85% of the variation found in the original 20 features. Diversity in *naturalness*, *movement*, *brightness*, *contrast* and *saturation* - in that order of importance - directly contributes towards producing a video database that is representative of BBC content in 2015.

While the work presented here represents a significant early step in the low-level modelling of consumer television content, a year of BBC programming should not pretend to represent the complete picture. Similar analyses using content from different networks around the world (including those published solely online) will help contribute to a more complete understanding of what consumer video looks like. Further work will also explore the identification of independent feature profiles across both professional and amateur content categories.

The procedure of sequence selection for video databases is currently ill-defined. Future research will use the findings presented here to evaluate the utility of more existing video databases while informing researchers how to make their future databases more representative of consumer content.

5. REFERENCES

[1] D R Bull, *Communicating pictures: A Course in Image and Video Coding*, Academic Press, Oxford, UK, 1 edition, 2014.

[2] VQEG, "Report on the Validation of Video Quality Models for High Definition Video Content," Tech. Rep., June 2010.

[3] K Seshadrinathan, A C Bovik, and L K Cormack, *LIVE Video Quality Database*, 2010, `http://live.ece.utexas.edu/research/quality/live_video.html`.

[4] F Zhang, S Li, L Ma, Y C Wong, and K N Ngan, *IVP subjective quality video database*, `http://ivp.ee.cuhk.edu.hk/research/database/subjective/`.

[5] S Péchard, R Pépion, and P Le Callet, *IRCCyN/IVC Videos 1080i Database*, `http://www.irccyn.ec-nantes.fr/~lecallet/platforms.htm`.

[6] M A Papadopoulos, F Zhang, D Agrafiotis, and D R Bull, "A video texture database for perceptual compression and quality assessment," in *Image Processing (ICIP), 2015 IEEE International Conference on*. IEEE, 2015, pp. 2781–2785.

[7] M H Pinson, M Barkowsky, and P Le Callet, "Selecting scenes for 2D and 3D subjective video quality tests," *EURASIP Journal on Image and Video Processing*, vol. 2013, no. 1, pp. 50, Aug. 2013.

[8] S Winkler, "Analysis of Public Image and Video Databases for Quality Assessment," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 6, no. 6, pp. 616–625, Oct. 2012.

[9] B Butterworth, *History of the 'BBC Redux' project*, 2008 (accessed January 30, 2016), `http://www.bbc.co.uk/blogs/bbcinternet/2008/10/history_of_the_bbc_redux_proje.html`.

[10] Broadcasting Audience Research Board, *BARB 2015 Viewing Summary*, 2015 (accessed January 30, 2016), `http://www.barb.co.uk/whats-new/monthly-viewing-summary?_s=4`.

[11] F Mercer Moss, K Wang, F Zhang, R J Baddeley, and David R Bull, "On the Optimal Presentation Duration for Subjective Video Quality Assessment," *IEEE Transactions on Circuits and Systems for Video Technology*, 2015.

[12] X Hou Hou and L Zhang, "Saliency Detection: A Spectral Residual Approach," *Journal of Applied Computational Harmonics Analysis*, pp. 1–8, June 2007.

[13] H S Scholte, S Ghebreab, L Waldorp, A W M Smeulders, and V A F Lamme, "Brain responses strongly correlate with Weibull image statistics when processing natural images.," *Journal of Vision*, vol. 9, no. 4, pp. 29.1–15, 2009.

[14] I T Jolliffe, *Principal Component Analysis*, Springer Series in Statistics. Springer Science & Business Media, New York, May 2006.

[15] G E P Box and D R Cox, "An Analysis of Transformations on JSTOR," *Journal of the Royal Statistical Society Series B*, 1964.

[16] H F Kaiser, *The Varimax Criterion for Analytic Rotation in Factor Analysis*, Tratto Da Psychometrika Vol. 23 N. 3 September 1958.

[17] H F Kaiser, "Varimax solution for primary mental abilities," *Psychometrika*, vol. 25, no. 2, pp. 153–158, 1960.