New Trends and Issues Proceedings on Humanities and Social Sciences

Volume 4, Issue 4 (2017) 12-20

ISSN 2421-8030
www.prosoc.eu

Selected Papers of 6th Cyprus International Conference on Educational Research (CYICER-2017), 04-06 May 2017, Acapulco Hotel and Resort Convention Center, North Cyprus.

Model of information density measuring in e-learning videos

Kristian Dokic [a]*, Department of Social Work, Faculty of Applied Sciences, Polytecnic University, 34000, Pozega, Croatia.

Robert Idlbek [b], Department of Social Work, Faculty of Applied Sciences, Polytecnic University, 34000, Pozega, Croatia.

Dubravka Mandusic [c], Department of Social Work, Faculty of Applied Sciences, Polytecnic University, 34000, Pozega, Croatia.

Suggested Citation:

Dokic, K., Idlbek, R. & Mandusic, D. (2017). Model of information density measuring in e-learning videos. *New Trends and Issues Proceedings on Humanities and Social Sciences.* [Online]. 4(4), pp 12-20. Available from: www.prosoc.eu

Selection and peer review under responsibility of Assoc. Prof. Dr. Cigdem Hursen, Near East University

Abstract

Educational video content authors must be careful that the content is adequate for the audience and the purpose especially on the term level. On the other hand authors must be careful that the content does not cause information overloading. Viewers can experience it if the videos have too much motions and movements. Information overload has become an increasingly popular area of study and a lot of authors indicated that information density of movies and videos have risen in the last few decades. High information density of videos and movies may cause information overloading and it has been proved by Lang et al in their theoretic model named Limited Capacity Model of Motivated Mediated Message Processing. This theoretic model suggests how to measure information density in videos and movies but this method is slow and just trained observers can do this job well. In this paper the new model for information density measuring in videos and movies is described and tested. This new model is based on special usage of background subtraction algorithms and in this paper multi-layer background subtraction algorithm based on color and texture is used. This algorithm is presented by authors Yao and Odobez on CVPR Visual Surveillance Workshop in 2007. Results of our model are compared with the trained observer's results and the same video clips that have been used by Lang et al are used for testing. It can be seen that there is a significant correlation between information densities measured by trained observers and our model. Our model can be used to determine some forms of information density in videos and movies and it can be used to indicate higher possibility of information overloading.

Keywords: Information density, visual activity, e-learning.

* ADDRESS FOR CORRESPONDENCE: **Kristian Dokic**, Department of Social Work, Polytecnic University, 34000, Pozega, Croatia.
E-mail address: kjdokic@vup.hr / Tel.: +385994503857

1. Introduction

Information density is a multifaceted term. Most of the researchers use it to measure a bit of density on magnetic or optical surfaces. This term can be found in the first papers about blue laser diodes used for digital recording discs. Those new discs have much bigger density then well-known CD or DVD systems (Schep, 2001). Spronck et al. (2004) wrote about the urgent need of increasing the information density on optical media when they described technologies that have been used in Blu-ray^{TM} disc. The term Information density can be found in the Principle of Constant Information Density concerning interfaces and maps optimization. This Principle is about the amount of information that should remain constant as the user zoom or pan (Frank, 1994). Jeffery wrote that “a high information density allows more information in the available space, or a given amount of information in a smaller space”. His field of research is screen usage in graphical interface programming (Jeffery, 2012).

One definition of the term Information density can be found in Lang’s papers where she defined it as information introduced in video per second. She used this term as a part of her theoretic model named, ‘The limited capacity model of mediated motivated message processing (LC4MP)’ and it is used in our paper as a stronghold (Lang, 2000). This term is in the subject of our paper and it is important especially in the field of educational videos because if we do not care about Information density of videos, it may cause viewers information overload. Lang’s LC4MP theory suggests that the result of information overload is poor encoding of the message so learning will suffer (Lang, 2007).

Video has been used in classrooms for decades. Hovland and colleagues (1949) wrote about filmstrips usage during World War II as a training tool for soldiers. Authors wrote that audio-video materials had been recognized as a good way to capture the attention of learners and enhance their experience. Deshpande (2001) described interactive virtual classrooms with two way audio and video streams and interaction before more than 15 years. He described issues that can be classified at 2 levels: technology and effectiveness. It is obviously that as technology has developed most problems with infrastructure and audio/video coders/decoders have disappeared. It is pointed out that recognition challenges of teaching topics in e-learning videos and their indexing are important. They asserted that instructional videos are different from other video genres and they wrote about differences and special issues (Lui & Kender 2004).

The paper about videos in LMS deals with the problem of e-learning videos speed. Song and colleagues realized that current technology allows manual video speed change but they suggested that there is a need for more dynamic control systems that are driven by viewer’s activities. They described a novel system based on a head tracking device that gives users the ability to control the pace of video clips based on the head position. This paper is interesting because it deals with a problem of video speed and the need to change it (Song et al., 2015). Our paper deals with information density in videos and this is somehow similar because high information density is sometimes a reason for slowing down video speed. In our paper a new approach for information density measuring will be presented.

2. Literature Overview

In the beginning of literature overview one paper that has had a big impact on the media will be presented. It is a Lillard and Petersons’ paper that analyzes the impact of different types of television programs on the executive functions of preschool children. The preschool children have been divided into three groups. The first group watched a dynamic (fast-paced) television cartoon, the second group watched a less dynamic (slow-paced) educational cartoon and the third group drew for 9 minutes. After that, all three groups have been tested of executive function. The results indicated the group that had watched a dynamic television cartoon solved tests significantly worse than the other two groups. Lillard and Peterson qualitatively described the differences between dynamic and educational animated film, and they did not use quantitative measures to show this difference (Lillard et al., 2011).

Many authors used the term pacing for video clips and television dynamic, but this term was interpreted in different ways. Watt and colleagues defined it as the frequency of verbal statements and background change, while the first author a decade later defined pacing as differences in the level of light on the screen over time (Watt et al., 1974). Cooper (2009) and colleagues interpreted pacing as a frequency of camera angle changes, while Lang and colleagues defined it as a number of transitions from one scene to another and they have quantified that number on three levels and (Lang et al., 1999). Anderson and colleagues used the same concept in their research and classification of dynamics or tempo based on: the frequency of camera movements and video processing, frequency of changes to a new scene, percentage of live music, percentage of movement activities and the length of the frame (Anderson et al., 1977). Huston and colleagues defined pacing as the frequency of scenes and character changes (Huston et al., 1981). Apparently there is no consensus among the authors and everyone interprets the term differently, although it is obvious that the descriptions are similar.

2.1. Average shot length

Most of the authors suggest that the value of pacing should be measured and categorized by a trained observer. In most situations subjectivity cannot be avoided in measurement or categorization. Most of the authors recognize that the length of the shot is associated with a concept of pacing. Average duration of the shot (Average Shot Length) is often mentioned in the literature. Bordwell (2002) provided an overview of the average duration of the shot for the movies produced in the last few decades. ASL was between 8 and 11 seconds between 1930's and 1960's. At the end of the sixties ASL was between 6 and 8 seconds. It slightly decreased in seventies to values between 5 and 8 seconds. At the turn of the century ASL was between 3 and 6 seconds. The author stated that in the last century the lowest ASL was 1.8 seconds. This paper dealt with movies produced in the United States (Bordwell, 2002).

Palmer analyzed advertisements for the movies (trailers) in the last seventy years. She measured the number of shots in a minute and she also noticed a constant increase of shots numbers in recent years. It was 12 in the fifties, but it increased to 38 in the nineties. To calculate ASL from the numbers presented by the author, 60 must be divided by the number of shots in the minute (Palmer, 2013).

Cutting and colleagues pointed out trends that they had observed in their sample of 150 movies. The ASL was bigger than 10 seconds in the middle of the last century but it fell to below 5 seconds at the end of the century. The authors stated that the same phenomenon had been observed and analyzed by the author Salt. He used a much larger sample of about 13,000 movies. Salt pointed out that ASL was 12 in the middle of the 20^{th} century, but it fell to 5 seconds at the end of the nineties. Both results can be seen on the chart on the figure number 1 (Cutting, 2011) and (Salt, 2006).

2.2. Visual activity index

Cutting and colleagues set up an advanced model to measure visual activity within the frame. They used Gibson's definitions for motion and movement to define visual activity in their paper. Gibson defined motion as a change of objects or people in the shot with a fixed background. He also defined movement as visual information caused by the movement of the observer or a camera in the case of the movie (Gibson, 1954). Cutting and colleagues defined visual activity as a sum of activities described as motion and movement. Authors calculated correlations between same pixels in every adjacent frame of a movie and then they calculated the median for that sequence. The last step is subtracting the median from the number one and the result is the Visual Activity Index. When a video clip is static without motion and movement then VAI is zero. VAI's greatest theoretical value is 2. Authors analyzed 145 movies and they provided a chart with average VAI values for the movies produced between 1935 and 2005. It can be seen that average VAI values between 1935 and 2005 increased about 3 times (Cutting, 2011).

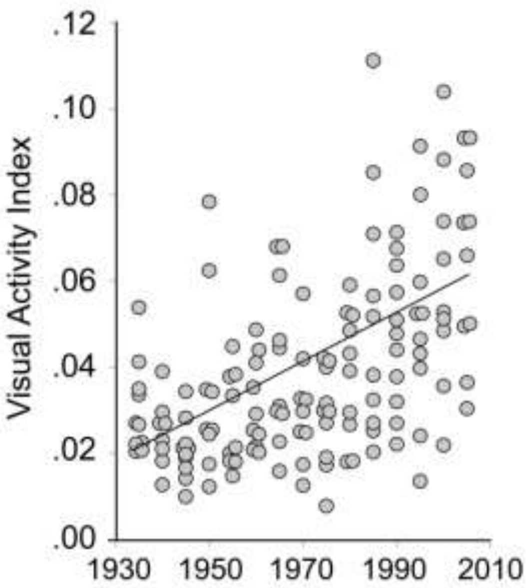

Figure 1. A plot of the grouped mean average shot lengths (ASLs) (Cutting, 2011)

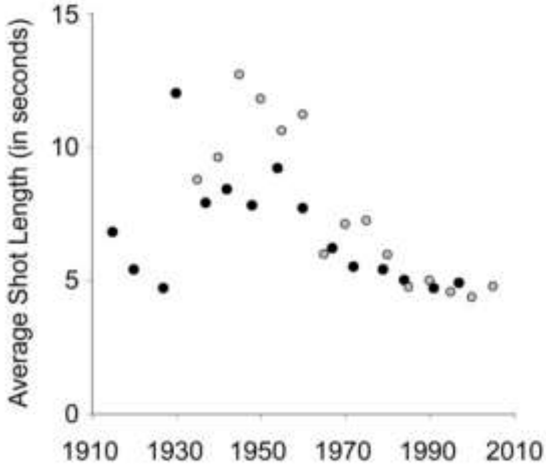

Figure 2. A scatter plot of whole-film visual activity indices (VAIs) by year for 145 films from 1935 to 2005. (Cutting, 2011)

Authors Cutting and DeLong are the most prominent authors in the field of analysis of visual activities in the media. Their model called the Visual Activity Index is the most comprehensive measure of visual activity. The authors suggest that there is a significant negative correlation between ASL and VAI values in the analyzed movies, and that r = -0.46 and r = -0.55 on logarithmic scale.

Visual Activity Index may not be always the best solution for visual activity quantification. First of all the model completely excludes viewers of media content. In some situations it is possible that the VAI value is quite high, but the viewers of media content do not even notice that. An example of this can be a video clip of wood where the leaves move on the wind. The second objection is related to noise sensitivity of the VAI model. Noise can affect a VAI value, but a viewer of media content might not notice activity at all.

2.3. Computer vision

Solutions of these potential shortcomings could be found in the field of computer vision. Jahne and Haußecker defined computer vision as "a set of techniques for the collection, processing, analysis and understanding of complex multidimensional data from our environment for technical and scientific applications" (Jahne et al., 2000). A lot of algorithms have been developed in the field of computer vision and one subgroup is used to remove background from video streams. Their name is background subtraction algorithms. The first algorithms for background subtraction was presented in the seventies (Jain, 1979), but today there are lot of algorithms available. Most of them are developed for specific applications. Algorithms handle color video content in a way that the output of processing is black and white videos where one color presents detected moving objects and the other presents a stationary background. The ratio of the number of pixels detected as moving objects and the total number of pixels in a video clip can be quantified as the amount of motion in video content and this is the key idea of this paper. This approach should be more robust in the analysis of video with a substantial amount of noise if algorithms that are used are resistant to noise. It can be expressed with this formula:

$$nVAI = \sum n_{fg} \;/ \sum n_{all} \tag{1}$$

Where nVAI is an abbreviation for new Visual Activity Index, n_{fg} is the number of foreground pixels in one frame and n_{all} is the number of all pixels in one frame (Dokic, 2015).

This approach has been used in our previous papers for motion quantification in a video content. The questions that arise are which of the tens of algorithms provide the best results in detection and how to define criteria of choosing the appropriate algorithm. These algorithms are mostly developed for the industry and for some specific purpose, and therefore they may not be the most suitable for our purpose. The logical way is to compare this model with some empirically proven model in the field of communication science and fortunately there is the theory called ‘Limited Capacity Model of Motivated Mediated Message Processing’ with well elaborated model (Lang et al., 2007).

2.4. LC4MP

This model assumed that recipients of audio visual content have limited resources to process information, and when all resources are used then information overload will occur. The process of resource allocation is complex and depends on many factors, and it takes place automatically or is controlled by the recipient. LC4MP model introduced new features like "Complexity of structure" (number of scene changes per second) as a measure of allocated resources and the "Information density" (newly introduced information) as a measure of necessary resources (Lang et al, 2007). To obtain the value of the variable "Complexity of structure" there are a number of automated methods well described in the literature (Hanjalic, 2002) and (Boreczky, 1996). To obtain the value of the variable "Information density" a trained observer has to analyze video content. (Lang et al, 2007). LC4MP and its empirical evidence (data) will be used to analyze one background algorithm in this paper.

3. Data and Methodology

In this paper 135 videos from Lang’s experiment were used. Those videos were used in her paper “*Cognition and Emotion in TV Message Processing: How Valence, Arousing Content, Structural Complexity, and Information Density Affect the Availability of Cognitive Resources*”. She introduced Information density in theory as a measure of necessary resources for video processing. This variable consists of seven factors and they are: object change, novelty, relatedness, distance, perspective, emotion and form change. Trained observers can measure them at the time of the video clip cut and those factors can be used to define when information overload will occur. Lang had used several methods to define the moment of information overload and she had good results with the method called ‘Second Task Response Time’. When the number of introduced information rises and time defined by STRT starts to fall then information overload occurred. Lang's data is interesting because our method can be proven without conducting research with people (Lang et al., 2007).

There are lots of background subtraction algorithms available and Sobral has divided them into 9 groups:

1) Basic methods.
2) Fuzzy based methods.
3) Single Gaussian based methods.
4) Multiple Gaussians based methods.
5) Type-2 Fuzzy based methods.
6) Multiple features based methods.
7) Non-parametric methods.
8) Subspace-based methods.
9) Neural and neuro-fuzzy methods (Sobral, 2013).

In this paper the algorithm called 'Multi-Layer Background Subtraction Based on Color and Texture' was used. Algorithm is from Sobral's group 'Multiple features based methods'. The algorithm was not analyzed in this paper and it was just used to subtract the background from Lang's videos. After subtraction our method previously described and expressed in formula number 1 was used. Finally, Spearman correlation was used to find correlation between Lang's "Information density" factors and our nVAI values (Yao et al., 2007).

4. Results

Complete results can be found on web page http://kristiandokic.from.hr/?page_id=69 and here is the table with Spearman's correlation values.

Table 1. Spearman's correlation values

Information density factor	Spearman's correlation	Correlation category
emotion	0,388499023	low
novelty	0,490135858	moderate
relatedness	0,452651034	moderate
object change	0,590530214	moderate
distance	0,612926829	high
perspective	0,697694303	high
form change	0,262696502	low

It can be seen that there are high correlations between nVAI and two factors, distance and perspective. On figure 3 those correlations are compared with results from another paper that had dealt with another background subtraction algorithm – temporal median algorithm by authors (Calderara et al, 2006; Cucchiara et al., 2003).

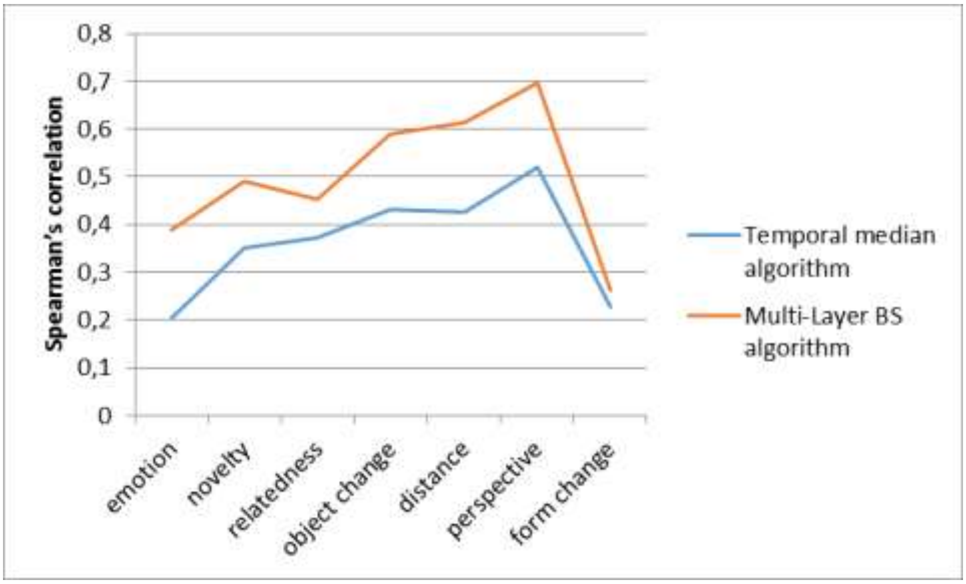

Figure 3. Spearman's correlations of Information Density factors and nVAI for 2 BS algorithms

It can be seen that there are correlations between our model called nVAI and all Lang's "Information density" factors but correlations with 'Multi-Layer Background Subtraction Based on Colour and Texture algorithm' values are bigger than values from previously used algorithm. Results indicate that our model can be used to replace some Lang's Information density factors, but not all. It is obvious that there is a weak correlation between nVAI and Lang's factors: emotion and form change.

5. Conclusion

There are lots of evidences that some videos cause information overload. If the video clips are on Learning Management System care must be taken to ensure that those clips do not have high information density. In this paper authors presented the new model for information density measurement that is based on computer vision algorithms. This new model is compared with Lang's Information density factors and it can be seen that there are correlations between our model and Lang's factors. Problem with Lang's factors is that only trained observers can measure them and it can be a big problem if there are a lot of videos on Learning Management System. The new model presented in this paper is the first step of automation of information density measuring. There's a lot of work to be done to finish this model. All background algorithms have to be tested and the best of them have to be prepared as a plug-in for most common Learning Management Systems. Finally, if we want to achieve maximum efficiency in e-learning, we have to personalize information density measuring algorithms but there is a long way to reach that goal.

References

Anderson, D. R., Levin, S. & Lorch, E. (1977). The effects of TV program pacing on the behavior of preschool children. *Educational Technology Research and Development, 25,* 159–166.

Bordwell, D. (2002). Intensified continuity - visual style in contemporary American film. *Film Quarterly, 55*(3), 16-28.

Boreczky, J. S. & Rowe, L. A. (1996). Comparison of video shot boundary detection techniques. *Journal of Electronic Imaging, 5*(2), 122-129.

Calderara, S., Melli, R., Prati, A. & Cucchiara, R. (2006) *Reliable background suppression for complex scenes.* VSSN '06 Proceedings of the 4th ACM international workshop on Video surveillance and sensor networks (pp. 211 -214).

Cooper, N., Uller, C., Pettifer, J. & Stolc, F. (2009). Conditioning attentional skills: Examining the effects of the pace of television editing on children's attention. *Acta Paediatrica, 98,* 1651–1655.

Cucchiara, R., Grana, C., Piccardi, M. & Prati, A. (2003). Detecting moving objects, ghosts, and shadows in video streams. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 25*(10), 1337-1342.

Cutting, J. E., DeLong, J. E. & Brunick, K. L. (2011). Visual activity in Hollywood film: 1935 to 2005 and beyond. *Psychology of Aesthetics, Creativity, and the Arts, 5*(2), 115-125.

Deshpande, S. G. & Hwang, J. N. (2001). A real-time interactive virtual classroom multimedia distance learning system. *IEEE Transactions on Multimedia, 3*(4), 432-444.

Dokic, K., Idlbek, R. & Raguz, M. J. (2015). *Measuring visual activity in a video content based on computer vision temporal median algorithm.* In The 4th Palestinian International Conference on Computer and Information Technology (pp. 03-15).

Frank, A. U. & Timpf, S. (1994). Multiple representations for cartographic objects in a multi-scale tree—an intelligent graphical zoom. *Computers & Graphics, 18*(6), 823-829.

Hanjalic, A. (2002). Shot-boundary detection: Unraveled and resolved?. *IEEE Transactions on Circuits and Systems for Video Technology, 12*(2), 90-105.

Hovland, C. I., Lumsdaine, A. A. & Sheffield, F. D. (1949). *Experiments on mass communication*. Princeton, NJ: Princeton University Press.

Huston, A. C., Wright, J. C., Wartella, E., Rice, M. L., Watkins, B. A. & Campbell, T. (1981). Communicating more than content: Formal features of children's television programs. *Journal of Communication, 31,* 32–48.

Jahne, B. & Haussecker, H. (2000), *Computer vision and applications - a guide for students and practitioners.* US: Academic Press.

Jain, R. & Nagel, H. H. (1979). On the analysis of accumulative difference pictures from image sequences of real world scenes. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 1*(2), 206-214.

Jeffery, C. L. (2012). *Program monitoring and visualization: An exploratory approach*. US: Springer Science & Business Media.

Joshua, F., Franklin, D. & Hammond, K, (2000). *Improving human computer interaction in a classroom environment using computer vision.* In Proceedings of ACM Multimedia (pp. 10-17).

Schep, K., Stek, B., Van-Woudenberg, R., Blüm, M., Kobayashi, S., Narahara, T. & Ogawa, H. (2001). Format description and evaluation of the 22.5 GB digital-video-recording disc. *Japanese Journal of Applied Physics, 40*(3), 1813.

Lang, A., Bolls, P., Potter, R. F. & Kawahara, K. (1999). The effects of production pacing and arousing content on the information processing of television messages. *Journal of Broadcasting & Electronic Media, 43*(4), 451-475.

Lang, A. (2000). The limited capacity model of mediated message processing. *Journal of Communication, 50,* 46–70.

Lang, A., Park, B., Sanders-Jackson, A. N., Wilson, B. D. & Wang, Z. (2007). Cognition and emotion in TV message processing: How valence, arousing content, structural complexity, and information density affect the availability of cognitive resources. *Media Psychology, 10*(3), 317-338.

Lillard, A. S. & Peterson, J. (2011). The immediate impact of different types of television on young children's executive function. *Pediatrics, 128,* 644–649.

Liu, T. & Kender, J. R. (2004). *Lecture videos for e-learning: Current research and challenges*. In Multimedia Software Engineering, 2004. Proceedings. IEEE Sixth International Symposium on (pp. 574-578).

Salt, B. (2006). *Moving into pictures*. London: Starword.

Sobral, A. (2013). BGSLibrary: *An OpenCV C++ Background Subtraction Library*. In IX Workshop de Visão Computacional (WVC'2013), Rio de Janeiro, Brazil (pp. 20-27).

Song, S., Hong, J. K., Oakley, I., Cho, J. D. & Bianchi, A. (2015). *Automatically adjusting the speed of e-learning videos.* In Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems (pp. 1451-1456).

Spronck, J. W., El-Husseini, M. H., Jabben, L., Overschie, P. M., Hobbelen, D. G. E., Du-Pau, P. & Van-Eijk, J. (2004). Mastering high-density optical disks: a new concept design. *Assembly Automation, 24*(4), 406-415.

Strunk, W. & White, E. B. (1979). *The elements of style.* (3rd Ed.). New York: Macmillan.

Vander-Geer, J., Hanraads, J. A. J. & Lupton R. A. (2000). The art of writing a scientific article. *Journal of Scientific Communications*, *163,* 51-59.

Watt, J. H. & Krull, R. (1974). An information theory measure for television programming. Communication *Research, 1,* 44–68.

Watt, J. H. & Welch, A. J. (1983). *Effects of static and dynamic complexity on children's attention and recall of televised instruction.* In J. Bryant & D. R. Anderson (Eds.), *Children's understanding of television: Research on attention and comprehension* (pp. 69–102). New York: Academic Press.

Yao, J. & Odobez, J. M. (2007, June). *Multi-layer background subtraction based on color and texture*. In Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on (pp. 1-8).