International Journal of Research in Marketing 36 (2019) 216–231

Contents lists available at ScienceDirect

IJRM

International Journal of Research in Marketing

journal homepage: www.elsevier.com/locate/ijresmar

Full Length Article

Video mining: Measuring visual information using automatic methods

Xi Li [a], Mengze Shi [b], Xin (Shane) Wang [c,*]

[a] City University of Hong Kong, Kowloon Tong, Hong Kong
[b] Rotman School of Management, University of Toronto, Canada
[c] Ivey Business School, Western University, London, Ontario N6G 0N1, Canada

ARTICLE INFO

Article history:
First received on January 9, 2018 and was under review for 7 months
Available online 25 March 2019

Area Editor: Michael Haenlein

Keywords:
Video mining
Visual information
Visual variation
Video content

ABSTRACT

Marketers are becoming increasingly reliant on videos to market their products and services. However, there is no standard set of measures of visual information that can be applied to large datasets. This paper proposes two standard measures that can be automatically obtained from videos: visual variation and video content. The paper tests the measures on crowdfunding videos from a leading online crowdfunding website, and shows that the proposed measures have explanatory power on the funding outcomes of the projects. These measures can be effectively implemented and used for large datasets. Further, researchers can apply these measures to other sets of visual information, and marketers could use the research to guide their video design and improve their video marketing effectiveness.

1. Introduction

Content marketing is becoming increasingly visual. With advances in information and media technologies, visual information is becoming more and more prevalent in online markets, and companies are relying more than ever on online videos to introduce, promote, and advertise their products and services. A recent report by Wochit finds that 88% of firms will increase their video spending in 2018 (Dreier, 2017), while by 2021, according to a study conducted by Cisco, 82% of all Internet traffic will be in the form of videos (Business Insider Intelligence, 2017).

Together with text and audio information, visual information is considered as a form of unstructured data (Sudhir, 2016). Today, unstructured data plays a key role in the consumer decision-making process. It is estimated that 80% of the data currently held by firms is unstructured (Rizkallah, 2017), including documents, messages, social media posts, pictures, videos, and audio. Unstructured data is growing 15 times faster than structured data (Nair & Narayanan, 2012). CIKLUM (2017) predicts that by 2022, 93% of all data in the digital universe will be unstructured.

While unstructured data is becoming increasingly common and available to firms and marketers, it is still underutilized in both academic research and industrial practice. For example, according to Howatson (2016), 87% of marketers "see data as their most underused asset" and realize that most of the massive number of unstructured information sites are poorly managed and largely untapped. Unstructured data is notoriously difficult to make sense of because it is not necessarily organized in a way that can be easily processed (Dorian, 2017).

[*] Corresponding author.
E-mail address: xwang@ivey.ca. (X.(S.) Wang).

https://doi.org/10.1016/j.ijresmar.2019.02.004
0167-8116/© 2019 Elsevier B.V. All rights reserved.

The sensemaking analysis of unstructured data is challenging for several reasons. First, unstructured data takes many different forms and must be cleaned and prepared before usage. As mentioned by Jude (2016), "Companies contemplating big data must first assume their data is in serious need of preparation." Preparing unstructured data is often time consuming and expensive, in part because it cannot just be done once: it must be done every time a new dataset is introduced to the company. Second, the massive volume of unstructured data means that highly efficient, automatic, and scalable methods must be developed. In the past, marketing researchers used controlled lab experiments to analyze small volumes of unstructured data (e.g., television commercials). Though this approach has been quite successful in the past, it is becoming prohibitive considering the sheer amount of unstructured data generated; for example, it is estimated that every 60 s, 300 h of video are uploaded to online video site YouTube (Flahive, 2017). Even for the same measures, researchers need to come up with new techniques that allow for fast and cheap feature extraction. Third, extracting *meaningful* information from unstructured data is a complex task (e.g., How can we extract human emotions from text and videos? What causes the emotions?). The context dependence of human behavior can further complicate the issue.

Due to the above challenges, analysis of unstructured data has been limited. Among all types of unstructured data, text information is perhaps the most widely studied. Existing studies have analyzed product reviews and consumer messages from forums, social tags, and tweets in different contexts using different methods (Klostermann, Plumeyer, Böger, & Decker, 2018). In the early stages of this research, researchers relied on simple measures of text information, such as review length and volume, or used human raters to evaluate the valence of review text (Godes & Mayzlin, 2004). Later, researchers introduced several machine learning algorithms to marketing and applied them to text mining. For example, Tirunillai and Tellis (2012) use two automatic methods, Naïve Bayesian and Support Vector Machines (SVM), to classify text reviews into two classes based on their valence (i.e., positive or negative). Tirunillai and Tellis (2014) implement latent Dirichlet allocation (LDA), an unsupervised machine learning algorithm, to automatically classify user-generated content on product reviews into several topics. Going further, Liu, Singh, and Srinivasan (2016) use principal component analysis (PCA) to extract several key components from movie reviews. Today, text mining is widely used to extract meaningful information from unstructured text (Dörre, Gerstl, & Seiffert, 1999; Feldman & Sanger, 2007), and is commonly used to analyze customer reviews and user-generated content (Archak, Ghose, & Ipeirotis, 2011; Ghose, Ipeirotis, & Li, 2012; Lee & Bradlow, 2011; Netzer, Feldman, Goldenberg, & Fresko, 2012). Thanks to these machine learning algorithms, researchers are now able to analyze millions of pieces of user-generated content in a single study.

While text information has been widely studied and used, academic research has lagged in analyzing visual information and provides little guidance on how to design an effective online video. Unlike text information, visual information must be processed with advanced information technologies, which were not available in the past. Early studies of visual information used controlled lab experiments to study print or television advertising. For instance, Rethans, Swasy, and Marks (1986) use lab experiments to analyze how the length and repetition of television commercials affect consumers' attitudes toward the commercials. Similarly, Gorn, Chattopadhyay, Yi, and Dahl (1997) use lab experiments to analyze the effect of colors (i.e., hue, chroma, and value) on consumers' feelings and attitudes toward print advertising. With the development of new technologies, researchers also resort to innovative techniques to obtain and analyze visual data. For example, eye tracking helps researchers understand consumers' entire decision journey during a store visit (see Wedel & Pieters, 2008 for a comprehensive review of eye tracking techniques in marketing research). However, despite its success in providing visual information to researchers which can hardly be obtained otherwise, eye tracking is still costly and not scalable.

To date, very few studies have adopted automatic methods to analyze visual information. Among these rare studies, Xiao and Ding (2014) explore the effects of facial features in print advertising. The authors take human faces as input, and construct a few "eigenfaces" from the faces. They then use PCA to analyze how these eigenfaces affect consumers' attitudes toward different products. Zhang, Lee, Singh, and Srinivasan (2017) construct several simple measures from Airbnb photos, such as hue and brightness, and investigate how these photo characteristics affect the property demand at Airbnb. Klostermann, Plumeyer, Böger, and Decker (2018) use Google Cloud Vision API to classify brand-associated images on social networks into several categories. They suggest that marketers should infer what consumers think and feel about the brand from the images that consumers post.

While all of the above studies focus exclusively on static images, Zhang, Wang, and Chen (2018) take videos into account. They analyze the shot length, camera motion, sound loudness, and sound pitch of videos, and then combine the data with live comments to construct a measure called "moment-to-moment synchronicity" (MTMS). In Table 1, we summarize the recent literature on unstructured data in marketing and other related fields. We also refer readers to Balducci and Marinova (2018) for a comprehensive review of unstructured data in marketing.

Despite the rise in studies analyzing unstructured data in the past few years, no tool or set of visual features has emerged as a standard, particularly for analyzing real-world visual data. As noted above, past research on visual information has focused on identifying the effectiveness of video advertising using controlled lab experiments. Analyzing videos is considered extremely costly and time consuming, and is thus limited to small-scale analysis.

The recent development of information technologies has made it possible for researchers and marketers to analyze visual information using automatic methods. Most notably, in the past a few years, the use of convolutional neural networks (CNN), a machine learning algorithm, has proven effective in analyzing visual information. CNN has been used to this effect in tasks such as autonomous driving and facial recognition. Inspired by these industrial practices, we attempt to introduce some of the same methods into marketing, and show how to extract meaningful measures from visual information using these techniques.

Table 1
Overview of literature on unstructured data.

	Type of data	Measures	Computational methods	Main findings
Godes and Mayzlin (2004)	Text	Volume, valence and length		A measure of the dispersion of conversations across communities has explanatory power in a dynamic model of television ratings.
Chevalier and Mayzlin (2006)	Text	Length		The length of reviews has an effect on product sales.
Mishne and Glance (2006)	Text	Sentiment		Blogger sentiment predicts movie sales.
Archak et al. (2011)	Text	Product features	Part-of-Speech (POS)	The textual content of product reviews is an important determinant of consumers' choices, over and above the valence and volume of reviews
Netzer et al. (2012)	Text	Occurrence, co-occurrence, and lifts		User-generated content can be used to understand market structure.
Tirunillai and Tellis (2012)	Text	Valence	Support Vector Machine (SVM)	The valence of online chatter predicts abnormal stock returns.
Tirunillai and Tellis (2014)	Text	Topics	Latent Dirichlet Allocation (LDA)	LDA provides a few dimensions with good face validity and external validity, and are enough to capture quality.
Liu et al. (2016)	Text	Volume, sentiment and N-gram	Principal Component Analysis (PCA)	Online platform content, such as Twitter, can be effectively used for forecasting.
Mayew and Venkatachalam (2012)	Audio	Cognition level and emotional level	Software (LVA)	Positive and negative affects displayed by managers are informative about the firm's financial future.
Hobson et al. (2012)	Audio	Cognitive dissonance	LVA	Vocal markers of cognitive dissonance are useful for detecting financial misreporting.
Xiao and Ding (2014)	Image (Faces)	Facial features	PCA	Different faces have an effect on people's attitude toward the advertisement, attitude toward the brand, and purchase intention.
Liu et al. (2018)	Image	Color, shape, texture	SVM	The resulting metrics are consistent across consumer- and firm-generated images, as well as large survey-based metrics of consumer perceptions
Mollick (2014)	Video	None		Having a video improves the success rate of a crowdfunding campaign.
Zhang et al. (2018)	Video and Text	Moment-to-moment synchronicity (MTMS)		They propose a measure (MTMS) that can be used to identify viewer engagement.

2. Measuring visual information

To overcome the above-mentioned challenges and limitations, in this research, we introduce and review some possible measures of visual information. We also discuss some complex visual measures which may become available in the future. In addition, we propose a standard set of measures of visual information that can be obtained automatically and effectively from videos. In doing so, we use visual information processing, an emerging area in computer science that combines techniques from data mining, computer vision, and machine learning to process and understand visual information. We then show how to obtain these measures and demonstrate their effectiveness using the example of crowdfunding videos. These measures can benefit academics and practitioners seeking to analyze visual information in other contexts as well.

2.1. Possible measures of visual information

Several approaches could be used to measure visual information. A video can be viewed as a collection of many frames (the typical frame rate ranges from 24 frames per second to 60 frames per second), each frame being an individual static image. Each individual frame consists of many pixels (a typical resolution is 1024 × 768 pixels), which are usually represented in the "RGB" color space (i.e., red, green, and blue). Sometimes the pixels are also represented in the "HSV" (i.e., hue, saturation, and value) color space. The transformation from the RGB color space to the HSV color space can be done via simple linear transformation. Next, we introduce some simple measures based on the pixel values. Most of these measures are covered by Datta, Joshi, Li, and Wang (2006) and Zhang, Lee., Singh, and Srinivasan (2017).

2.1.1. Resolution

Resolution is perhaps the simplest measure that can be derived from an image. It captures the size of an image and is simply the number of pixels in the image. Higher resolution means greater image detail. Researchers may use other measures to study the size of an image as well (i.e., the sum of the image's width and height).

2.1.2. Aspect ratio

Aspect ratio is the width to height ratio of an image. Most videos have an aspect ratio of 4:3 or 16:9, which approximates the "golden ratio," and are standard for television screens or 70 mm files. Other relatively common aspect ratios include 3:2, 7:6, and 5:4.

2.1.3. Hue

Hue (i.e., RGB) is an important measure of image colors, and is believed to affect people's emotions. It is simply the first coordinate in the HSV color space. Warm hues such as red and yellow can trigger excitement, whereas cool hues such as blue and black make viewers more relaxed.

2.1.4. Brightness

Brightness is a simple and useful measure of visual information. When a pixel is represented in the RGB color space, brightness is the arithmetic mean of the red, green, and blue coordinates, i.e.,

$$\mu = \frac{R+B+G}{3}$$

The brightness of an image is the average of the brightness of all its pixels. An image with low brightness is considered a dark image, whereas an image of high brightness is considered as bright image. Other similar measures, such as luminance ($=0.2126 \times R + 0.7152 \times G + 0.0722 \times B$), are adjusted for psychological aspects.

2.1.5. Saturation

Saturation indicates chromatic purity. Some scholars argue that pure colors in an image tend to appeal more to viewers than dull or impure ones. When represented in the HSV color space, the saturation of an image can be easily obtained by averaging the saturation of each pixel of the image.

2.1.6. Contrast

Contrast is the difference in luminance or color that makes an object (or its representation in an image or display) distinguishable. Contrast is determined by the difference in the color and brightness of the object and other objects within the same field of view. There are several different formulas for calculating contrast (e.g., RSM contrast and Weber contrast).

2.1.7. The rule of thirds

"The rule of thirds" is a well-known principle in photography. It implies that centers of interest (i.e., a large part of a main object) should lie in the center of an image. One can calculate the above measures for the center object following the rule of thirds. For example, for an image of X × Y pixels, the average hue is calculated as follows:

$$I_H = \frac{9}{XY} \sum_{x=X/3}^{2X/3} \sum_{y=Y/3}^{2Y/3} I_H(x, y)$$

where $I_H(x,y)$ is the hue of pixel (x, y). The average brightness and saturation under the rule of thirds can be calculated in a similar fashion.

The above measures are simple and can be easily obtained using various software packages. More complex measures of visual information are outlined below.

2.1.8. Graininess

Graininess (or smoothness, texture, etc.) of an image is defined in different ways. In general, a grainy image is usually a photo taken with a grainy film or under high ISO settings. In contrast, a smooth image is often out-of-focus. Graininess also captures the presence of natural textures within the image. Researchers frequently use Daubechies wavelet transformation to calculate the graininess of an image (Datta et al., 2006).

2.1.9. Segmentation

An image often captures multiple objects, and can be decomposed into a number of distinct segments. Segmentation groups similar pixels into the same segment and different pixels into different segments; then, one can derive measures (e.g., hue and saturation) for each segment of the image. Segmentation can be done automatically using classification algorithms such as the K-means algorithm.

While the above measures can be adopted to measure static images, they may not be very powerful when analyzing dynamic videos. Since a video has many frames, the values of most static measures change constantly. For example, the color of the frames is seldom stable and is usually changing rapidly over the course of the video. Yet, simply averaging the colors of all frames provides little information to researchers. Accordingly, new measures must be developed to measure videos.

Therefore, in this study, we focus on two measures that are not static and can capture the dynamics of a video: (1) visual variation, which captures the change in visual information of a video; and (2) video content, which measure the features that the video contains. We also consider the duration (i.e., length) of a video. It is worth noting that, when needed, the static measures of images (e.g., the color composition and average brightness of a video) can be used together with the visual measures.

While our measures can be helpful, they are only the start of research in video mining. With the rapid development of machine learning and artificial intelligence techniques, new measures of visual information will become increasingly available. Hence, we also discuss some measures that may be available for analyzing videos in the future.

2.1.10. Facial expression

When an individual speaks in front of the camera, viewers can learn a great deal of information from his or her face (e.g., Is this person confident/nervous? Is this person trustworthy? Is there frequent eye contact between the person and the audience?) Although facial expression has not been included in the present study, measures of facial expression will be feasible with the advance of machine learning techniques.

2.1.11. Body language

Viewers often judge a person based on his or her body language. For example, crossed arms can indicate anxiety, vulnerability, or a closed mind. However, if crossed arms are accompanied by a genuine smile and overall relaxed posture, then the stance can indicate a confident, relaxed attitude. Like facial expression, body language will be a measurable factor for researchers in the future.

2.1.12. Event detection

Event detection allows researchers to detect what is happening in a video. For instance, the video may feature a musician playing guitar in front of the screen, a photographer operating a drone in the sky, or an autonomous driving vehicle traveling on a street. Event detection is based on image recognition techniques but goes further. It provides researchers with additional information regarding the objects displayed in the video.

2.2. Our measures

2.2.1. Visual variation

The first measure that we propose is called "visual variation," which is a normalized measure of the changes in visual information in a video. Visual variation differs from the existing static measures of images (e.g., color and color dispersion; see Zhang, Lee, Sing, and Srinivasan 2017) in that it captures dynamic changes. Intuitively, when the video screen changes rapidly (e.g., images move quickly from one place to another, or feature numerous different individuals, objects, and places), the video has a high visual variation level; otherwise, the video is likely to have a low visual variation level. We adopt this measure because visual variation has some relevance to the levels of visual stimulation studied by consumer psychologists, who posit that a medium level of visual stimulation is perceived as most satisfying (Berlyne, 1970; Hebb, 1955).

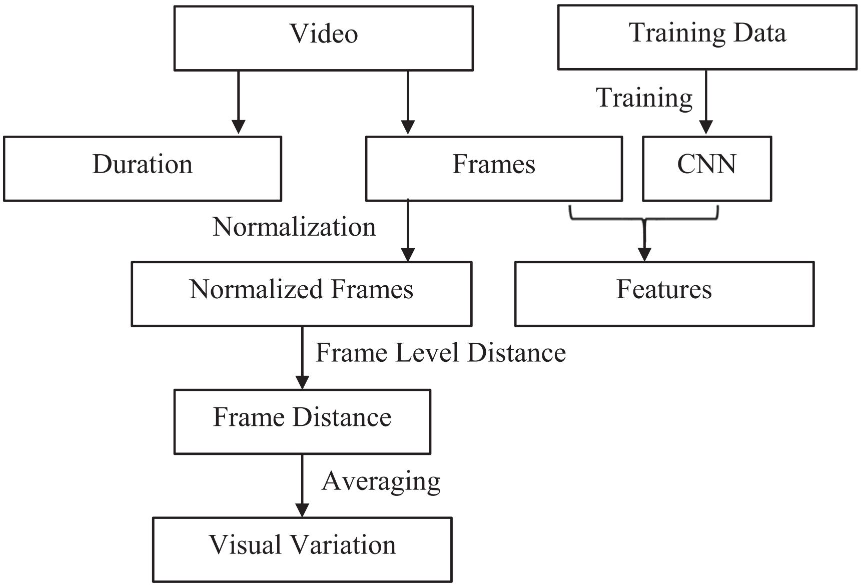


Fig. 1. Steps for calculating visual measures.

The construction of this measure has an intuitive appeal. We decompose a video into a number of static frames and then calculate the distance between consecutive video frames. Fig. 1 shows the step-by-step process of constructing the visual variation measure.

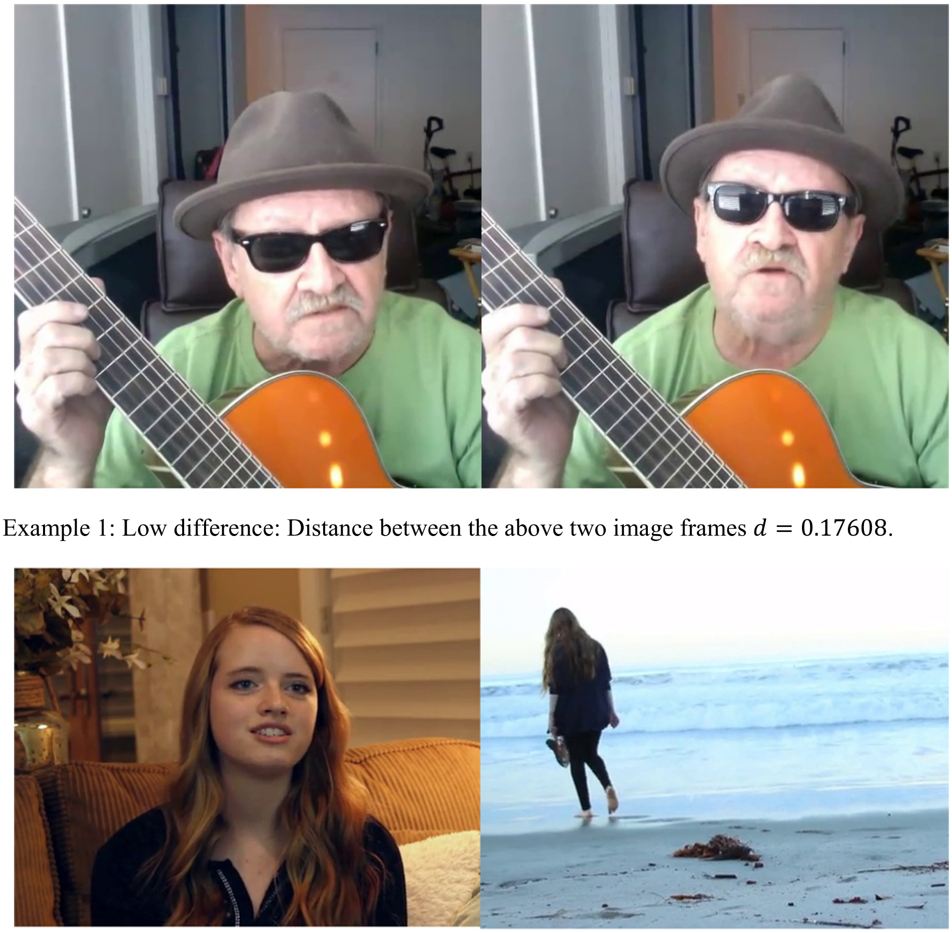
First, we read color images as 3D arrays and average the RGB channels to obtain an intensity measure—that is, we convert color images into grayscale. Although the conversion is not essential to our results, it simplifies the problem immensely and reduces the computational overhead (the measure is not qualitatively altered with other transformations). For example, for a color pixel (49, 15, 14), the grayscale value of the pixel will be converted to $(49 + 15 + 14) \div 3 = 26$.

Second, we normalize the image scores from the interval [0, 255] to [0, 1]. For an image with n pixels of value $x_1, \dots, x_n$ shown at time t, normalization is done through the following transformation:

$$x_i' = \frac{x_i - x_{min}}{x_{max} - x_{min}}$$

where $x_{max} = \max\{x_1, \dots, x_n\}$, $x_{min} = \min\{x_1, \dots, x_n\}$, and x_i' is the normalized grayscale value for the pixel x_i. The normalization compensates for possible exposure difference. For instance, if we take videos of the same object in the early morning and in the afternoon, the two video frames will be different in grayscale, but similar in content. Normalization reduces such exposure-induced differences between image frames.

Third, we calculate the pixel-level distance between the images, where the distance is defined as the Manhattan norm, $d(x_i', y_i') = |x_i' - y_i'|$. The difference between the images is the sum over the absolute distance of every pair of pixels, i.e., $d(X, Y) = \sum_{i=1}^{n} d(x_i', y_i')$. This aggregation assumes that the sizes of image frames are the same, which is always true for different frames of the same video.



Example 1: Low difference: Distance between the above two image frames $d = 0.17608$.

Example 2: High difference: Distance between the above two image frames $d = 1.16550$.

Fig. 2. Examples of image distance.

The algorithm-calculated distance score is consistent with our intuition, as illustrated with two examples in Fig. 2. In the first set of images, a musician is playing a guitar without any significant changes of background. In contrast, in the second set of images, the background changes dramatically from one setting to another. The distance of the first pair of images is equal to 0.176, which is much smaller than that of the second pair (equal to 1.04). In these two sets of images, the difference in distance scores clearly reflects the extent of the visual changes between the two images.

We then calculate the *visual variation* of a video in the following way. For a given video of duration *T*, we divide it into 10 equal-distanced clips and take the frame in the middle of each clip.[1] In other words, we extract 10 frames at times $\frac{T}{20}, \frac{3T}{20}, \ldots, \frac{19T}{20}$. We denote these 10 frames by $F_1, F_2, \ldots, F_{10}$. The visual variation score is the average distance between consecutive frames—that is,[2]

$$\text{visual variation level} = \frac{1}{9}\sum_{i=1}^{9} d\left(F_i, F_{i+1}\right).$$

The visual variation level is maximized when a video changes its scenes frequently and significantly, and is minimized when its images are relatively stable and predictable over time.

The visual variation measure bears some similarity to the stimulation level in consumer psychology. According to the stimulation theory, the emotional experience of pleasant arousal produced by visual stimuli should be central to the effect of a video. An excessive amount of visual stimulation can lead to disequilibrium that viewers strive to avoid. There is general agreement in the literature that the stimulation level obtained and the affective response to stimulation by a viewer follows an inverted U-shaped curve (Berlyne, 1970), with an intermediate level of stimulation perceived as the most satisfying. While stimulation level is much more complex than visual variation, visual variation can be used as a proxy for stimulation level in most contexts.

It is worth noting that visual variation is not the only proxy of visual information. Researchers can also construct similar measures based on variations in frame brightness and other qualities.

2.2.2. Video content

While visual variation captures the dynamics of a video, our second measure focuses on the *content* of the video. More specifically, we examine whether or not a certain feature (e.g., an individual) presents in a video frame. Compared to the traditional measures obtained from static images (e.g., color and brightness), video content is usually stable across video frames. Thus, it provides useful information about the entire video.

Although all video features can be recognized by humans, the task can quickly become difficult and even impossible when the amount of visual information scales up, making it impractical for commercial purposes. We overcome this limitation by using CNNs. A CNN is a machine learning algorithm that has wide applications in computer vision and artificial intelligence. CNNs are usually pre-trained for a specific task, such as image recognition, recommendation, or natural language processing. In our paper, a CNN takes an image frame as the input, performs a series of nonlinear transformations, and outputs a variable indicating whether or not a specific feature is present in the input image frame. The CNN technique is revolutionary in the sense that it dramatically reduces the amount of human labor required in visual analysis, while still maintaining a high-level performance; for example, Simonyan and Zisserman (2015) and Markoff (2014) show that CNNs can achieve an accuracy rate of 94%. While CNNs are relatively new to marketing, they are based on neural networks, which have already been used by marketing researchers to predict consumer choice (Bentz & Merunka, 2000; West, Brockett, & Golden, 1997).

At the time of this study, we did not have a vast dataset nor the necessary computational power to train our own CNNs. As a result, we outsourced the image recognition work to Imagga,[3] a start-up specializing in CNN-based image recognition at affordable prices. It is worth noting that there are other firms offering similar image recognition services (e.g., Clarifai).

2.3. Summary

Past research suggests that videos can be very effective tools of communication. This paper contributes to the marketing literature and practice on exactly *how* videos matter—specifically, on how different types of video features may affect overall video effectiveness. We propose a set of standard features that can be obtained from visual data. All of these features are measured automatically and can therefore be easily applied to large datasets at a marginal cost. In the next section, we demonstrate that these features have a significant impact on the effectiveness of videos. Accordingly, marketers and advertisers could use these measures to guide their daily marketing activities.

[1] One concern with this approach is that the distance between two consecutive frames can be different in different videos. Thus, in our empirical application, as a robustness check, we also constructed stimulation level with equal-distanced frames in all videos, and the results are qualitatively robust.

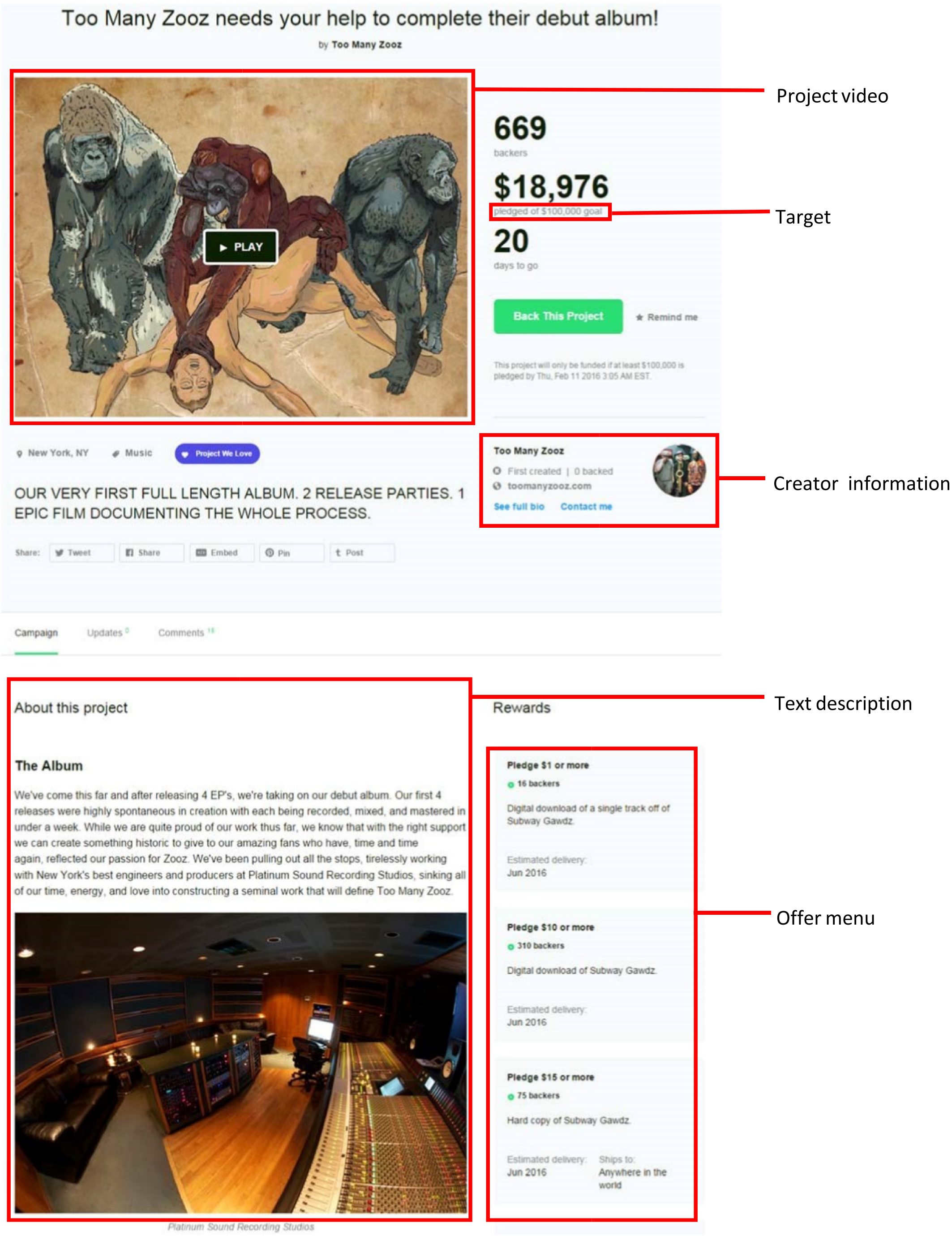
[2] The stimulation level of a video ad is usually constant over its duration. The results would remain the same if we chose only the first or last few frames to construct the stimulation measure.

[3] https://imagga.com

3. Data and Preliminary Analysis

To demonstrate the effectiveness of the proposed visual measures, we use crowdfunding data as our application context. The data is scraped from Kickstarter, the leading online crowdfunding website, where sellers or entrepreneurs (the “creators”) raise funds from potential buyers in order to initiate their new projects. In return, the creators offer products (e.g., a music album) or services (e.g., access to a concert) to be delivered on a future date.

Creators on Kickstarter can be industrial designers, musicians, software developers, writers, and so on. Here, unlike in conventional business models, a buyer commits to purchasing the product or service and prepays to fund the project. A project will be



Too Many Zooz needs your help to complete their debut album!

by Too Many Zooz

669
backers

$18,976
pledged of $100,000 goal

20
days to go

Back This Project ★ Remind me

This project will only be funded if at least $100,000 is pledged by Thu, Feb 11 2016 3:05 AM EST.

Too Many Zooz
First created | 0 backed
toomanyzooz.com
See full bio Contact me

New York, NY Music Project We Love

OUR VERY FIRST FULL LENGTH ALBUM. 2 RELEASE PARTIES. 1 EPIC FILM DOCUMENTING THE WHOLE PROCESS.

Share: Tweet Share Embed Pin Post

Campaign Updates 0 Comments 18

About this project

The Album

We've come this far and after releasing 4 EP's, we're taking on our debut album. Our first 4 releases were highly spontaneous in creation with each being recorded, mixed, and mastered in under a week. While we are quite proud of our work thus far, we know that with the right support we can create something historic to give to our amazing fans who have, time and time again, reflected our passion for Zooz. We've been pulling out all the stops, tirelessly working with New York's best engineers and producers at Platinum Sound Recording Studios, sinking all of our time, energy, and love into constructing a seminal work that will define Too Many Zooz.

Platinum Sound Recording Studios

Rewards

Pledge $1 or more
16 backers
Digital download of a single track off of Subway Gawdz.
Estimated delivery:
Jun 2016

Pledge $10 or more
310 backers
Digital download of Subway Gawdz.
Estimated delivery:
Jun 2016

Pledge $15 or more
75 backers
Hard copy of Subway Gawdz.
Estimated delivery: Jun 2016
Ships to: Anywhere in the world

Fig. 3. A sample music project posted on Kickstarter.

Table 2
Summary statistics.

Variable	N	Mean	St. Dev.	Min	Max
Video (dummy)	8327	0.819	0.385	0	1
Success (dummy)	8327	0.536	0.499	0	1
Target ($)	8327	17,404.150	346,258.300	1	21,474,836
Project duration	8327	36.149	14.864	1	92
Word count	8327	5.893	361.181	16	6546
Menu length	8327	8.978	5.636	0	73
Creator experience	8327	0.195	0.396	0	1
Positive	8327	0.028	0.012	0	0.111
Negative	8327	0.006	0.006	0	0.057
Price	8327	128.096	357.935	0	10,000
Funds pledged	8327	4864.492	12,359.450	0	600,874

successfully funded if and only if the total value of committed purchases exceeds a prespecified target within a prespecified time period (e.g., a target of $10,000 within a month). For a typical Kickstarter project, the creators design offerings, choose prices (e.g., a digital album for pledges of $15 or more, a CD and a bonus track for $25 or more, etc.), and set a funding target. When potential buyers arrive at the project site, they watch a video, followed by a paragraph of project description, before making their purchase decisions.

Our empirical research focuses on the projects in the music category of Kickstarter. Within this category, our data includes all music projects in the following three major U.S. markets: New York, Los Angeles, and Texas. It comprises all completed projects in these markets from the inception of Kickstarter to December 2015. In total, we amassed 8327 observations, among which 6822 projects had a video. With few exceptions, most of the projects aimed to offer a music album. We also replicate the analysis on the technology category to demonstrate the robustness of our results.

Crowdfunding sites provide an ideal setting to study the effect of visual information for a number of reasons. First, crowdfunding platforms offer a direct measure of video effectiveness, the funding outcome, which allows us to monetize the value of videos. This measure seems more relevant than traditional measures such as view-through rate, ad completion rate, and ad abandon rate. Second, most crowdfunding campaigns launch a video; in our dataset, 81.9% of projects had an accompanying video, providing us with a large number of videos to study the effect of video ads. (Videos are also commonly observed in other crowdfunding sites such as Indiegogo and Patreon.) Third, an online crowdfunding site is a fairly closed environment, where the potential buyers learn about the projects within the site. We are able to gather most key information presented to the buyers. (Note, however, that we are not able to obtain information concerning the social relationships between the creators and the potential buyers.) Relative to other purchase environments (e.g., supermarkets), there are fewer external and unobservable factors in crowdfunding that could affect buyer decisions. Finally, the buyer decision journey is simpler. Once the potential buyers arrive at a crowdfunding site, they browse the projects and then make their funding decisions. In comparison, the decision journeys in traditional purchase settings are much more complex, often making it difficult to isolate the advertising effect on sales from the effects of many other factors.

Fig. 3 shows a typical Kickstarter project page for "Too Many Zooz", a New York band. The band had spent a great deal of time playing music in subway stations, and asked for $100,000 to help complete their debut album. The project page contained a video, a product description, a menu of rewards and prices, and a target, as well as the number of backers and the amount of money pledged so far.

Like Too Many Zooz, most crowdfunding projects use a video to communicate with potential buyers and supporters. It is generally agreed that posting a video will increase the success rate. In an industry study, Strickler (2009) found that projects with videos have had a success rate of 54% while ones without have had a success rate of 39%. However, not all the crowdfunding videos were designed in the same way, nor did they make the same impact on potential buyers. In the rest of this section, we extract the visual measures we introduced as well as the control variables, and show how they affect the effectiveness of the videos.

3.1. Dependent variable

We choose the final outcome of the project as our dependent variable. We measure project outcome by a dummy variable (*success* in Table 2), indicating whether or not the project succeeded by raising enough funds to meet the target. For the projects listed on Kickstarter, their outcomes are "all or nothing": if the amount of funds pledged for a project is below its target, regardless of closeness to the target, the creators receive nothing. Therefore, the creators' primary goal is to reach their targets. As shown in Table 2, among all 8327 projects, 4466 of them successfully reached their goals (all others failed), leading to an overall success rate of 53.6%.

We choose to measure the project outcome by the indicator variable for success because, first, this is how project success is typically judged in practice. This success measure is also commonly adopted in the literature (e.g., Mollick, 2014). Second, alternative outcome measures, such as the total amount of funds pledged for each project, follow irregular distributions. Fig. 4 plots the amount of funds pledged as a percentage of target for all of the projects. The distribution of results is skewed in both the below- and above-target areas, with a large mass in the interval slightly above one (at the target). The explanation for this distribution pattern is discussed by Kuppuswamy and Bayus (2013) in their paper on herding and bystander effects among buyers.[4]

[4] All of the main results hold in the regression analyses using the logarithm of the total amount pledged as the dependent variable.

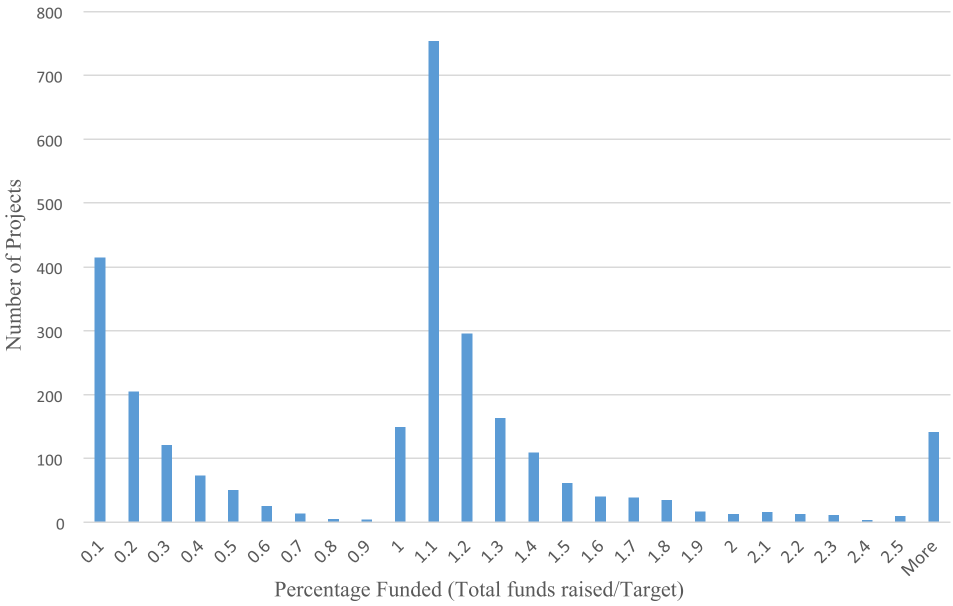


Fig. 4. Funding outcome.

Next, we describe the visual measures extracted from the crowdfunding videos.

3.2. Visual Variation

In the data, the *visual variation* level ranges from 0 to 1.51, with an average of 0.435. To visualize the relation between *visual variation* and *project success*, we plot the visual variation and product success rate in Fig. 5.

The figure shows a concave shape curve: the marginal effect of visual stimulation is declining as visual stimulation level increases.

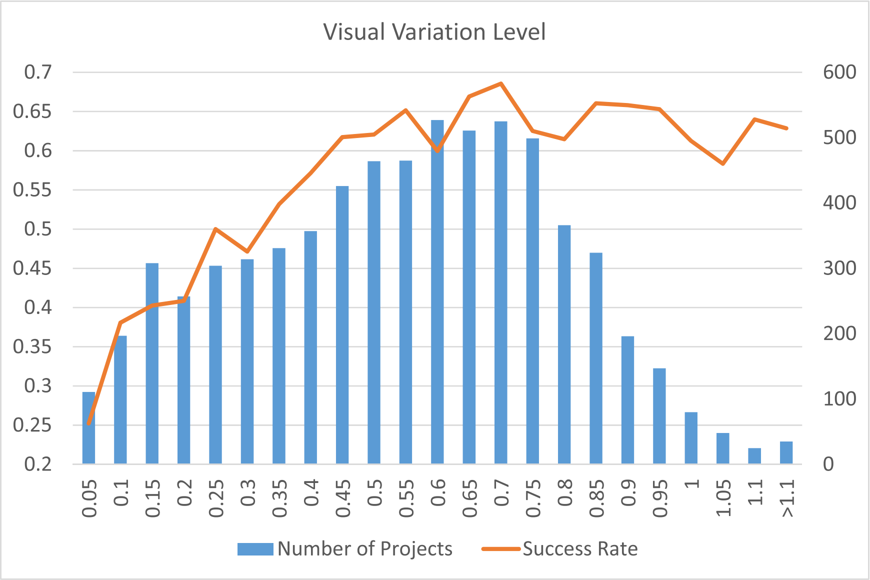


Fig. 5. Visual variation, number of projects, and success rate.

3.3. Video Content

We first extract three frames from the beginning, middle, and end of each video (i.e., video frames at times $\frac{T}{6}, \frac{T}{2}, \frac{5T}{6}$).[5] Then, using the CNN algorithm, we extract two features from project videos: *human* and *instruments*. *Human* is a dummy variable that is set to one when the video features human beings, and *instruments* is a dummy variable indicating whether the video contains any of the following instruments: guitar, wind instruments, piano, bass, or banjo. We focus on these five instruments because they are the most common among Kickstarter music projects. Other musical instruments, such as drum and violin, appear much less frequently (accounting for <1% of all videos). Hence, we excluded these instruments from our analysis. As we later discuss in greater details, these two variables are important for the music category and could help communicate credibility to buyers.[6]

Based on our video image analysis, in our data, a video has (on average) a 78.5% chance of including human beings, and a 27.3% chance of featuring one of the five common musical instruments. We conduct survey to validate our findings using Amazon Mechanical Turk. We randomly select 12 photos in our sample; some photos have humans and various instruments depicted in them whereas other photos have neither humans nor instruments in them. We then ask each respondent a series of questions for each picture in a randomized order. The questions asked whether there is at least one person in the photo and whether there are the following instruments: guitar, wind instruments, piano, bass, or banjo. One hundred and two respondents completed the survey. The aggregated recall and precision for these photos are 94% and 73%.

3.4. Video duration

In addition to visual variation and video content, we also study *video duration*, the number of seconds of the video. A lengthy video carries both advantages and disadvantages: while a longer video offers consumers more information about the product or service, after reaching a certain threshold, the incremental learning from additional frames typically diminishes, and consumers feel increasingly bored with watching the video. We use the log transformation of video duration in our empirical analysis.

3.5. Controlling for audio content

Videos provide not only visual content, but audio content as well. If the audio content is not orthogonal to the visual content, our estimates of visual measures could be biased. In this study, we follow the standard computer science and acoustic approach to control for audio content. In the analysis, we include the following variables that control for audio content: zero crossing rate (ZCR), energy entropy, brightness, and spectral entropy. (See Appendix B for a detailed description of these variables.)

3.6. Sentiment analysis of project description

A project description is always included in the project page. A typical buyer views both the video and product description before making the purchase decision. The communications through the text and video can substitute for each other or be complementary in attempting to persuade the buyer. Thus, it is necessary to account for the text information when studying the effect of videos. In the present study, we limit the text analysis to *word count* and *simple sentiment analysis* (commonly conducted in the literature). First, we use the variable log(*words*) to denote the logarithm of the number of words in the project description as a proxy for the amount of text information.

Second, we adopt sentiment analysis to study the sentiment expressed in the project description. Prior literature suggests that the frequency of positive and negative words captures the tone of a text. Following this tradition, we use positive and negative word lists complied by Hu and Liu (2004) and Liu, Hu, and Cheng (2005). We construct two measures, *positive* and *negative*, the fractions of positive and negative words in a descriptive text.

3.7. Other control variables

3.7.1. Projects and offers

First, each project has a *target amount of funds* that the creator is requesting for the project. If the pledged funds for a project fail to meet this target, then the project is deemed unsuccessful and all of the pledged funds are returned to the buyers. The targets vary significantly among projects, ranging from $1 to $21,474,836, as shown in Table 2. Targets are also skewed in distribution, with an average of $17,404, and a median target of $5000. In the regression analysis, we take log transformation of the target.

Second, each project specifies a *project duration*, the number of days within which the project remains valid and can continue accepting fund pledges. For example, if a project ran from March 30, 2015, to May 2, 2015, *then project duration is* equal to 33 days. A project closes automatically after reaching the end of its project duration. In our data, Kickstarter allowed a maximum

[5] Note that our selection of video frames for the CNN analysis is not without loss of generality. Researchers may choose different numbers of frames at different time points based on their research contexts.

[6] More features can be obtained using CNNs (e.g., table, vehicle, and flower). We did not include these features in our analysis because they appear less frequently and are less important for a music project, though they may be useful in other contexts. Moreover, we did not include other standard image features that can be obtained without CNNs (e.g., color and brightness).

project duration of 90 days. Most of the projects (i.e., about 80% in our data sample) had durations of between 30 and 60 days, with an average duration of 36 days.

Third, most projects offer multiple levels of rewards and/or recognitions. Hu, Li, and Shi (2015) suggest that a well-designed menu of offerings can help creators improve their project success rates. We include the variable *menu length* in Table 2 to measure the number of different rewards in the form of products or services offered to buyers. In our data, two projects offered no rewards, 478 projects offered one single reward, and the remaining 7847 projects offered at least two levels of rewards. On average, each project offers about eight levels of rewards.

Fourth, the *price* variable indicates the median price of the offerings in the menu. As discussed above, most projects have multiple offerings, each associated with a particular price. For instance, consider a project that offers four versions of products at prices $15, $30, $50, and $100; then the median price is ($30 + $50)/2 = $40. The average price in our dataset is $128. The two projects that do not offer any rewards have prices equal to zero.

Finally, we construct a number of indicator variables for the music *genres*, following the classifications used by Kickstarter: *Blues*, *Chiptune*, *Classical Music*, *Country & Folk*, *Electronic Music*, *Faith*, *Hip-Hop*, *Indie Rock*, *Jazz*, *Kids*, *Latin*, *Metal*, *Music* (for those without a specific genre), *Pop*, *Punk*, *R&B*, *Rock*, and *World Music*.

3.7.2. The creators

We measure the project creators' level of experience with Kickstarter by using a dummy variable, *experience*, to indicate whether or not the creators of a project have previous experience posting projects on Kickstarter. In our data, 19.5% of creators

Table 3
Logistic regression results.

	Success			
	(1)	(2)	(3)	(4)
Video		0.646***		
		(0.074)		
Log(duration)				1.803***
				(0.466)
Log(duration) – squared				−0.198***
				(0.047)
Visual variation				2.722***
				(0.455)
Visual variation - squared				−1.633***
				(0.425)
ZCR			−1.280	1.505
			(2.520)	(2.566)
Energy			−9.822***	−10.133***
			(1.546)	(1.574)
Entropy			0.296	0.088
			(0.257)	(0.284)
Brightness			−7.719***	−8.850***
			(2.103)	(2.138)
Spectral entropy			1.101***	0.907***
			(0.258)	(0.263)
Log(target)	−0.633***	−0.661***	−0.655***	−0.665***
	(0.029)	(0.029)	(0.034)	(0.034)
Project duration	−0.016***	−0.015***	−0.017***	−0.016***
	(0.002)	(0.002)	(0.002)	(0.002)
Menu length	0.137***	0.128***	0.123***	0.115***
	(0.007)	(0.007)	(0.007)	(0.007)
Creator experience	−0.296***	−0.291***	−0.338***	−0.329***
	(0.065)	(0.066)	(0.073)	(0.074)
Price	−0.0001	−0.0001	−0.0002	−0.0002
	(0.0001)	(0.0001)	(0.0002)	(0.0002)
Log(words)	4.662***	4.328***	5.549***	5.275***
	(0.470)	(0.468)	(0.572)	(0.576)
Log(words) – squared	−0.329***	−0.305***	−0.398***	−0.376***
	(0.040)	(0.040)	(0.048)	(0.048)
Positive	−12.281***	−11.924***	−14.379***	−14.636***
	(2.204)	(2.217)	(2.545)	(2.571)
Negative	−53.940***	−51.692***	−57.320***	−55.829***
	(4.683)	(4.713)	(5.437)	(5.505)
Observations	8327	8327	6822	6822
Log likelihood	−4462.977	−4424.722	−3598.829	−3541.517
Akaike Inf. Crit.	8987.954	8913.445	7269.657	7163.034

Note: Columns (1) and (2) use all projects data. Columns (3) and (4) use data on projects with a video. Regression includes genre and gender fixed effects. Standard error in parentheses.

*** $p < 0.01$.

Table 4
The effects of featuring humans and/or instruments in videos on project success.

	Success		
	(1)	(2)	(3)
Human	0.372^{**}		0.348^{**}
	(0.161)		(0.161)
Instrument		0.205^{***}	0.199^{***}
		(0.064)	(0.064)
Log(duration)	1.734^{***}	1.805^{***}	1.741^{***}
	(0.466)	(0.466)	(0.466)
Log(duration) – squared	-0.192^{***}	-0.199^{***}	-0.193^{***}
	(0.047)	(0.047)	(0.047)
Visual variation	2.620^{***}	2.598^{***}	2.508^{***}
	(0.458)	(0.457)	(0.459)
Visual variation – squared	-1.537^{***}	-1.547^{***}	-1.461^{***}
	(0.427)	(0.426)	(0.428)
Observations	6822	6822	6822
Log likelihood	−3538.867	−3536.338	−3534.008
Akaike Inf. Crit.	7159.733	7154.675	7152.015

Note: This table uses data on projects with a video. Regression includes audio controls, target, project duration, menu length, creator experience, price, word count, sentiments, genre, and gender. Standard error in parentheses.
***: $p<0.01$ **: $p<0.05$

had prior experience with Kickstarter. Past research on offline ventures shows that experienced entrepreneurs are more likely to succeed (Roure & Maidique, 1986). Therefore, buyers may perceive a higher success rate with more experienced creators. However, buyers on Kickstarter may also have a tendency to support newer, less experienced musicians. As a result, those musicians with no prior experience using Kickstarter may attract more funds. Overall, the direction of relationship between *experience* and *project success* is unclear.

We also control for the gender of the creators by recognizing the gender from the first name of the creator. For example, the first name “Alice” corresponds to a female whereas “Bob” corresponds to a male. Names in non-English or non-names are not recognized.

4. Results

In this section, we present the results of our main empirical analysis. Since the dependent variable (*project success*) is a binary variable, we use logistic regression to estimate the effect of visual information on project success rate. In the regression analysis, we control for genre fixed effects by including the *genre* dummies. We provide initial evidence for the validity of the video measures. Although we cannot claim a causal relationship, this helps demonstrate the effectiveness of the video measures.

We first present the regression results regarding the effect of *video duration* and *visual variation*. The results are presented in Table 3.

Column (1) of Table 3 represents the regression result, excluding any video-related variables. Column (2) is the regression result including *video*, the variable indicating whether or not a project had a video. Both regressions were run on the entire dataset of 8327 projects. Column (3) presents the regression results controlling for audio information but not for video content. The last column shows the regression results including the *video duration*, *visual variation*, and *visual variation-squared* variables. Both columns (3) and (4) are run on the 6822 projects with videos.

4.1. Effect of visual information on project success: visual variation

Column (4) of Table 3 is the regression of success on video characteristics, including only projects with a video. It shows that the impact of *visual variation* on the likelihood of project success has a significant linear effect (mean 2.722, p-value <0.01) and a significant nonlinear effect (mean −1.633, p-value <0.01).

This finding is consistent with the optimal stimulation level (OSL) theory in consumer behavior, according to which, the emotional experience of pleasant arousal produced by visual stimuli is central to the effectiveness of videos. The above result holds after we control for other elements known to influence the success rate of a project. For practical implications regarding video design, one would need to resort to our definition of *visual variation* and infer the corresponding specifications for the video images.[7]

4.2. Effect of visual information on project success: video content

Using the CNN techniques, we create two variables from the video content—*human* and *instruments*—for 6822 projects with videos. We run logistic regressions on *project success* with respect to each of these two variables, separately and together. We

[7] It is worth noting that our data only shows that the effect of visual variation is decreasing, and does not show an inverted U-shape curve. One possible reason is that the visual variation level is generally below the optimal level, and we only observe the left half of the inverted U-shape curve.

summarize the estimation results in Table 4. The coefficients of *human* and *instruments* are both significant and positive (p-value <0.01 in all cases), suggesting that featuring humans and/or instruments is positively related to project success rates.

Table 4 provides evidence that the content of a project's video affects the funding decisions made by the potential buyers. We manually inspected a large number of videos, and found that almost all humans featured in the videos are the project creators. When they are featured in the videos, these creators seek to develop a personal connection with potential buyers, making buyers feel like they have a real and credible relationship with the musicians. Our result is also consistent with the suggestions provided on Kickstarter's official blog: "And don't be afraid to put your face in front of the camera and let people see who they're giving money to. We've watched thousands of these things, and you'd be surprised what a difference this makes."[8]

Musical instruments are the essential elements of music making and performance. Therefore, showing the instruments in the video ads should enhance the perceived credibility of the musical projects. This feature is key to the music category. Due to technical constraints, we were not able to confirm through the programming that the creators actually played the instruments in each video (although manual inspection of several hundred videos indicates that they were indeed playing the instruments). Our results confirm that the projects featuring instruments in a video are associated with a higher success rate.

Overall, the creators/musicians who feature instruments in their videos are more likely to succeed in crowdfunding. In addition, the coefficient of *human* is greater than that of *instruments*. This finding indicates that when establishing credibility to improve the project success rate, it is more effective if one can achieve that credibility through personal connection with the project creators.

5. Discussion

With the rapid development of information technologies, visual information is becoming increasingly prevalent among firms. Today, practitioners and researchers are relying more than ever on the marketing power of online videos. Traditional research analyzes videos using controlled lab experiments, which are costly, time consuming, and limited to small scales. Further, there has been no standard set of measures of visual information. These challenges substantially restrict the study of videos.

In this paper, we present several important measures that can be automatically obtained from videos. We first review some standard and simple measures of images such as hue, brightness, and saturation, and then propose two new measures —*visual variation* and *video content*. Visual variation is calculated based on the difference between consecutive video frames and reflects the dynamic changes in the video frames. Visual variation bears some similarities to visual stimulation level, and can be obtained completely automatically and applied to the aggregate population. The video content measure captures whether a specific object appears in the video (e.g., whether the video features humans and/or musical instruments). The measure is obtained using CNN, a reliable and scalable machine learning algorithm. We also include the length of the video as another measure of visual information. We then take Kickstarter, an online crowdfunding website, as our empirical context. We take the final outcome of the crowdfunding campaigns as the dependent variable and the video measures of campaign videos as the independent variables. Analysis shows that the video measures have a significant effect on the final outcome of projects, thereby demonstrating the explanatory power of video measures.

5.1. Potential applications

We outline several video measures that can potentially benefit marketers and researchers in many different contexts. Below, we describe several scenarios where video mining can be helpful.

1. **Television and online video advertising**. In 2018, firms worldwide spent $178 billion on television advertising and more than $20 billion on online video advertising. Given this huge market size, it is critical to examine how to design an effective advertising video. As noted, in the past, researchers have used lab experiments to analyze the effectiveness of television advertising, but this is costly, time consuming, and limited to a small scale. With video mining techniques, practitioners and researchers can now use automatic methods to extract meaningful features from video advertisements and investigate the effects of these measures.
2. **Product videos**. Just like creators of crowdfunding projects, sellers and manufacturers of innovative and technical products often post videos online describing the possible usage of their products. For example, DJI, the world's largest drone manufacturer, has uploaded several product videos on YouTube and Facebook, as well as its official website. Consumers who are new to drones can assess the value of a drone by watching these product videos. By applying video mining, sellers could optimize their product videos and improve consumers' attitudes toward their products. DJI, for instance, may want to find out whether a video focus exclusively on the drone product or feature a person operating the drone as well.
3. **User-generated videos in online reviews**. Today, most online retailers (e.g., Amazon and Taobao) allow buyers to upload user-generated videos with their reviews, which help other consumers better understand the products (e.g., how a product works or how a buyer looks in a new piece of clothing). While these online retailers do not offer videos themselves, they may still take advantage of video mining to examine the helpfulness or usefulness of these reviews and decide how to rank the reviews and make specific recommendations. Moreover, they can provide users with a guidance on how to prepare an attractive video for the review.
4. **Motion pictures**. Video mining could be extremely useful for the motion picture industry. A movie is essentially a video, and researchers have spent decades analyzing how to make a great movie. Now, equipped with video mining techniques, producers and researchers may extract features from movies and analyze how these features affect box office takings and consumer

[8] Source: https://www.kickstarter.com/blog/how-to-make-an-awesome-video.

preferences. For instance, what features do consumers like most? What is the optimal visual variation level for a specific genre of movie? Answering these questions will help companies create movies that attract more viewers.

5. **Video games**. The video game industry continues to grow at a remarkable pace, with a worldwide revenue of $138 billion in 2018, according to Newzoo's Global Games Market Report. In 2017, >7000 video games were released on Steam alone. Yet, the precise factors that lead to a successful video game are relatively unknown. Some articles even attribute the success of certain video games to random chance. With video mining techniques, however, game makers could be able to discover the features that players value most.

In addition to the scenarios mentioned above, video mining can also be used for other purposes, such as online education and live streaming.

5.2. Limitations and future research

While we believe that video mining will become ever more important and prevalent in the future, due to technical and data constraints, our current study is subject to a few caveats.

First, as discussed, we did not have a comprehensive training set and outsourced the image recognition task to an application programming interface (API). As of the time of this research, image recognition techniques are still in their infancy, though they continue to improve rapidly. We expect that in the future, with the development of new techniques and larger training sets, new algorithms will be introduced with much higher levels of precision and faster performance. In addition, future studies may collect their own training sets, which will greatly improve the performance of the image recognition algorithms.

Second, due to technology limitations, our analysis ignores many features of videos, some of which can be very important. For example, the effectiveness of a video may depend on the emotional features (e.g., whether the speaker is confident enough), which are not captured in the present research. With the development of facial recognition techniques, it may soon be possible for researchers and practitioners to extract emotional features from videos as well. This will greatly enhance our understanding of videos.

Third, in our empirical example, we consider the effect of video measures on the total funds pledged by all individual contributors; however, viewers are heterogeneous, and the effects of a video may be different across different user groups (e.g., males and females may react differently to the same object in a video). When the reactions of all audiences are available, it would be helpful to incorporate individual heterogeneity into the analysis of videos. This development could also be important for practitioners: with the ability to personalize recommendations, firms could deliver different video content to different users.

Fourth, although we can show that video features are related to the success of crowdfunding projects, we cannot claim casual relationships. For instance, funders may support a music project because of the creators' outstanding performance in front of the camera, which is not reflected in our video measures. The same problem can also affect other video mining tasks, such as analyzing the effectiveness of online video advertising. If researchers need to establish a causal relationship between video measures and other variables of interest, it may still be necessary to include lab experiments in the study.

Finally, our empirical application focuses solely on the crowdfunding industry. The reasons for this choice include the fact that we can easily quantify the effectiveness of the videos (from the total contributions pledged by funders), and the platform provides us with thousands of videos to analyze. However, our findings may not be applied to other industries, and these measures may not have the same predictive power when explaining other dependent variables.

We do not view our research as the final word on video mining. Instead, we consider our work a starting point for further research on this important topic. Our research can be extended in several general directions: (1) applying new algorithms to extract more features such as facial expressions from videos; (2) generalizing the results to other categories such as the online gaming industry; and (3) using advanced econometric models to better understand the effect of the videos. We believe these extensions will significantly benefit practitioners and expand our understanding of videos' effects on viewers.

Acknowledgements

Financial support from Social Science and Humanities Research Council of Canada is gratefully acknowledged.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ijresmar.2019.02.004.

References

Archak, N., Ghose, A., & Ipeirotis, P. G. (2011). Deriving the pricing power of product features by mining consumer reviews. *Management Science*, *57*(8), 1485–1509.

Balducci, B., & Marinova, D. (2018). Unstructured data in marketing. *Journal of Academy of Marketing Science*, *46*(4), 1–34.

Bentz, Y., & Merunka, D. (2000). Neural networks and the multinomial logit for brand choice modelling: A hybrid approach. *Journal of Forecasting*, *19*(3), 177–200.

Berlyne, D. E. (1970). Novelty, complexity, and hedonic value. *Perception & Psychophysics*, *8*(5), 279–286.

Business Insider Intelligence (2017). Video will account for an overwhelming majority of internet traffic by 2021. http://www.businessinsider.com/heres-how-much-ip-traffic-will-be-video-by-2021-2017-6.

Chevalier, J. A., & Mayzlin, D. (2006). The effect of word of mouth on sales: Online book reviews. *Journal of Marketing Research*, *43*(3), 345–354.

CIKLUM (2017) Big Data and the Challenge of Unstructured Data. CIKLUM, August 29. https://www.ciklum.com/blog/big-data-and-the-challenge-of-unstructured-data/

Datta, R., Joshi, D., Li, J., & Wang, J. Z. (2006). Studying aesthetics in photographic images using a computational approach. *European Conference on Computer Vision*, 288–301.

Dorian S (2017) Data: The world's most underused valuable resource. Medium, December 1. https://medium.com/@dselz/data-the-worlds-most-underused-valuable-resource-bb4177f79933

Dörre, J., Gerstl, P., & Seiffert, R. (1999). Text mining: Finding nuggets in mountains of textual data. *In Proceedings of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 398–401.

Dreier T (2017) 88% will increase online video spending in 2018, Finds Wochit. http://www.onlinevideo.net/2017/12/increase-online-video-spending-2018/.

Feldman, R., & Sanger, J. (2007). *The text mining handbook: Advanced approaches in analyzing unstructured data.* Cambridge: Cambridge University Press.

Flahive, E. (2017). 36 mind blowing YouTube facts, figures and statistics – 2017. *December*, *13*, 2017 http://videonitch.com/2017/12/13/36-mind-blowing-youtube-facts-figures-statistics-2017-re-post/.

Ghose, A., Ipeirotis, P. G., & Li, B. (2012). Designing ranking systems for hotels on travel search engines by mining user-generated and crowdsourced content. *Marketing Science*, *31*(3), 493–520.

Godes, D., & Mayzlin, D. (2004). Using online conversations to study word-of-mouth communication. *Marketing Science*, *23*(4), 545–560.

Gorn, G. J., Chattopadhyay, A., Yi, T., & Dahl, D. W. (1997). Effects of color as an executional cue in advertising: They're in the shade. *Management Science*, *43*(10), 1387–1400.

Hebb, D. O. (1955). Drives and the CNS (conceptual nervous system). *Psychological Review.*, *62*(4), 243.

Hobson, J. L., Mayew, W. J., & Venkatachalam, M. (2012). Analyzing speech to detect financial misreporting. *Journal of Accounting Research*, *50*(2), 349–392.

Howatson, A. (2016). How to unlock the power of unstructured data. *Marketing Tech News, December*, *13*, 2016 https://www.marketingtechnews.net/news/2016/dec/13/how-unlock-power-unstructured-data/.

Hu, M., Li, X., & Shi, M. (2015). Product and pricing decisions in crowdfunding. *Marketing Science*, *34*(3), 331–345.

Hu, M., & Liu, B. (2004). Mining and summarizing customer reviews. *In Proceedings of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 168–177.

Jude M (2016) Unstructured data analysis is critical, but difficult. TechTarget, November 2016. https://searchnetworking.techtarget.com/tip/Unstructured-data-analysis-is-critical-but-difficult.

Klostermann, J., Plumeyer, A., Böger, D., & Decker, R. (2018). Extracting brand information from social networks: Integrating image, text, and social tagging data. *Interntional Journal of Research in Marketing*, *35*, 538–556.

Kuppuswamy, V., & Bayus, B. L. (2013). Crowdfunding creative ideas: The dynamics of project backers in Kickstarter. Working paper. *UNC Kenan-Flagler School*.

Lee, T. Y., & Bradlow, E. T. (2011). Automated marketing research using online customer reviews. *Journal of Marketing Research*, *48*(5), 881–894.

Liu, B., Hu, M., & Cheng, J. (2005). Opinion observer: Analyzing and comparing opinions on the web. *In Proceedings of the 14th International Conference on World Wide Web*, 342–351.

Liu, X., Singh, P. V., & Srinivasan, K. (2016). A structured analysis of unstructured big data by leveraging cloud computing. *Marketing Science*, *35*(3), 363–388.

Liu, L., Dzyabura, D., & Mizik, N. (2018). Visual listening in: Extracting brand image portrayed on social media. *In Workshops at the Thirty-Second AAAI Conference on Artificial Intelligence*.

Markoff, J. (2014). Computer eyesight gets a lot more accurate. *New York Times, August*, *29*, 2014 http://bits.blogs.nytimes.com/2014/08/18/computer-eyesight-gets-a-lot-more-accurate.

Mayew, W. J., & Venkatachalam, M. (2012). The power of voice: Managerial affective states and future firm performance. *Journal of Finance*, *67*(1), 1–43.

Mishne, G., & Glance, N. S. (2006). Predicting movie sales from blogger sentiment. *AAAI spring symposium: computational approaches to analyzing weblogs* (pp. 155–158).

Mollick, E. (2014). The dynamics of crowdfunding: An exploratory study. *Journal of Business Venturing*, *29*(1), 1–16.

Nair, R., & Narayanan, A. (2012). *Benefitting from big data: Leveraging unstructured data capabilities for competitive advantage.* (Booz & Company).

Netzer, O., Feldman, R., Goldenberg, J., & Fresko, M. (2012). Mine your own business: Market-structure surveillance through text mining. *Marketing Science*, *31*(3), 521–543.

Rethans, A. J., Swasy, J. L., & Marks, L. J. (1986). Effects of television commercial repetition, receiver knowledge, and commercial length: A test of the two-factor model. *Journal of Marketing Research*, *21*(3), 50–61.

Rizkallah, J. (2017). The big (unstructured) data problem. *Forbes, June*, *5*, 2014 https://www.forbes.com/sites/forbestechcouncil/2017/06/05/the-big-unstructured-data-problem.

Roure, J. B., & Maidique, M. A. (1986). Linking prefunding factors and high-technology venture success: An exploratory study. *Journal of Business Venturing*, *1*(3), 295–306.

Simonyan K, Zisserman A (2015) Very deep convolutional networks for large-scale image recognition. In Proceedings of the 2015 International Conference on Learning Representations (ICLR).

Strickler, Y. (2009). The Importance of Video. *The Kickstarter Blog*https://www.kickstarter.com/blog/the-importance-of-video.

Sudhir, K. (2016). Editorial: The exploration-exploitation tradeoff and efficiency in knowledge production. *Marketing Science*, *35*(1), 1–9.

Tirunillai, S., & Tellis, G. J. (2012). Does chatter really matter? Dynamics of user-generated content and stock performance. *Marketing Science*, *31*(2), 198–215.

Tirunillai, S., & Tellis, G. J. (2014). Mining marketing meaning from online chatter: Strategic brand analysis of big data using latent Dirichlet allocation. *Journal of Marketing Research*, *51*(4), 463–479.

Wedel, M., & Pieters, R. (2008). Eye tracking for visual marketing. *Foundations and Trends in Marketing*, *1*(4), 231–320.

West, P. M., Brockett, P. L., & Golden, L. L. (1997). A comparative analysis of neural networks and statistical methods for predicting consumer choice. *Marketing Science*, *16*(4), 370–391.

Xiao, L., & Ding, M. (2014). Just the faces: Exploring the effects of facial features in print advertising. *Marketing Science*, *33*(3), 338–352.

Zhang Q, Wang W, Chen Y. (2018) Extracting and utilizing in-consumption moment-to-moment dynamics: The case of movie appreciation and live comments. Working paper, Hong Kong University of Science and Technology.

Zhang, S., Lee, D., Singh, P. V., & Srinivasan, K. (2017). How much is an image worth? *Airbnb property demand estimation leveraging large scale image analytics. Working paper, Tepper School of Business*. CMU.