Marketing Deep Dive: Enriching the Digital Involvement Cycle, Part I of III

Rick W. Lentz, M.S., M.S.

University of Wisconsin, La Crosse: DS780

27 Sept, 2017

Abstract

This three-paper series that focuses on published methods for exploiting and generating engaging user content from social media data to improve marketing decisions. The first paper deep dives on methods that can effect transformative change in the content generation and delivery of an individual marketing experience.

What is the shift towards individual content generation and delivery? The real difference is that individual dynamic content generation uses technology to generate an interactive experience that uses predicted preferences, to increase excitement and engagement. It is also different in that it requires a continual tracking effort to detect shifts in user preferences, rather than point-in-time estimate. The systems and methods that track, reward, connect, and measure the user experience are the same.

Data: A Foundation for the dynamic content generation experience

Users generate troves of unstructured data. Parsing this data allows machines to 'learn' about likes and dislikes, beyond those binary icon attributes that systems let users select. Within unstructured data, there is a story to interpret. People can do this analysis quickly, but not at scale, for not for all customers.

This paper deep dives on techniques for understanding customer's unstructured data and extracting preferences. The article concludes with an example showing how the digital marketing experience is improved by using preferences to drive a dynamically generated content segment, increasing its appeal and engagement.

Types of Data: the Video Stream

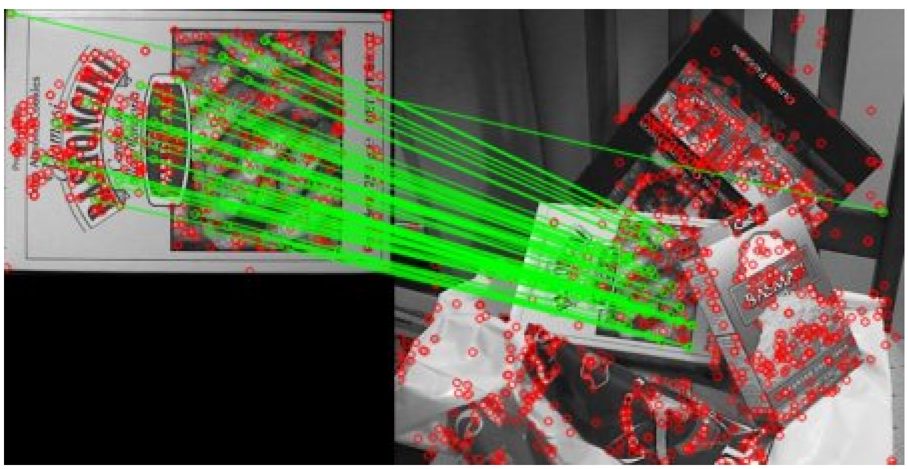
Videos are produced and shared by thousands of customers on their timelines every day. Individual customers typically don't retain certain rights, specifically to derivative information when sharing information on most social media platforms. By 'following you', your customers reveal a trove of unstructured content and that contains specific information that can be parsed to improve customer tracking, increase preference understanding, and aid in dynamic content generation.

Recovering Details

The content we care about in a video stream is who, what, where, and when. Deep neural networks can be trained to make predictions over a variety of tasks directly from the image frames. This paper chooses a different path, with focus on recovering the lost 3D structure for the benefit of derivative analysis. We begin the deep dive by looking at how this is done.

Image Features: SIFT, SURF, DAISY...

Each video contains a series of 2D frames. Features can be extracted from each 2D frame. Features can be analyzed directly. An example is identifying a rotated and scaled product in a scene. They can also be used for more complex analysis such as recovering structure from motion. In the below image we can see a direct use of features for finding a product in a 2D frame. Although the product is partially covered, many feature match. These are shown as green lines. The spatial relationship between matched features use useful for ensuring valid matches.



3D Retrieval Method Based on Spatial+LDA Model

Latent Dirichlet Allocation (LDA) is a generative probabilistic model for the collections of discrete data. LDA is composed of a three-layer hierarchical Bayesian model. One such method takes image features, places them into a topic model using LDA, and then uses the spatial distances as a descriptor for similarity measurement (Nie et al., 2017). This method is useful when searching, comparing, or retrieving segments of customer video content.

Going Deeper, Beyond 2D Frames to Spatial Reconstruction

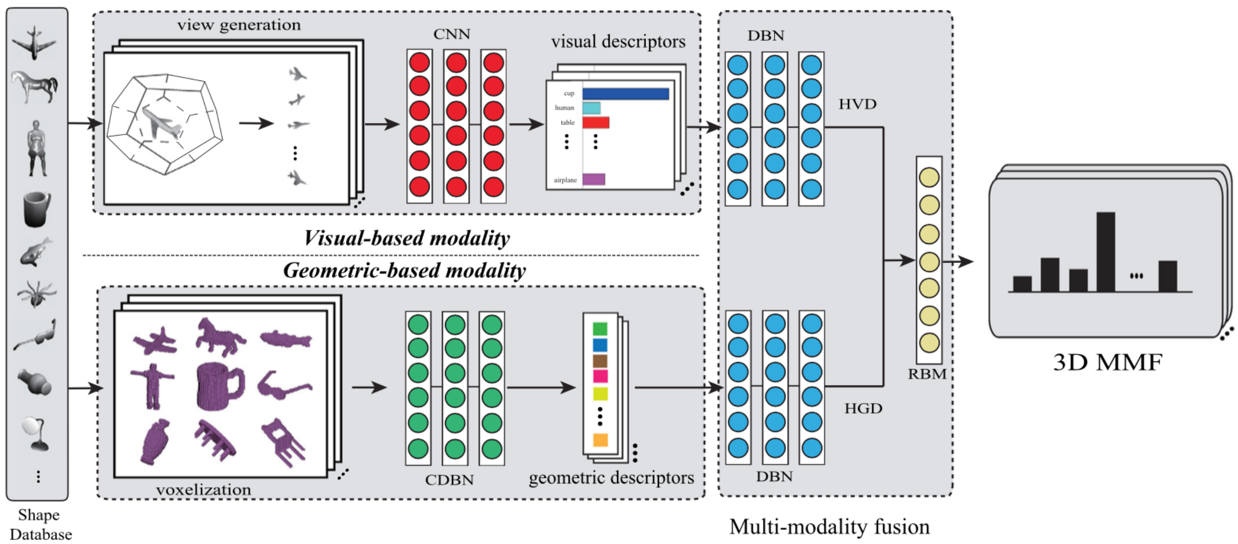
Features have utility, but to gain scene understanding, towards customer preference understand, requires going deeper. The literature points to spatial reconstruction as the next layer in our deep dive journey. Spatial reconstruction allows extraction of lost spatial information from a sequence of images. Meshes are geometric structures that capture relative spatial relationships. Meshes are surfaces and surfaces represent what we know as objects or groups of objects.

Meshes are not an end but rather an intermediate product in our marketing analysis workflow. We can find meshes using ‘N-view automatic image orientation’. This process starts by identifying feature matches between images. These are filtered using Random Sample Consensus algorithms, and spatial reconstruction is carried out using inlier points. The next step in this process is called bundle adjustment, an activity with the purpose of minimizing model errors (Somogyi, 2017).

The final step uses the sparse point cloud to calculate the camera positions. With the camera position known, energy-based methods can generate a dense reconstruction. This means that each pixel in the source image can aid in generating a detailed surface mesh, creating a spatial accurate model of an object as compared to laser scans that produce the same relative geometry. The author references that this processing is nearly identical across several available software libraries: COLMAP, Visual SfM, CMPMVS, and MeshRecon (Somogyi, 2017).

Objects From Mesh: a Multi-Modal Approach

The next step in this journey is moving from mesh structures to objects. The approach researched by Bu et al., uses Convolutional Deep Belief Networks (CDBNs) to learn 3D shapes from geometry-based modality and Convolutional Neural Networks (CNNs) to learn from the view-based modality.



The above flowchart shows how the renderings of the 3D mesh or the original images are input into the CNNs. The lower frame shows how the mesh shapes are converted to a volumetric representation which is input into the CDBN. The output of this process is two high-level descriptors, one for the visual information, and another for the geometric information. Deep Belief Networks are used to fuse the two descriptors, which are then processed by a Restricted Boltzmann Machine. The output of this process is a muti-modality feature, like those discussed earlier, that represents the scene (Bu et al., 2017).

Mining Descriptors To Drive Dynamic Content Generation

The next step in this journey is moving from extracting preference from descriptors to enhancing customer experience. Dynamic content generation is driven by an understanding preference. As an example, encoded preference information can be an input to a dynamically generated car advertisement and used to change the attributes of the person riding in the back ofa car (Armin, 2017).

Conclusion

This paper covering methods useful for generating deep machine understanding of consumer-produced video content. Individual unstructured data contains information that can be encoded in the form of descriptors. By capturing these descriptors, we can unlock a wealth of specific information about customers, including their preferences. The visual illustration from the Automotive Industry marketing journal showed how altering attributes of dynamically generated car media content could increase individual appeal and engagement.

References

Nei, W., Li X., and Su, Y. (2017). 3D object retrieval based on Spatial+LDA model. *Multimed Tools Appl, 76, 4091-4104,* doi 10.1007/s11042-015-2840-x

Somogyi, A., Lovas, T., and Barsi, A. (2017). Comparison of Spatial Reconstruction Software Package Packages Using DSLR Images. *Pollack Periodica: An International Journal for Engineering and Information Sciences, Vol 12 No 2, 17-27,* doi 10.1556/606.2017.12.2.2

Bu, S. et. al. (2017). 3D shape recognition and retrieval based on multi-modality deep learning. *Neurocomputing, 259, 183-193,* doi 10.1016/2016.06.088

Armin Pohl. (2017). Computer generated images are the next marketing platforms for motor behicles. *Automotive Industry Online, 259, 48-49,* accessible via http://www.ai-online.com/Adv/Previous/show_issue.php?id=7018