Week 8: Video Editing With AI

Nate Bachmeier

TIM-7211: Research Design Methodologies

February 14, 2021

Northcentral University

# Video Editing with AI

Movie and television studios spend millions of dollars annually removing objects from videos (Shimamura et al., 2020). This painstaking process requires humans to transpose pixels between different shots and then handle any differences, such as variations in lighting or color pallets. Businesses need mechanisms that automate this challenging process. This constructive research seeks to produce an algorithm that serves as that mechanism.

# Problem Statement

Production studios enjoy shooting video in the real world because it creates an authentic feeling and saves money. However, the camera also records licensed content, pedestrians, and other distracting objects within these shots. Video Effects (VFX) teams clean up the images by carefully transposing pixels between different frames (Trinh et al., 2019). When the scene contains fast-paced action, a compounding of effort is required. For instance, VFX teams need to account for changes to these objects due to different lighting, object orientation, and dynamic positioning.

Today, VFX teams meet these requirements through painstaking manual processes that require 3-D modeling and sophisticated software. These programs are complex and come with a high barrier to entry, which in turn increases costs. Businesses would prefer to have artificial intelligence solutions manage these tedious tasks. This capability would reduce costs, enable faster time-to-market, and free up VFX resources for value-differentiating services.

The most powerful artificial intelligence applications use machines to enhance human capabilities rather than replace them (Heer, 2019; Boire, 2017). For instance, a person can write a more profound business case than a machine; however, the same machine will have fewer misspellings and grammatical errors. This dichotomy exists because humans specialize in contextualizing thought versus automation uses patterns to make predictions (Schleer et al., 2019).

# Methodological Approach

There are four approaches to studying a business use-case or phenomena (see Table 1). Software engineers are builders, and we learn best through hands-on experimentation, making constructive research the most appealing selection. Constructive design is also one of the most common research methods for information systems and technology (Silvestrini et al., 2012). These studies identify a problem, build solution artifacts, and communicate the implementation’s unique value (Hevner et al., 2004). Typically, these results (artifacts) originate from specific Proof-Of-Concept (POCs) or directed case studies.

Table 1: Research Strategies

|  |  |  |
| --- | --- | --- |
| Approach | Description | Study Example |
| Quantitative | Studies the magnitude of a phenomena | Measure the time required for object removal tasks |
| Qualitative | Explores a concept without a numerical basis | Exploration of reasons objects require removal |
| Mixed-Method | Combines exploration and studying the magnitude of these issues | What preparation steps reduce the costs of object removal |
| Constructive | Produce artifacts to study a scenario | Create an algorithm for removing objects |

## Artifacts Production

A critical challenge with a constructive design is properly scoping the problem. If the research is too broad, then it will not complete on time. This research project seeks to produce a reusable library of video transforms powered by artificial intelligence. However, building something entirely from scratch is out-of-scope. Instead, these capabilities must extend mainstream open-source packages like Open Computer Vision (OpenCV) or similar frameworks.

The study will also produce a comprehensive review of state-of-the-art approaches. For instance, Generative Adversal Networks (GAN) and Auto Encoders (AE) use machine learning to create deep fake images. Can these sophisticated systems learn to identify commonly removed objects and “auto clean” the video? Alternatively, are there algorithms that accelerate the processing times of more traditional manual strategies?

## Addressing the Problem

The initial proof-of-concept needs to provide capabilities that identify and remove objects from video files. These features will reduce the efforts of VFX teams and allow them to focus on value-differentiating tasks. When a project requires fewer resources, it translates into lower costs for businesses and faster time-to-market. All of these benefits directly address the core problem statement.

Due to time and resource constraints, the artifacts will not address the entire problem. For instance, removing a static positioned object could require ten units of effort. Handling a dynamically moving object within 3-D space might be four times more complex. An initial investigation must take place to ensure the sub-problem(s) are helpful and adequately sized. Ideally, another researcher or engineering team can consume these artifacts to build their custom solutions.

## Measuring Success

The Key Performance Indicator (KPI) is the amount of human time required to clean up a video. The most direct way of obtaining this value is by measuring the time it takes VFX experts to process scenes under traditional methods and AI-enhanced methods. Hence, the differences in time and effort are quantifiable.

Performing a job quickly but poorly does not provide value. It is out-of-scope to create a custom grading solution to this problem. Instead, there needs to be research into an existing mechanism for grading the accuracy of pixel-level predictions.

# Ethical Issues

Ethical challenges exist across multiple fronts, such as malicious use-cases, potentially copyrighted content, and ensuring verifiable results.

## Malicious Uses

The primary use-case of this research is to simplify removing objects from video files. There are potential risks that end-users abuse this technology to create misleading content and spread it across social media. Another malicious attack vector exists where the tooling removes copyrights and other watermarks, weakens low-tech anti-piracy protections.

## Copyrighted Content

Sample videos used during the study could be copyrighted, and that limits other researcher's access. There are several options for mitigating these risks, such as using public domain content or filming proprietary videos. Several footages produced with governmental funding or held by a deceased copyright holder are also fair use. Given the breadth of choices, the study can easily avoid these legal pitfalls.

## Result Validity

The four major threats to research validity are internal, external, statistical conclusions, and construct validity (see Table 2). According to Parker, “it is widely accepted truism that all published research to some extent is flawed. Because the research enterprise is fraught with many pitfalls, researchers must become well-versed in recognizing, and when possible avoiding design shortcomings (Parker, 1993, p. 1).”

Table 2: Threat Sources

|  |  |
| --- | --- |
| Source | Description |
| Internal Threat | Contamination by the research team |
| External Threat | Contamination outside of the study’s controls |
| Statistical Conclusion Validity | Results are arbitrary or non-reproducible |
| Construct Validity | Controls are not enforceable or consistent |

The study’s effectiveness requires comparing cleaning a video with the study’s artifacts against existing software applications. Those sophisticated software packages have expensive licensing fees, preventing some participants from having access. It is also possible that the VFX professionals find external factors (e.g., the UI) more natural and influential toward cleaning up the scenes. These limitations could make it impossible for external researchers to reproduce the findings.

While these factors are not controllable, it is possible to identify and record them through a structured survey. For example, each participant must provide their background information and level of comfort with relevant software packages. That will aid future artifact consumers in understanding their likelihood of achieving similar results.

# Conclusions

Production studios spend millions of dollars annually on post-processing films. One of the most common tasks is removing objects from videos, which is tedious and complicated. Businesses can improve time-to-market and reduce costs by relying on artificial intelligence solutions to make those changes. After freeing up VFX resources, those employees can focus on value-differentiating tasks that benefit from human creativity.

There are multiple methods for studying this problem. However, a constructive design is the preferred option. Software engineers are builders, and we learn best through hands-on experimentation. Under a constructive design, the research needs to produce reusable artifacts, such as proof-of-concept tools and software libraries. While creating these resources, researchers need to be cognizant of any ethical challenges that prevent reproducible results.

# References

Boire, R. (2017). Artificial Intelligence, automation, and its impact on data science. *IEEE International Conference on Big Data* (pp. 3571-3574). Boston, MA: Institute of Electrical and Electronics Engineers. doi:10.1109/BigData.2017.8258349

Heer, J. (2019). Agency plus automation. *Proceedings of the National Academy of Sciences of the United States of America, 116*(6), 1844-1850. doi:10.1073/pnas.1807184115

Hevner, A., March, S., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS Quarterly, 28*(1), 75-105. doi:10.2307/25148625

Le, T., Almansa, A., Gousseau, Y., & Masnou, S. (2019). Object removal from complex videos using a few annotations. *Computational Visual Media, 5*(3), 267-291. doi:10.1007/s41095-019-0145-0

Parker, R. (1993). Threats to the validity of the research. *Rehabilitation Counseling Bulletin, 36*(3), 130-138. Retrieved from https://search-ebscohost-com.proxy1.ncu.edu/login.aspx?direct=true&db=eric&AN=EJ458938&site=eds-live

Schleer, P., Drobinsky, S., de la Fuente, M., & Radermacher, K. (2019). Toward versatile cooperative surgical robotics: a review and future challenges. *International Journal of Computer Assisted Radiology and Surgery, 14*(10), 1673-1686. doi:10.1007/s11548-019-01927-z

Shimamura, R. F. (2020). Audiovisual object removal in 360-degree videos. *The Visual Computer: International Journal of Computer Graphics, 10-12*. doi:10.1007/s00371-020-01918-1

Silvestrini, R. P., & Sammito, G. (2012). Design of Experiments for Information Technology Systems. *Defense AT&L, 41*(5), 30-35. Retrieved from https://search-ebscohost-com.proxy1.ncu.edu/login.aspx?direct=true&db=bth&AN=80409129&site=eds-live