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 offset any differences, such as variations of lighting or color pallets. Businesses require a mechanism to 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 shots 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 different lighting, object orientation, and non-static positioning.

Today, VFX teams meet these requirements through painstaking manual processes that require 3-D modeling and sophisticated software projects. 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 survey of the current state-of-the-art approaches. For instance, Generative Adversal Networks (GAN) and Auto Encoders (AE) use machine learning to produce 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 exist to ensure the sub-domain useful and adequately sized. Ideally, another researcher or engineering team can consume these artifacts to build their custom solution with them.

## Measuring Success

There need to be benchmarks that measure the effectiveness of the proposed solution. For this particular use-case, the Key Performance Indicator (KPI) is the amount of human time required to clean-up a shot. The most direct way of obtaining this value is with VFX experts. They will need to process scenes under traditional methods and AI-enhanced strategies. Hence, the differences in time and effort are quantifiable.

# Ethical Issues

Ethical challenges exist across multiple issues like malicious use-cases, potentially copyrighted content, and ensuring verifiable results.

## Malicious Uses

The primary use-case of this constructive design approach reduces the effort required to remove objects from video files. There are potential risks that end-users abuse this technology to create misleading content and spreading it across social media. Another malicious attack vector could exist by removing copyrights and watermarks within videos. That capability would weaken low-tech anti-piracy protections.

## Copyrighted Content

Sample videos used during the study might be copyrighted, limiting others’ ability to reproduce the results. Mitigating these challenges requires the researcher to film example content or use public domain content. For instance, footage produced with governmental funding often comes with a fair-use policy. Many older movies are also fair-use if the copyright holder is deceased for at least seventy years.

## 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 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 a comparison of cleaning a video with the 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 study’s 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 film. 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 these changes. After freeing up VFX resources, those employees can focus on value-differentiating 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. The most likely candidate is library