Video Editing Tools
with Artificial Intelligence
Video Editor

आर्टिफिशियल इंटेलिजेंस के साथ वीडियो संपादन उपकरण

हाचिक हांzadzhian
सोफिया तकनीकी विश्वविद्यालय, शाखा
प्लोवदीव
कंप्यूटर सिस्टम और प्रौद्योगिकी विभाग

अमूर्त . वीडियो संपादन एक बहुत ही कठिन कार्य हो सकता है, इसलिए इसमें कोई आश्चर्य की बात नहीं है कि वर्कफ़्लो को सुव्यवस्थित करने या कठिन कार्यों को स्वचालित करने के लिए कृत्रिम बुद्धिमत्ता का उपयोग तेजी से किया जा रहा है। हालाँकि, शोध साहित्य में कौन से बुद्धिमान वीडियो संपादन उपकरण मौजूद हैं और वीडियो संपादकों की स्वचालन आवश्यकताओं का अवलोकन करना बहुत मुश्किल है। इसलिए, हमने शोध में स्मार्ट वीडियो संपादन टूल के क्षेत्र की पहचान की है और पेशेवर वीडियो संपादकों की राय का सर्वेक्षण कर रहे हैं। हमने कृत्रिम बुद्धिमत्ता अनुसंधान में कला की वर्तमान स्थिति को भी संक्षेप में प्रस्तुत किया है ताकि यह पता लगाया जा सके कि वास्तव में बुद्धिमान वीडियो प्रसंस्करण उपकरणों की दिशा में क्या संभावनाएं और वर्तमान तकनीकी सीमाएं हैं। निष्कर्ष बुद्धिमान वीडियो संपादन टूल के क्षेत्र की समझ में योगदान करते हैं, स्वचालन, अनुसंधान के लिए अधूरी जरूरतों को उजागर करते हैं, और बुद्धिमान वीडियो संपादन टूल पर आगे के शोध के लिए सामान्य सुझाव प्रदान करते हैं।

कीवर्ड: एआई, वीडियो संपादन, एडोब प्रीमियर,
दा विंची संकल्प

I. परिचय

वीडियो इंटरनेट पर सामग्री का सबसे लोकप्रिय रूप है। विज़ुअल नेटवर्क इंडेक्स के अनुसार, 2023 में इंटरनेट ट्रैफ़िक का 82.5% वीडियो सामग्री था। मोबाइल फ़ोन, वीडियो शेयरिंग और सोशल मीडिया प्लेटफ़ॉर्म वीडियो लेना और पोस्ट करना पहले से कहीं अधिक आसान और तेज़ बनाते हैं। हालाँकि, इन वीडियो को संपादित करने में अभी भी बहुत अधिक समय लगता है। वीडियो को संपादित करना कठिन है क्योंकि इसमें ऑडियो और छवि दोनों के साथ दो-ट्रैक माध्यम होने के अलावा, अलग-अलग फ्रेम पर काम करने की आवश्यकता होती है। वीडियो संपादन को आसान बनाने के लिए कई प्रयास किए जा रहे हैं। एक दृष्टिकोण कृत्रिम बुद्धिमत्ता (एआई) का उपयोग करके संपादन प्रक्रिया को स्वचालित करना है।

यहां हम वीडियो संपादन स्वचालन में अत्याधुनिक में रुचि रखते हैं, विशेष रूप से जो वांछित है और जो वर्तमान एआई तकनीक के साथ प्राप्त करने योग्य है, के बीच बेमेल पर ध्यान केंद्रित कर रहे हैं।

ओकुन एट अल. [ओकुन एट अल., 2015] वीडियो संपादन को एकल संपादित फिल्म बनाने के लिए एक या अधिक स्रोतों के हिस्सों को काटने और जोड़ने के कार्य के रूप में परिभाषित किया गया है। वीडियो संपादन टूल को मोटे तौर पर (कंप्यूटर) प्रोग्राम के रूप में परिभाषित किया जा सकता है जिसका उपयोग लोग वीडियो संपादन का कार्य करने के लिए कर सकते हैं, अर्थात। वीडियो खंडों का संयोजन. वीडियो संपादन एक ऐसा क्षेत्र है जहां मानव वीडियो संपादकों के कार्यों को स्वचालित करने या बढ़ाने के लिए एआई का उपयोग किया जा रहा है।

स्मार्ट वीडियो एडिटिंग टूल डिजिटल वीडियो एडिटिंग की शुरुआत से ही वीडियो एडिटिंग को आसान बनाने की कोशिश कर रहे हैं। बुद्धिमान वीडियो संपादन टूल में प्रमुख विषयों में से एक फ़्रेम के बजाय फ़्रेम और डायलॉग जैसे उच्च-स्तरीय अमूर्त से वीडियो हेरफेर की अनुमति देने की समस्या है। इस तरह के टूल का एक प्रारंभिक उदाहरण 2002 का सिल्वर [कैसारेस एट अल., 2002] है, जो वीडियो मेटाडेटा का उपयोग करके वीडियो क्लिप के बुद्धिमान चयन के साथ-साथ वीडियो संपादन के अमूर्त दृश्य प्रदान करता है। स्मार्ट वीडियो एडिटिंग टूल का एक हालिया उदाहरण रफकट है [लीक एट अल., 2017]। रफकट उपयोगकर्ता द्वारा दर्ज किए गए दृश्य संवाद, कच्चे फुटेज और संपादन मुहावरों का उपयोग करके संवाद-संचालित दृश्यों के लिए कम्प्यूटेशनल संपादन को सक्षम बनाता है। एक ओपन सोर्स टूल ऑटोएडिट [पासरेल्ली, 2019] और रिसर्च [बर्थौज़ोज़, 2012] है जो वीडियो के साथ टेक्स्ट ट्रांसक्रिप्ट को जोड़कर वीडियो साक्षात्कार के टेक्स्ट संपादन की अनुमति देता है।

पूरी तरह से एआई-नियंत्रित वीडियो उत्पादन ने हाल ही में बहुत अधिक शोध रुचि को आकर्षित किया है [ज़ू एट अल.; हुआ एट अल., 2004]। वर्तमान में, एआई-नियंत्रित वीडियो उत्पादन स्वचालित वीडियो सारांश या मिश्रण बनाने पर केंद्रित है। ये पूरी तरह से स्वचालित वीडियो संपादन विधियां, जैसा कि वीडियो सारांश और समग्र एप्लिकेशन बनाने के लिए उपयोग किया जाता है, को बुद्धिमान वीडियो संपादन उपकरण नहीं माना जाता है, क्योंकि वे सरल एल्गोरिदम हैं जो एक बहुत ही संकीर्ण और विशिष्ट क्वेरी निष्पादित करते हैं जिसके लिए खुफिया या उपयोगकर्ता इंटरैक्शन की आवश्यकता नहीं होती है।

छवि प्रसंस्करण, दृष्टि और प्राकृतिक भाषा प्रसंस्करण में कृत्रिम बुद्धिमत्ता में प्रगति हुई है

वीडियो संपादन. लेकिन क्या वीडियो संपादन के कठिन परिश्रम से छुटकारा पाने का सपना सच हो गया है? बेशक, इसका उत्तर इस बात पर निर्भर करता है कि हम किसके सपने के बारे में बात कर रहे हैं। स्वचालित वीडियो संपादन में प्रगति का मूल्यांकन कैसे किया जा सकता है? स्वचालित वीडियो संपादन में प्रगति का आकलन करने के इस प्रश्न पर दो कोणों से विचार किया जा सकता है। पहला है साहित्य की समीक्षा करना, जबकि दूसरा है मानव वीडियो संपादकों की अपेक्षाओं की जांच करना और उनकी तुलना कृत्रिम बुद्धिमत्ता में कला की स्थिति से करना। इस लेख में, हम दोनों करते हैं।

स्मार्ट वीडियो संपादन टूल ने जिस मुख्य चुनौती को हल करने का प्रयास किया वह उपयोगकर्ताओं के लिए वीडियो संपादन प्रक्रिया को सुव्यवस्थित करना था। यह आमतौर पर वीडियो को फ़्रेम दर फ़्रेम देखने के कठिन कार्यों को समाप्त करके किया जाता है। प्रस्तावित समाधान और उपकरण कई मायनों में भिन्न हैं जो समस्या के दृष्टिकोण, इच्छित लक्ष्य, अंतर्निहित तकनीक, अमूर्तता के स्तर, प्रस्तावित इंटरैक्शन और उक्त इंटरैक्शन के तौर-तरीकों से उत्पन्न होते हैं।

हमने दो दृष्टिकोणों से अत्याधुनिक वीडियो एआई अनुप्रयोगों की समीक्षा की: i) सामान्य वीडियो एआई तकनीक;

ii) विशिष्ट एआई वीडियो संपादन तकनीक। वीडियो के लिए सामान्य एआई तकनीक में विभिन्न प्रकार के वीडियो कार्य शामिल हैं जैसे ऑब्जेक्ट ट्रैकिंग, ऑब्जेक्ट डिटेक्शन, स्पीच रिकग्निशन, वीडियो रीजनिंग, एक्शन डिटेक्शन, वीडियो में भावना का पता लगाना। वीडियो संपादन के लिए विशिष्ट एआई तकनीक बहुत संकीर्ण है जैसे वीडियो स्क्रिप्ट, फ़्रेम और दृश्यों को संसाधित करना और वीडियो संपादन खनन नियम।

हमने 13 वीडियो संपादकों का सर्वेक्षण किया जिनका वीडियो संपादन अनुभव 1 वर्ष से 22 वर्ष तक था। सर्वेक्षण में वीडियो संपादन में उनके अनुभव, एआई वीडियो संपादक पर विचार और वीडियो संपादन स्वचालन आवश्यकताओं को शामिल किया गया है। सर्वेक्षण प्रतिक्रियाओं का उपयोग वीडियो संपादन टूल में स्वचालन के संबंध में अपेक्षाओं, आवश्यकताओं और मुद्दों का अवलोकन संकलित करने के लिए विषयगत विश्लेषण करने के लिए किया जाता है। हम सर्वेक्षण में सामने आई राय और अपेक्षाओं की तुलना सामग्री निर्माण/हेरफेर, स्वचालित वीडियो संपादन और अन्य एआई टूल के लिए मशीन लर्निंग के वर्तमान ज्ञान से करते हैं। हम चर्चा करते हैं कि कैसे नवीनतम एआई वीडियो निर्माताओं के लिए एक आदर्श एआई वीडियो संपादन उपकरण बना सकता है। यह पेपर इस प्रकार संरचित है: खंड 2 वीडियो में बुद्धिमान वीडियो संपादन टूल और एआई तकनीकों का अवलोकन प्रदान करता है। साहित्य में इंटेलिजेंट वीडियो संपादन टूल की तुलना और सारांश धारा 3 में किया गया है। प्रक्रिया और सारांशित परिणामों सहित (मानव) वीडियो संपादकों का सर्वेक्षण धारा 4 में है। धारा 5 में, इंटेलिजेंट के लिए समाधान पर पिछले कार्य वीडियो संपादन टूल और सर्वेक्षण से उपयोगकर्ता की अपेक्षाओं की तुलना की गई है और कुछ एआई तकनीकों को उपयोगकर्ता की अपेक्षाओं को पूरा करने के लिए संभावित समाधान के रूप में प्रस्तावित किया गया है। अंत में, पेपर सारांश, निष्कर्ष और भविष्य का कार्य खंड 6 में प्रस्तुत किया गया है।

II. उपयोग में आसानी

डिजिटल वीडियो की शुरुआत के बाद से वीडियो संपादन की सुविधा के लिए बेहतर उपकरण बनाना हमेशा एक शोध एजेंडा रहा है।

सबसे पहले, हमने बुद्धिमान और स्वचालित वीडियो प्रकाशन उपकरण बनाने के विभिन्न तरीकों की समीक्षा की। इसके बाद, हम एआई तरीकों की ओर रुख करते हैं जिन्हें प्रत्यक्ष या अप्रत्यक्ष रूप से वीडियो संपादन में लागू किया गया है।

वीडियो संपादन को सरल बनाने की कोशिश करने वाली पहली परियोजनाओं में से एक सिल्वर थी [मायर्स एट अल., 2001; लॉन्ग एट अल., 2004]। सिल्वर प्रोजेक्ट के पहले संस्करण में, विभिन्न प्रकार के दृश्य हैं, अर्थात् ट्रांसक्रिप्ट, टाइमलाइन, पूर्वावलोकन और स्क्रिप्ट दृश्य।

यह एडिटिंग टूल फ़ोटो और दृश्य पहचान का उपयोग करके स्मार्ट चयन, कट, डिलीट, कॉपी, पेस्ट और री-अटैच के साथ स्मार्ट एडिटिंग की भी खोज करता है। मेटाडेटा उत्पन्न करने के लिए टेक्स्ट ट्रांसक्रिप्शन, शॉर्ट बाउंड्री डिटेक्शन और ऑप्टिकल कैरेक्टर रिकग्निशन (ओसीआर) के रूप में वीडियो की मेटाडेटा परत का उपयोग करके स्मार्ट संपादन बनाया गया है। सिल्वर टूल के दूसरे पुनरावृत्ति में [लॉन्ग एट अल।, 2004], टूल ने दो वीडियो सेगमेंट में शामिल होने पर कट करने के लिए सही फ्रेम को विज़ुअलाइज़ करने जैसे आसान कार्यों को बनाने के लिए लेंस (क्लिप, फोटो, फ्रेम) और सिमेंटिक स्केलिंग को लागू किया। अधिकांश स्मार्ट वीडियो टूल केवल एक विशिष्ट प्रकार के वीडियो बनाने के लिए डिज़ाइन किए गए हैं। उदाहरण के लिए, QuickCut [Truong et al., 2016] को वर्णित वीडियो की रचना के लिए बनाया गया था, और वीडियो डाइजेस्ट्स [Pavel et al., 2014] को वीडियो व्याख्यानों को सारांशित करने के लिए बनाया गया था। इसके बाद, हम पूरी तरह से स्वचालित वीडियो संपादन के तरीकों को देखेंगे। स्वचालित वीडियो संपादन मानव संपादक से किसी भी इनपुट के बिना वीडियो खंडों की कम्प्यूटेशनल प्रोसेसिंग और कंपोज़िंग है। स्वचालित वीडियो संपादन रिकॉर्ड किए गए वीडियो पर या, बहुत बड़े पैमाने पर, वीडियो संग्रह पर किया जा सकता है। स्वचालित वीडियो संपादन पर प्रारंभिक कार्य वीडियो अनुक्रम उत्पन्न करने के लिए नियम-आधारित रणनीतियों पर केंद्रित था [बटलर और पार्क्स, 1997] या वीडियो वृत्तचित्रों के क्षेत्र में उपयोगकर्ता अनुरोधों को चुनने और स्वचालित रूप से संपादित करने के लिए एक अर्थ-आधारित विधि [बोकोनी, 2004]। मैशअप, एक ही इवेंट के लिए कई वीडियो का संयोजन, एक अन्य प्रकार का स्वचालित वीडियो संपादन है। वर्चुअल डायरेक्टर [श्रेष्ठ एट अल., 2010] नामक कार्य ने कॉन्सर्ट रिकॉर्डिंग के लिए मिश्रण तैयार करने की एक विधि बनाई

यह वीडियो संपादकों और फिल्म व्याकरण साहित्य के साक्षात्कार से प्राप्त नियमों के आधार पर एक अच्छे कॉन्सर्ट वीडियो को अधिकतम बनाता है। लाइव इवेंट प्रसारित करने के लिए स्वचालित वीडियो संपादन का भी उपयोग किया जा सकता है। [रादुत एट अल., 2020] के काम में न केवल लाइव इवेंट के लिए प्रोटोटाइप एआई वीडियो संपादक पर चर्चा की गई, बल्कि एआई-संपादित लाइव इवेंट की गुणवत्ता को मापने के लिए मूल्यांकन विधियों के मूल्यांकन और चर्चा भी की गई। AI का उपयोग स्वचालित रूप से संपादित वीडियो बनाने के लिए भी किया जा सकता है। मशीन द्वारा निर्मित जब एआई को बीबीसी के संग्रह का सामना करना पड़ा, तो इसने बीबीसी संग्रह से 150 छोटी जटिलताएँ पैदा कीं [आर एंड डी, 2018]। [तस्किर एट अल., 2006] ने वाक् पहचान प्रणालियों से वाक् प्रतिलेख का उपयोग करके वीडियो कार्यक्रमों को स्किमिंग करने के लिए एक सारांश विधि प्रस्तुत की। [ह्यूओंग एट अल., 2016] वीडियो अमूर्तन या सारांश विधियों का सारांश प्रदान करता है, जैसे कि कम से कम समय में किसी वीडियो के बारे में जानकारी प्रदान करने के लिए कीफ्रेम या चलती छवियों की एक श्रृंखला तैयार करना। [वू एट अल., 2020] द्वारा कॉर्पोरेट बैठकों के स्वचालित वीडियो संपादन पर काम मानव-संपादित वीडियो से सीखे गए संपादन निर्णयों का उपयोग करता है और ऑडियो और वीडियो दोनों पर दो ध्यान मॉडल का उपयोग करता है। ईडीएल- संपादन निर्णय सूची एक पाठ-आधारित भाषा है जो क्लिप और टाइमकोड डेटा की क्रमबद्ध सूची के साथ रचनात्मक निर्णयों को एन्कोड करती है। इसका उपयोग कई स्वचालित वीडियो संपादन प्रणालियों के आउटपुट के रूप में किया जाता है [टास्किर एट अल., 2006; वू एट अल., 2020; पासरेली, 2019], जिसके बाद वीडियो संपादक एडोब प्रीमियर प्रो या डेविंसी रिजॉल्व जैसे सॉफ्टवेयर में अपने संपादन को जारी रखने के लिए ईडीएल के साथ एन्कोड किए गए स्वचालित संपादन समाधान का उपयोग कर सकते हैं।

वीडियो से जानकारी निकालने के लिए एआई तकनीकों का उपयोग कैसे किया जा सकता है, यह शोध का एक बहुत ही विविध क्षेत्र है। हम विशेष रूप से चेहरा पहचान, वस्तु पहचान, वस्तु ट्रैकिंग, दृश्य पहचान, भावना विश्लेषण, वीडियो तर्क और वीडियो कैप्शनिंग में रुचि रखते हैं। चेहरा पहचानना यह पहचानने की समस्या को संदर्भित करता है कि क्या एक मानव चेहरा एक छवि में मौजूद है और संभवतः किसका है, जबकि वस्तु का पता लगाना एक छवि में एक विशिष्ट वस्तु की पहचान करने की समस्या है। ऑब्जेक्ट ट्रैकिंग किसी विशिष्ट ऑब्जेक्ट को पहचानने और उसका पता लगाने और वीडियो के फ्रेम के माध्यम से उसकी गति को ट्रैक करने की समस्या है। दृश्य पहचान या वीडियो विभाजन उन खंडों की पहचान है जो किसी वीडियो में शब्दार्थ या दृश्य रूप से संबंधित हैं। भावना विश्लेषण उस मनोदशा से मेल खाने की समस्या है जिसे सामग्री के एक टुकड़े द्वारा व्यक्त किया जाएगा: चाहे वह खुश हो, उदास हो, विडंबनापूर्ण हो, आदि। वीडियो कैप्शनिंग [वू एट अल।, 2016] एक एआई तकनीक है जो प्राकृतिक विवरण उत्पन्न करती है जो कैप्चर करती है वीडियो की गतिशीलता.

ऐसी वीडियो एआई तकनीकें हैं जो वीडियो संपादन के लिए अधिक विशिष्ट हैं। [मात्सुओ एट अल.] ने प्रतिलिपि प्रस्तुत करने योग्य संपादन पैटर्न बनाने के लिए वीडियो से संपादन पैटर्न (ढीले, मध्यम-तंग फ्रेम और नियमों से बना) का पता लगाने के लिए एक डेटा माइनिंग तकनीक प्रस्तुत की। [बटलर और पार्क्स, 1997] के पहले के काम ने वीडियो संपादन को स्वचालित करने के लिए एक नियम और क्वेरी-आधारित दृष्टिकोण प्रस्तुत किया था। संपादन प्रक्रिया को मॉडलिंग करके और शब्दार्थ का उपयोग करके स्वचालित वीडियो संपादन [नैक एंड पार्क्स, 1997] में प्रस्तुत किया गया है।

III. स्मार्ट वीडियो संपादन उपकरण

इस अध्याय में, हम साहित्य की समीक्षा करते हैं और संक्षेप में बताते हैं कि पिछले बुद्धिमान वीडियो संपादन उपकरण वीडियो संपादन को आसान कार्य बनाने की समस्या को कैसे तैयार और हल करते हैं। बुद्धिमान वीडियो संपादन टूल के लिए साहित्यिक खोज कंप्यूटर विज्ञान साहित्य डेटाबेस में कीवर्ड (बुद्धिमान या बुद्धिमान या स्वचालित या एआई) और (वीडियो संपादक या वीडियो संपादन) का उपयोग करके की गई थी, जो डीबीएलपी, एसीएम डिजिटल लाइब्रेरी और Google विद्वान हैं। शीर्षक और सार को तब समावेशन मानदंड के आधार पर पढ़ा और फ़िल्टर किया गया था, जिसमें यह निर्धारित किया गया था कि शामिल कागजात को उपयोगकर्ताओं के लिए वीडियो संपादन टूल बनाने के लिए एक बुद्धिमान दृष्टिकोण का वर्णन करना चाहिए। शामिल दस्तावेज़ों में उपयोगकर्ता इंटरफ़ेस का विवरण और/या कार्यान्वयन भी शामिल होना चाहिए। अधिक संबंधित साहित्य खोजने के लिए शामिल कागजात के संदर्भों को भी स्कैन किया गया। परिणामी पेपरों को संक्षेप में प्रस्तुत किया गया और तीन विषयों में समूहीकृत किया गया, अर्थात् वीडियो संपादन कार्य, एक स्वचालित संपादक के साथ बातचीत (मानव-कंप्यूटर इंटरैक्शन), और एआई तकनीक। इस अनुभाग में शामिल कागजात की सूची तालिका 1 में है

A. वीडियो संपादन कार्य - स्मार्ट वीडियो संपादकों के

यह उपखंड वीडियो संपादन वर्कफ़्लो में संबोधित विभिन्न कार्यों के संदर्भ में बुद्धिमान वीडियो संपादन टूल के क्षेत्र का परिचय देता है और प्रत्येक कार्य के लिए दृष्टिकोण का सारांश देता है। वीडियो विभाजन सबसे आम कार्य है जिसे स्मार्ट वीडियो संपादक हल करने का प्रयास करते हैं। इस अनुभाग में समीक्षा किए गए सभी पिछले कार्यों में किसी न किसी प्रकार की वीडियो विभाजन विधि का उपयोग किया गया है, लेकिन विभाजन करने के लिए विभिन्न तरीकों का उपयोग किया जाता है। अब से, एक वीडियो को एकल स्रोत फ़ाइल से वीडियो के एक सतत खंड के रूप में परिभाषित किया गया है। पहचानी गई पहली विभाजन विधि शॉट डिटेक्शन का उपयोग थी।

एक फ्रेम ओकुन एट अल की छवियों का एक निरंतर निरंतर अनुक्रम है। [2015]। शॉट डिटेक्शन सेगमेंटेशन छवि विश्लेषण विधियों के साथ किया जाता है [कैसरेस एट अल., 2002; लॉन्ग एट अल., 2004] और [शिपमैन, 2008] में कैमरे की गति, चमक और अवधि जैसी सुविधाओं का उपयोग करके शॉट का पता लगाया जाता है। [वू एट अल., 2015] उपयोगकर्ता-जनित वीडियो को खंडित करने के लिए फ्रेम और सबफ्रेम का उपयोग करता है (एक सबफ्रेम को वीडियो की एक मूल इकाई के रूप में परिभाषित किया गया है जिसमें लगातार कैमरा मूवमेंट और स्व-निहित शब्दार्थ शामिल हैं)। [कैसरेस एट अल., 2002] एल-कट के मामले में विभिन्न स्थानों पर वीडियो और ऑडियो के विभाजन पर भी विचार करें।

दूसरी विभाजन विधि वीडियो ऑडियो की सामग्री या अर्थ पर निर्भर कट बनाने के लिए टेक्स्ट ट्रांसक्रिप्ट के साथ सिंक्रनाइज़ेशन का उपयोग करती है। [लीक एट अल., 2017] ने वीडियो को खंडित करने के लिए दृश्य संवाद स्क्रिप्ट से पूर्व-लिखित पंक्तियों का उपयोग किया, जबकि [पावेल एट अल., 2014] ने बायेसियन विषय विभाजन के माध्यम से पाठ प्रतिलेखन चलाया [ईसेनस्टीन और बार्ज़िले, 2008] पहचानने के लिए बीएसएक सूचनात्मक वीडियो के अनुभाग और उपखंड।

अन्य विभाजन दृष्टिकोणों में टकटकी का उपयोग करना जैसे दृष्टिकोण शामिल हैं [किमुरा एट अल।, 2005], उपयोगकर्ता-चिह्नित बिंदुओं से शुरू करना, उन बिंदुओं के लिए फ्रेम की समानता के उपाय [ची एट अल।, 2013], और बात करने वाले सिर साक्षात्कार में कमी के लिए फिटनेस मूल्यांकन वीडियो [बर्टुसो, 2012]। [कैटेलन एट अल., 2008] द्वारा उपयोगकर्ता-जनित वीडियो सारांश ने उपयोगकर्ता की देखने की गतिविधियों और टिप्पणियों का उपयोग करके विभाजन बनाया। अगले कार्य पर चर्चा करने के लिए वीडियो विभाजन भी आवश्यक है, जो कि वीडियो खंडों की रचना करना है।

एआई वीडियो संपादन टूल की हमारी सूची में वीडियो सेगमेंट बनाना दूसरा सबसे आम कार्य है। वीडियो खंडों की संरचना को सुव्यवस्थित करने का सबसे आम तरीका इन-सीन संवादों [लीक एट अल., 2017] या टेक्स्ट ट्रांसक्रिप्शन्स [बर्थौज़ोज़, 2012; वू एट अल., 2015] रचना के शुरुआती बिंदु के रूप में। वीडियो संवाद को इनपुट के रूप में लिखा और प्रदान किया जाना चाहिए, लेकिन वाक् पहचान तकनीक का उपयोग करके एक पाठ प्रतिलेख तैयार किया जा सकता है।

उदाहरण के लिए, [ट्रूओंग एट अल., 2016] ने मैन्युअल रूप से बनाए गए टेक्स्ट ट्रांसक्रिप्ट के बजाय सुनाई गई ऑडियो या वॉयस-ओवर से परिवर्तित टेक्स्ट ट्रांसक्रिप्ट का उपयोग किया। रफकट [लीक एट अल., 2017] में, ऑटोमेशन का उपयोग करके प्रत्येक संबंधित संवाद और स्पीकर के लिए वीडियो सेगमेंट बनाए गए। संवाद क्रम प्रदान की गई स्क्रिप्ट का अनुसरण करता है। हालाँकि, रचनात्मक संरचना को उस उपयोगकर्ता द्वारा बदला जा सकता है जो वीडियो संपादन मुहावरों का संयोजन चुनता है। इसी तरह, संपादक द्वारा बनाई गई कहानी स्कीमा का उपयोग खंडों की रचना के लिए किया गया था [ट्रूओंग एट अल., 2016]। वीडियो रिकॉर्डिंग से संबंधित पाठ का चयन करके अवांछित भागों, जैसे साक्षात्कार के कुछ वाक्यांश या दोहराए गए शब्दों को काटकर विभाजन किया जाता है [बर्थौज़ोज़, 2012]।

[ची एट अल., 2013] सेगमेंट को व्यवस्थित करने और बनाने के तरीके के रूप में वीडियो क्लिप में उपयोगकर्ता द्वारा प्रदत्त टैग का उपयोग करता है, और संपादक डेमो वीडियो में चरणों के अनुरूप टैग को व्यवस्थित करके समग्र सेगमेंट की संरचना को बदल सकते हैं। उपयोगकर्ता-जनित वीडियो सारांश बनाने के उद्देश्य से संरचना टकटकी से मॉडल किए गए दर्शक इरादे का उपयोग करके की जाती है [किमुरा एट अल।, 2005] और गतिविधियों और उपयोगकर्ता टिप्पणियों को देखने के साथ [कैटेलन एट अल।, 2008]।

समयरेखा और वीडियो को विज़ुअलाइज़ करना समयरेखा और वीडियो को अमूर्तता के विभिन्न स्तरों पर देखने और वैकल्पिक तरीके से वीडियो समयरेखा के विभिन्न प्रतिनिधित्व, जैसे कि पाठ, के रूप में आता है। वीडियो रेंडरिंग एब्स्ट्रैक्शन फ्रेम, फ्रेम और क्लिप के रूप में हो सकते हैं।

फ्रेम एक वीडियो की स्थिर छवि है, जबकि फ्रेम छवियों का एक सतत अनुक्रम है [ओकुन एट अल., 2015]। वीडियो एब्स्ट्रैक्शन के विभिन्न रूपों का अवलोकन [ट्रूओंग और वेंकटेश] में उपलब्ध है। एक प्रतिनिधि फ्रेम का उपयोग करके क्लिप के विज़ुअलाइज़ेशन पर टाइमलाइन पर वीडियो सामग्री के त्वरित निर्णय की अनुमति देने की एक विधि के रूप में चर्चा की गई है [लॉन्ग एट अल।, 2004]।

[कैसारेस एट अल., 2002] अमूर्त के विभिन्न स्तरों में टाइमलाइन विज़ुअलाइज़ेशन प्रदान करता है, जो स्क्रिप्ट, संपादन योग्य ट्रांसक्रिप्ट और टाइमलाइन दृश्य हैं। समयरेखा विज़ुअलाइज़ेशन के लिए दूसरा दृष्टिकोण पाठ्य प्रतिलेखों के संदर्भ में समयरेखा का प्रतिनिधित्व करना है [कैसरेस एट अल।, 2002; ट्रूओंग एट अल., 2016; बर्टुसो, 2012; पावेल एट अल., 2014]। जैसा कि पिछले पैराग्राफ में बताया गया है, वीडियो फ्रेम की वास्तविक संरचना में परिवर्तन करने के लिए समयरेखा के पाठ्य प्रतिनिधित्व को कभी-कभी शब्द स्तर पर हेरफेर किया जा सकता है।

बुद्धिमान वीडियो हेरफेर केवल दो बुद्धिमान वीडियो संपादकों में चर्चा की गई है। पहला काम [कैसारेस एट अल., 2002] में स्मार्ट चयन, क्लिपिंग, क्लिपिंग, पेस्टिंग और री-पिनिंग का उपयोग किया गया। ये सभी क्रियाएं छवि विश्लेषण के साथ फोटो बॉर्डर का उपयोग करके की जाती हैं। दूसरा कार्य केवल वीडियो ट्रांसक्रिप्ट का उपयोग करके वीडियो खंडों का बुद्धिमान चयन या बुद्धिमान कटिंग प्रदान करता है [बर्थौज़ोज़, 2012]। इस कार्य के कई उदाहरणों और नीचे उल्लिखित कार्यों की कमी इस तथ्य के कारण हो सकती है कि सभी स्मार्ट वीडियो संपादन उपकरण अवधारणाओं के प्रमाण हैं, इस प्रकार इन बहुत महत्वपूर्ण लेकिन गैर-आवश्यक सुविधाओं का अभाव है।

परिवर्तन बनाएँ. ट्रांज़िशन बनाने की सरल विधि दो अलग-अलग कार्यों में शामिल है। [डूओंग एट अल., 2016] ने पारगमन कार्यों को गतिशील प्रोग्रामिंग के रूप में तैयार करके सौंदर्यपूर्ण रूप से सुखदायक संक्रमण को स्वचालित करने के लिए एक विधि बनाई, जहां खराब संक्रमण बिंदु, जैसे कि छलांग, को दंडित किया जाता है। [बर्थौज़ोज़, 2012] फ़्रेम के पदानुक्रमित क्लस्टरिंग का उपयोग करके और लेनदेन बिंदुओं के रूप में फ़्रेम के बीच सबसे छोटा रास्ता ढूंढकर छिपे हुए संक्रमण बनाने के लिए एक अलग दृष्टिकोण का उपयोग करता है।

वीडियो रिकॉर्ड करें. [डूओंग एट अल., 2016] ने कैप्चर प्रक्रिया के दौरान वीडियो को पंजीकृत करके ऑडियो एनोटेशन के साथ वीडियो क्लिप को पंजीकृत करने के लिए एक नया दृष्टिकोण प्रस्तुत किया। उनके काम में, रिकॉर्डिंग समीक्षा के दौरान टैग के साथ लॉगिंग के अलावा, ऑडियो द्वारा भी लॉगिंग की जा सकती है।

B. स्वचालन के साथ सहभागिता

इस खंड में, बातचीत के तरीके के साथ-साथ वीडियो अमूर्तता के स्तर [डूओंग और वेंकटेश] को संक्षेप में प्रस्तुत किया जाएगा। हमारे द्वारा जांचे गए अधिकांश स्मार्ट वीडियो संपादन टूल में उपयोग किया जाने वाला इंटरैक्शन का प्राथमिक तरीका कीबोर्ड और माउस के साथ ग्राफिकल यूजर इंटरफेस (जीयूआई) के माध्यम से होता है। एक अपवाद [किमुरा एट अल., 2005] का एक मामला है, जो टकटकी-आधारित बातचीत की जांच करता है। हालाँकि, उपयोगकर्ताओं के पास नियंत्रण की अमूर्तता और सूक्ष्मता का स्तर विभिन्न उपकरणों में भिन्न होता है। वीडियो संपादन टूल में बिना किसी अमूर्तता के, संपादन व्यक्तिगत फ्रेम स्तर पर किया जाना चाहिए, जो बहुत श्रम-गहन है। हालाँकि, कुछ स्मार्ट वीडियो संपादन उपकरण अमूर्तता के कई स्तरों पर वीडियो हेरफेर की पेशकश करते हैं। एकाधिक अमूर्तता प्रदान करने वाले उपकरणों के दो उदाहरण हैं सिल्वर [कैसरेस एट अल., 2002] और क्विककट [डूओंग एट अल., 2016]। चाँदी जो तीन सार प्रस्तुत करता है जो हैं: क्लिप, शॉट और फ़्रेम। क्विककट बोले गए शब्दों और फ्रेम के संदर्भ में अमूर्तता प्रदान करता है।

कुछ वीडियो संपादन उपकरण बहुत उच्च स्तर के अमूर्तन पर काम करते हैं। उदाहरण के लिए, डैमोकट [ची एट अल., 2013] में, वीडियो संपादन उन चरणों के लिए चरणों और मार्करों को अमूर्त करके किया जाता है। इसी तरह, रफकट [लीक एट अल., 2017] में, उपयोगकर्ता संवाद स्क्रिप्ट में संवाद पंक्तियों का उपयोग करके और मुहावरों के रूप में निर्णयों को संपादित करके समयरेखा में हेरफेर कर सकता है। हालाँकि, उच्च स्तर पर हेरफेर, फ्रेम स्तर पर बेहतर समायोजन करने की क्षमता की कीमत पर आता है। हालाँकि, तीन बुद्धिमान वीडियो संपादन टूल में [लीक एट अल., 2017; डूओंग एट अल., 2016; पासरेली, 2019] वीडियो संपादन कार्य को ईडीएल (एडिट डिस्क्रिप्शन लैंग्वेज) के रूप में निर्यात किया जा सकता है, जिसका उपयोग फ्रेम-स्तरीय सुधार करने और वीडियो पर संपादन प्रक्रिया को पूरा करने के लिए अन्य वाणिज्यिक वीडियो संपादन सॉफ्टवेयर के साथ किया जा सकता है।

C. एआई तकनीक का इस्तेमाल किया गया है

वीडियो विभाजन. पहले इंटेलिजेंट वीडियो एडिटिंग टूल पर काम फ्रेम सीमाओं का पता लगाने और प्रत्येक फ्रेम के लिए प्रतिनिधि फ्रेम ढूंढने के लिए छवि विश्लेषण पर निर्भर था [कैसरेस एट अल।, 2002] या छवि विश्लेषण के संयोजन का उपयोग करके, पृष्ठभूमि ज्ञान और मॉडल मिलान [शिपमैन, 2008] में किया गया था। .

फ्रेम पहचान नियम लक्षित वीडियो प्रकारों के लिए हाथ से तैयार किए गए हैं। [कैसारेस एट अल., 2002] में, वाक् पहचान का उपयोग करके हाथ से तैयार किए गए प्रतिलेखों को वीडियो के साथ जोड़ा गया था। विषयगत रूप से सुसंगत इकाइयों में वीडियो व्याख्यानों का विभाजन [पावेल एट अल., 2014] में वीडियो के पाठ्य प्रतिलेख पर विषय विभाजन द्वारा किया जाता है। ऑडियो एनोटेशन के साथ विभाजन का दूसरा रूप [डूओंग एट अल., 2016] में खोजा गया है। यह शब्दार्थ की दृष्टि से प्रासंगिक खंडों में ऑडियो एनोटेशन के माध्यम से गति-आधारित विभाजन और परिशोधन का उपयोग करके काम करता है। गति-आधारित विभाजन वीडियो में निरंतर गति का पता लगाकर किया जाता है, जबकि सिमेंटिक विभाजन वीडियो में क्रियाओं या विषयों से मेल खाता है।

कम्प्यूटेशनल तकनीकों का उपयोग करके इमेजिंग के सिद्धांतों का परिचय [वू एट अल., 2015] और [लीक एट अल., 2017] में पाया जा सकता है। वीडियो क्लिपिंग बिंदुओं, चयनित वीडियो और चयनित ऑडियो अंशों का पता लगाने के लिए डोमेन-विशिष्ट सिद्धांतों को साक्षात्कार के माध्यम से चुना जाता है और अनुकूलन समस्याओं के रूप में प्रस्तुत किया जाता है [वू एट अल।, 2015]। [लीक एट अल., 2017] में, 12

बुनियादी फिल्म संपादन मुहावरों (कूद से बचाव, भावना में वृद्धि, आदि) को फीचर मापदंडों के संदर्भ में दर्शाया जाता है जो संपादन समाधान उत्पन्न करने के लिए एक छिपे हुए मार्कोव मॉडल में इनपुट के रूप में दर्ज होते हैं।
हिडन मार्कोव मॉडल (एचएमएम) मॉडलिंग अनुक्रमों के लिए एक सांख्यिकीय दृष्टिकोण है जिसमें आंतरिक राज्यों की श्रृंखला छिपी हुई है। एचएमएम में उपयोग की जाने वाली सुविधाओं में भाषण-से-पाठ, चेहरे की पहचान और क्लिप से संरचनात्मक जानकारी का उपयोग करके उत्पन्न लेबल शामिल हैं।

IV. संपादक एक स्मार्ट वीडियो संपादन सहायक से क्या चाहते हैं

इस अनुभाग में, हम एक आदर्श एआई वीडियो संपादक के गठन पर (मानव) वीडियो संपादकों की राय जानने के लिए किए गए सर्वेक्षण पर रिपोर्ट करते हैं।

अध्ययन प्रक्रिया
सामान्य जानकारी.
हमारे सर्वेक्षण प्रतिभागियों के बीच औसत वीडियो संपादन अनुभव 9.75 वर्ष था, जिसमें सबसे कम 1 वर्ष से लेकर सबसे अधिक 22 वर्ष तक का अनुभव था। प्रतिभागियों के साथ काम करने वाले वीडियो के प्रकार के संदर्भ में, प्रत्येक प्रतिभागी ने लगभग 3 प्रकार के वीडियो सूचीबद्ध किए। सबसे आम प्रकार के वीडियो विज्ञापन, वृत्तचित्र, प्रस्तुति, खेल, सोशल मीडिया और समाचार वीडियो हैं।
सॉफ्टवेयर प्रोग्राम के संदर्भ में, प्रतिभागियों ने औसतन 5 वीडियो संपादन प्रोग्राम का उपयोग किया। सबसे अधिक उल्लिखित संपादन प्रोग्राम Adobe Premier Pro और DaVinci रेजोल्यूशन हैं। इसके अलावा, Avid, Flimora 9 और VizStory जैसे कम-ज्ञात कार्यक्रमों के साथ-साथ Rev.com और Descript.com जैसी वीडियो सेवाओं का भी उल्लेख किया गया है। जब पूछा गया कि वीडियो संपादन के संदर्भ में उन्होंने किस एआई तकनीक के बारे में सुना है, तो 13 में से 8 उत्तरदाताओं ने एडोब प्रीमियर प्रो सीसी और मैजिस्टो जैसी ब्रांडेड और वाणिज्यिक पेशकशों में एआई तकनीक के साथ उत्तर दिया। शेष 5 उत्तरदाताओं के लिए, उत्तरों में एआई तकनीकों को शामिल किया गया है जो स्वचालित सुधार, शोर में कमी, पृष्ठभूमि हटाना, वीडियो स्थिरीकरण, ऑब्जेक्ट डिटेक्शन/छवि एनोटेशन, सेगमेंटेशन, ज़ूमिंग, गहरी मिथ्याकरण, स्वचालित वीडियो सिग्नल, चेहरे की पहचान और पाठ के लिए भाषण हैं।

उत्तम एआई संपादक। इस प्रश्न का उत्तर "आप कौन सा आदर्श AI वीडियो संपादन उपकरण चाहेंगे?" उत्तर एक दूसरे से बहुत भिन्न होते हैं। हालाँकि, हमने प्रतिक्रियाओं में पाँच विषयों की पहचान की। वे वीडियो संपादन कार्यों के लिए एक उपकरण के रूप में एआई, प्रोजेक्ट के लिए एक उपकरण के रूप में एआई हैं प्रबंधन कार्य, सौंदर्य गुणवत्ता में सुधार के लिए स्वचालित उपकरण, मानव एआई समस्याएं और सामग्री खोज के लिए एआई।
वीडियो संपादन कार्यों के लिए एक उपकरण के रूप में AI सबसे बड़ी श्रेणी है जिसमें अधिकांश प्रतिक्रियाएँ आती हैं। इसमें ऐसे कीवर्ड शामिल हैं जो शॉट रिकग्निशन, वीडियो कंपोज़िटिंग, डायलॉग लाइनों के आधार पर खराब वीडियो को फ़िल्टर करना, गाने और उपशीर्षक को सिंक करना, अनुवाद और भाषा की समझ हैं। परियोजना प्रबंधन एआई विषय में वीडियो मेटाडेटा निर्माण, डेटा प्रबंधन और अपनाने जैसे शब्द शामिल हैं। अगला विषय सौंदर्य गुणवत्ता में सुधार के लिए एआई है, जिसमें स्वचालित रंग ग्रेडिंग और स्वचालित ऑडियो समकरण शामिल है। इसके अलावा, एआई के संबंध में मानव-एआई संबंधी चिंताएं भी हैं, जैसे नियंत्रण और स्वचालन का संतुलन, उपयोगकर्ता-केंद्रित एआई और वैयक्तिकरण। अंत में, सामग्री खोज एआई विषय की शर्तों में टाइमलाइन पर मौजूदा सामग्री के आधार पर स्टॉक वीडियो और स्टॉक संगीत वीडियो का सुझाव देना शामिल है।
जब पूछा गया कि एआई वीडियो संपादक के साथ कैसे बातचीत की जाए, तो अधिकांश प्रतिक्रियाओं में उल्लेख किया गया कि वे आवाज के माध्यम से बातचीत करना चाहेंगे, इसके बाद वे लोग थे जो कीबोर्ड और माउस जीयूआई के माध्यम से बातचीत करना चाहते हैं। बातचीत के इन दो मुख्य तरीकों के अलावा, कई अन्य मोड जैसे स्पर्श इंटरफ़ेस, इशारे, मस्तिष्क कंप्यूटर इंटरफ़ेस का कई बार उल्लेख किया गया है। कुछ उत्तरों में परियोजना की स्थिति के आधार पर संचार के लिए आवश्यक प्रासंगिक आदेशों के साथ-साथ एक एआई ऑफ बटन का भी उल्लेख किया गया है जो स्वचालन को आसानी से बंद करने की अनुमति देता है।
संपूर्ण एआई संपादक के विषय पर अंतिम प्रश्न उन अमूर्तताओं (मानव) के स्तर के बारे में पूछता है जिनके साथ संपादक काम करना चाहेंगे। अधिकांश उत्तरदाताओं ने कहा कि वे एक आदर्श एआई वीडियो संपादन टूल के अपने दृष्टिकोण में कीफ़्रेम स्तर पर वीडियो में हेरफेर करना चाहेंगे। अमूर्तता का दूसरा और तीसरा सबसे लोकप्रिय स्तर क्लिप और फ़्रेम हैं। अन्य प्रकार के अमूर्तन जिनका एक बार उल्लेख किया गया है वे अनुक्रम, इतिहास और फ्रेम हैं। दो उत्तरदाताओं ने उल्लेख किया कि वे एक लचीले प्रकार का अमूर्तीकरण चाहेंगे जहां वे अंतःक्रिया के अंतर्निहित सिद्धांतों के लिए अमूर्तन के स्तर को समायोजित कर सकें। एआई और वर्कफ़्लो। अध्ययन का दूसरा भाग निम्नलिखित विषयों की पड़ताल करता है: वीडियो संपादन वर्कफ़्लो में कार्य जिन्हें प्रतिभागी स्वचालित करना चाहते हैं, और इन कार्यों के लिए स्वायत्तता के स्तर और बातचीत के तरीकों से संबंधित मुद्दे। वर्कफ़्लो अनुभाग में प्रश्नों के उत्तर में चार थीम वाले कार्य शामिल हैं: वीडियो संपादन कार्य, सौंदर्य संवर्धन, वीडियो पूर्व-संपादन कार्य और सुझाव कार्य।

वीडियो संपादन कार्यों का वर्णन करने के लिए उपयोग किए जाने वाले सबसे लोकप्रिय कीवर्ड विभाजन और उपशीर्षक थे, जिनका उल्लेख तीन प्रतिक्रियाओं में किया गया था। वीडियो संपादन कार्यों के लिए दूसरे सबसे लोकप्रिय कीवर्ड वीडियो विभाजन और खराब फ्रेम फ़िल्टरिंग हैं (जिनका दो बार उल्लेख किया गया था)। इसके अलावा, निम्नलिखित वीडियो संपादन कार्यों का एक बार उल्लेख किया गया है: सामग्री विश्लेषण, वीडियो संग्रह, चेहरे का पता लगाना, फसल प्लेसमेंट, संक्रमण फ्रेम चयन, ऑडियो-टू-वीडियो सिंक्रनाइज़ेशन, और दो-चैनल ऑडियो और वीडियो चयन।

उत्तरों में अगला अक्सर उल्लिखित विषय कार्य जिसे हम स्वचालित करना चाहते हैं वह सौंदर्य गुणवत्ता में सुधार है। इनके लिए सबसे लोकप्रिय शब्द रंग सुधार और ऑडियो इक्वलाइज़ेशन हैं। इसके अलावा, दृश्य सुधार, पृष्ठभूमि हटाना और हकलाना हटाना जैसे कार्यों का एक-एक बार उल्लेख किया गया था। पूर्व-संपादन कार्यों के लिए, वीडियो रिकॉर्डिंग का दो बार और स्वचालित टाइमकोड निर्माण का एक बार उल्लेख किया गया है। विचारोत्तेजक कार्यों के संदर्भ में, उत्तरों में संपादन शैलियों, सामान्य सहायता और संगीत सुझावों के साथ एक क्लिप सुझाव शामिल है।

अंतिम प्रश्न यह है कि संपादकों को उनके द्वारा उपयोग किए गए टूल की तुलना में संपादक के एआई होने की कितनी अच्छी उम्मीद है। इस प्रश्न में, चार उत्तरदाताओं ने कहा कि वे चाहते थे कि यह बहुत समान या परिचित हो। दो प्रतिक्रियाओं में कहा गया कि वे चाहते थे कि समानता का बुनियादी स्तर हो। दो उत्तरों में पाया गया कीवर्ड "उपयोग में आसान" एक और शब्द है जिसका अर्थ "परिचित" के समान हो सकता है। एक प्रतिवादी ने एआई वीडियो संपादन टूल को मौजूदा टूल में एकीकृत करने के लिए प्लग-इन दृष्टिकोण का उल्लेख किया है। केवल दो उत्तरदाताओं ने कहा कि उन्हें उम्मीद है कि भविष्य में एआई वीडियो संपादक बहुत अलग होंगे या बिल्कुल भी नहीं होंगे।

V. एआई वीडियो संपादन टूल के लिए चुनौतियाँ

वीडियो संपादन कार्य साहित्य में पहचाने गए और हमारे अध्ययन के परिणामों में उल्लिखित वीडियो संपादन कार्यों के बीच काफी समानता है। यहां हम वीडियो संपादन वर्कफ़्लो के अनछुए हिस्सों पर ध्यान केंद्रित करते हैं। पहला है विभिन्न ट्रैक से ऑडियो और वीडियो को सिंक करना। इस कार्य की स्वचालित वीडियो मिश्रण पीढ़ी के एक अलग लेकिन संबंधित संदर्भ में पहले ही जांच की जा चुकी है [वू एट अल., 2015; श्रेष्ठ एट अल., 2010]। वीडियो में खराब फ्रेम या खराब सेगमेंट को फ़िल्टर करने का अध्ययन नहीं किया गया है, लेकिन यह समझने के लिए सूचित शोध के साथ किया जाना चाहिए कि वीडियो संपादकों का क्या मतलब था जब उन्होंने कहा कि वे खराब फ्रेम थे। अंततः, हमारे सामने भाषा जैसे मुद्दे हैं वीडियो संपादित करते समय स्वचालित अनुवाद, उपशीर्षक और भाषा की समझ। उपशीर्षक और स्वचालित अनुवाद प्रणाली को मशीन स्वचालन का उपयोग करके स्वचालित किया जा सकता है। हालाँकि, वीडियो संपादन के संदर्भ में भाषा को समझने के लिए प्राकृतिक भाषा प्रसंस्करण में प्रगति और वीडियो संपादन के संदर्भ में वीडियो संपादन शब्दों और भाषा के उपयोग की समझ दोनों की आवश्यकता होती है।

वीडियो लॉगिंग और मेटाडेटा निर्माण वीडियो लॉगिंग एक वीडियो रिकॉर्डिंग देखना और टाइमकोड का उपयोग करके उसकी सामग्री को टैग करना है। इसे बुद्धिमान वीडियो संपादन साहित्य और हमारे शोध दोनों में उन कार्यों में से एक के रूप में पहचाना गया है जिन्हें उपयोगकर्ता स्वचालित करना चाहेंगे। साहित्य में वर्तमान वीडियो रिकॉर्डिंग तकनीक विशिष्ट अनुप्रयोगों तक बहुत सीमित हैं; अर्थात् डेमो वीडियो और संवाद आधारित वीडियो। वीडियो को टेक्स्ट में बदलने और टेक्स्ट प्रोसेसिंग तकनीकों को लागू करने के लिए वाक् पहचान के वीडियो रिकॉर्डिंग और मेटाडेटा निर्माण में कई संभावित उपयोग के मामले हैं। एआई का एक और दिलचस्प क्षेत्र जिस पर ध्यान देना चाहिए वह है वीडियो तर्क और दृश्य और भाषाई इनपुट दोनों से पैटर्न के संयोजन को समझना [वू एट अल।, 2016]।

वीडियो संपादन के लिए ध्वनि-आधारित इंटरैक्शन बातचीत का यह सबसे आम तरीका है जिसके बारे में लोग कहते हैं कि वे एआई वीडियो संपादन टूल के साथ संवाद करने के लिए इसका उपयोग करना चाहेंगे। वीडियो संपादन टूल में वॉयस यूजर इंटरफेस की क्षमता का पता नहीं लगाया गया है। [चांग एट अल., 2019] ने निर्देशात्मक वीडियो को नेविगेट करने के लिए आवाज-आधारित इंटरैक्शन के डिजाइन स्थान की खोज प्रस्तुत की। चूंकि वीडियो संपादन में वॉयस इंटरेक्शन पूरी तरह से अज्ञात क्षेत्र है, इसलिए शुरुआती बिंदु डिजाइन अनुसंधान होना चाहिए। एक अन्य संभावना इंटरैक्शन के बजाय संदर्भ-मुक्त एकल वॉयस कमांड का अध्ययन करना है। ऐसे कार्य जिन्हें उपयोगकर्ता पूरी तरह से स्वचालित करना चाहेगा, जैसे कि सौंदर्य गुणवत्ता में वृद्धि, फ़ाइल प्रबंधन या पूर्व-संपादन कार्य, वॉयस कमांड डिज़ाइन के लिए आदर्श होने चाहिए।

इस संदर्भ में वैयक्तिकरण एक स्मार्ट वीडियो संपादन टूल की क्षमता है जो सॉफ्टवेयर में संसाधित वीडियो, उत्पादित वीडियो और उपयोग पैटर्न से सीखकर उपयोगकर्ता के अनुकूल होता है। यह विषय साहित्य से अनुपस्थित है लेकिन हमारे अध्ययन में वीडियो संपादन और वैयक्तिकरण के संदर्भ को समझने के रूप में पाया गया है। वीडियो संपादन नियमों की नियम-आधारित शिक्षा पर [मात्सुओ एट अल.] में चर्चा की गई है।

यह ध्यान रखना महत्वपूर्ण है कि धारा 3 में चर्चा किए गए स्मार्ट वीडियो संपादन उपकरण एक विशिष्ट प्रकार के वीडियो के संपादन को हल करने के लिए डिज़ाइन किए गए हैं: वास्तव में, उनमें से आठ केवल एक प्रकार के वीडियो को लक्षित करते हैं। हालाँकि, वीडियो संपादकों के हमारे सर्वेक्षण में प्रत्येक द्वारा औसतन तीन प्रकार के वीडियो संसाधित किए गए हैं। चूंकि अधिकांश स्मार्ट वीडियो संपादन उपकरण एक प्रकार के वीडियो के लिए बनाए गए हैं, इसलिए परिणामों में उल्लिखित तकनीकों की सामान्य प्रयोज्यता के बारे में चिंताएं हैं।

स्मार्ट वीडियो एडिटिंग टूल साहित्य में वर्णित वीडियो संपादन कार्य मूल वीडियो संपादन कार्य के दौरान होने वाली बातचीत पर केंद्रित हैं। हालाँकि, अनुसंधान प्रतिभागियों को फ़ाइल/मीडिया संगठन, सौंदर्य गुणवत्ता में सुधार, पूर्व-संपादन कार्यों और सामग्री सुझाव जैसे अतिरिक्त कार्यों के लिए स्वचालन की आवश्यकता होती है। वीडियो संपादन सामग्री सुझाव किसी मौजूदा वीडियो या कहानी में जोड़ने के लिए अच्छे वीडियो या संगीत खंड का सुझाव देता है।

सर्वेक्षण के परिणामों में, प्रतिभागियों द्वारा वांछित बातचीत का सबसे आम तरीका आवाज बातचीत थी। हालाँकि, एक बुद्धिमान वीडियो संपादक पर केवल एक काम में आवाज शामिल है [ची एट अल।, 2013], जहां वीडियो को टैग करने के लिए आवाज एनोटेशन का उपयोग किया जाता है। वॉइस इंटरेक्शन की इस लोकप्रियता को मोबाइल फोन और स्मार्ट होम स्पीकर में एआई-आधारित वॉयस सहायता की लोकप्रियता के साथ-साथ विज्ञान कथा में एक आवाज के रूप में एआई के प्रतिनिधित्व के लिए जिम्मेदार ठहराया जा सकता है।

स्मार्ट वीडियो संपादन टूल में उपयोग की जाने वाली AI तकनीक अनुमान-आधारित प्रणालियाँ हैं। अन्य दृष्टिकोण जैसे तंत्रिका नेटवर्क और मशीन लर्निंग-आधारित दृष्टिकोण का कम अध्ययन किया जाता है। [डोव एट अल., 2017] के अध्ययन ने निष्कर्ष निकाला कि उपयोगकर्ता अनुभव बनाने के लिए मशीन लर्निंग एक कठिन डिज़ाइन सामग्री है, क्योंकि मशीन लर्निंग-आधारित प्रोटोटाइप बनाना मुश्किल है और इसे निष्पादित करने के लिए मशीन लर्निंग सहयोगियों के प्रशिक्षण की आवश्यकता होती है।

VI. निष्कर्ष

इस पेपर में, हमने बुद्धिमान वीडियो संपादन टूल को परिभाषित किया है और (बुद्धिमान) वीडियो संपादन, उपयोगकर्ता इंटरैक्शन और एआई तकनीक पर मौजूदा साहित्य का अवलोकन प्रस्तुत किया है।

हमने वीडियो संपादकों से उनके वीडियो संपादन वर्कफ़्लो में स्वचालन की आवश्यकताओं के बारे में भी सर्वेक्षण किया। अध्ययन के इस क्षेत्र में शोध करने के लिए ज्ञान की आवश्यकता होती है वीडियो संपादन, मानव कंप्यूटर इंटरैक्शन, और एआई या मशीन लर्निंग। इन तीन अलग-अलग विशेषज्ञताओं की आवश्यकता वाले बुद्धिमान वीडियो संपादन टूल एक कारण है कि व्यक्तिगत रूप से प्रत्येक क्षेत्र की तुलना में उन पर साहित्य बहुत सीमित है।

इस अध्ययन में, वीडियो संपादन कार्यों के क्षेत्र में साहित्य और शोध निष्कर्षों के बीच बड़ी मात्रा में क्रॉसओवर है। हालाँकि, वीडियो रिकॉर्डिंग या वीडियो संपादन परियोजनाओं का आयोजन, सौंदर्य गुणवत्ता सुधार और सामग्री सुझाव जैसे क्षेत्र हैं जिन्हें अनुसंधान आवश्यकताओं को पूरा करने के लिए और अधिक तलाशने की आवश्यकता है। हमारा निष्कर्ष यह है कि मशीन लर्निंग समुदाय की अधिक भागीदारी के साथ, आदर्श एआई संपादक प्राप्त किया जा सकता है। भविष्य के कार्यों में, हम इस लक्ष्य में योगदान देने का इरादा रखते हैं।

संदर्भ

[1] एफ़. बर्थौज़ोज़। साक्षात्कार वीडियो में कट्स और ट्रांज़िशन लगाने के लिए उपकरण। पृष्ठ 8, 2012. एस. बीकोनी। शब्दार्थ-जागरूक स्वचालित वीडियो संपादन। मल्टीमीडिया पर 12वें वार्षिक एसीएम अंतर्राष्ट्रीय सम्मेलन की कार्यवाही में - मल्टीमीडिया '04, पृष्ठ 971, 2004।

[2] एस. बटलर और ए. पार्क्स। स्वचालित बुद्धिमान वीडियो संपादन के लिए फिल्म अनुक्रम निर्माण रणनीतियाँ। एप्लाइड आर्टिफिशियल इंटेलिजेंस, 11(4):367-388, जून 1997।

[3] जे. कैसरेस, एसी लॉन्ग, बीए मायर्स, आर. भट्नागर, एसएम स्टीवंस, एल. डैबिश, डी. योकम, और ए. कार्बेट। मेटाडेटा का उपयोग करके वीडियो संपादन को सरल बनाना। पृष्ठ 10, 2002।

[4] आरज़ी कैटेलन, सी. टेक्सेरा, आर. गौलार्ट, और एमडीजीसी पिमेंटेल। सर्वव्यापी इंटरैक्टिव वीडियो संपादन की दिशा में एक आदेश के रूप में देखें और टिप्पणी करें। मल्टीमीडिया कम्प्यूटिंग, संचार और अनुप्रयोगों पर एसीएम लेनदेन, 4 (4):1-24, अक्टूबर 2008.

[5] एम. चांग, ए. इंग, ओ. वांग, एम. अग्रवाला, और जे. किम। कैसे-करें वीडियो के लिए ध्वनि आधारित नेविगेशन कैसे डिज़ाइन करें। कम्प्यूटिंग सिस्टम में मानव कारकों पर 2019 सीएचआई सम्मेलन की कार्यवाही में, पृष्ठ 1-11, मई 2019।

[6] पी.-वाई. ची, जे. लियू, जे. लिंडर, एम. डोनचेवा, डब्ल्यू. ली, और बी. हार्टमैन। डेमोकट: भौतिक प्रदर्शनों के लिए संक्षिप्त निर्देशात्मक वीडियो तैयार करना। पृष्ठ 10, 2013. वी. सिस्को। सिस्को विज़ुअल नेटवर्किंग इंडेक्स: पूर्वानुमान और रुझान, 2017-2022। श्वेत पत्र, 1:1, 2018।

[7] जी. डोव, के. हल्सकोव, जे. फ़ोर्लिज़ी, और जे. ज़िम्मरमैन। यूएक्स डिज़ाइन इनोवेशन: डिज़ाइन सामग्री के रूप में मशीन लर्निंग के साथ काम करने की चुनौतियाँ। पृष्ठ 278-288, 2017। जे. ईसेनस्टीन और आर. बरज़िले। बायेसियन अप्रशिक्षित विषय विभाजन। प्राकृतिक भाषा प्रसंस्करण में अनुभवजन्य तरीकों पर सम्मेलन की कार्यवाही में - ईएमएनएलपी '08, पृष्ठ 334, 2008।

[8] एक्स.-एस. हुआ, एल. लू, और एच.-जे. झांग. अनुकूलन-आधारित स्वचालित होम वीडियो संपादन प्रणाली। वीडियो प्रौद्योगिकी के लिए सर्किट और सिस्टम पर आईईईई लेनदेन, 14(5):572-583, मई 2004।

[9] टी. किमरा, के. समिया, और एच. तनाका। उपयोगकर्ता की निगाहों का उपयोग करते हुए एक वीडियो संपादन सहायता प्रणाली। में

पैक्रिम। 2005 संचार, कंप्यूटर और सिग्नल प्रोसेसिंग पर आईईईई पैसिफिक रिम सम्मेलन, 2005, पृष्ठ 149-152, 2005।

[10] एम. लीक, ए. डेविस, ए. इंग, और एम. अग्रवाला। संवाद-संचालित दृश्यों के लिए कम्प्यूटेशनल वीडियो संपादन। ग्राफिक्स पर एसीएम लेनदेन, 36(4):1-14, जुलाई 2017।

[11] एसी लॉन्ग, बीए मायर्स, जे. कैसरेस, एसएम स्टीवंस, और ए. कॉर्बेट। लेंस और सिमेंटिक ज़ूमिंग का उपयोग करके वीडियो संपादन। पृष्ठ 10, 2004।

[12] यू. मात्सुओ, एम. अमानो, और के. उएहारा। वीडियो स्टीम में माइनिंग वीडियो संपादन नियम। पेज 4, बीए मायर्स, जेपी कैसरेस, एस. स्टीवंस, एल. डैबिश, डी. योकम, और ए. कॉर्बेट। डिजिटल वीडियो लाइब्रेरी के लिए एक बहु-दृश्य बढ़ाधिमान संपादक। डिजिटल पुस्तकालयों पर 3 पहले एसीएम/आईईईईसीएस संयुक्त सम्मेलन की कार्यवाही में - जेसीडीएल '01, पृष्ठ 106-115, 2001।

[13] एफ. नैक और ए. पार्क्स। स्वचालित वीडियो संपादन में वीडियो सिमेंटिक्स और थीम प्रतिनिधित्व का अनप्रयोग। एचजे झांग, पी. एंग्रेन, और डी. पेटकोविच, संपादकों में, मल्टीमीडिया सिस्टम में वीडियो डेटा का प्रतिनिधित्व और पुनर्प्राप्ति, पृष्ठ 57-83। 1997.

[14] जेए ओकन, एस. ज़्वर्मन, के. रैफ़र्टी, और एस. स्क्वॉयर, संपादक। दृश्य प्रभावों की वीईएस हैंडबुक: उद्योग मानक वीएफएक्स प्रथाएं और प्रक्रियाएं। 2015.

[15] पी. पसारेली. ऑटोएडिट फास्ट टेक्स्ट आधारित वीडियो संपादन, 2019।

[16] ए. पावेल, सी. रीड, बी. हार्टमैन, और एम. अग्रवाला। वीडियो डाइजेस्ट: सूचनात्मक व्याख्यान वीडियो के लिए एक ब्राउज़ करने योग्य, स्किम करने योग्य प्रारूप। यूज़र इंटरफ़ेस सॉफ्टवेयर और प्रौद्योगिकी पर 27वीं वार्षिक एसीएम संगोष्ठी की कार्यवाही में, पृष्ठ 573-582, अक्टूबर 2014।

[17] एम. रैडट, एम. इवांस, के. टू, टी. नूनी, और जी. फिलिप्सन। कितना . अच्छा है, यहाँ काफ़ी है? लाइव इवेंट के एआई-संपादित वीडियो कवरेज की व्यक्तिपरक गुणवत्ता का मूल्यांकन करने की चुनौती। इंटेलिजेंट सिनेमैटोग्राफी और एडिटिंग पर कार्यशाला, पृष्ठ 8 पृष्ठ, 2020।

[18] कलाकृति का आकार: 8 पृष्ठ आईएसबीएन: 9783038681274 प्रकाशक: द यूरोग्राफिक्स एसोसिएशन संस्करण संख्या: 017-024।

[19] बी. अनुसंधान एवं विकास. एआई और आर्काइव - मेड बाय मशीन का निर्माण, 2018। एफ शिपमैन। हाइपरवीडियो का लेखन, देखना और निर्माण करना: हाइपर-हिचकॉक का एक अवलोकन। 5(2):19, 2008.

[20] पी. श्रेष्ठ, पीएच डी विथ, एच. वेदा, एम. बारबेरी, और ईएच आर्ट्स। एकाधिक कैमरा कॉन्सर्ट रिकॉर्डिंग से स्वचालित मैशअप पीढ़ी। मल्टीमीडिया पर अंतरराष्ट्रीय सम्मेलन की कार्यवाही में - एमएम '10, पृष्ठ 541, 2010।

[21] सी. तास्किर, जेड. पिज्लो, ए. अमीर, डी. पोल्सेलियन, और ई. डेलप। भाषण प्रतिलेखों का उपयोग करके स्वचालित वीडियो प्रोग्राम साराशीकरण। मल्टीमीडिया पर आईईईई लेनदेन, 8(4):775-791, अगस्त 2006।

[22] ए. डूओंग, एफ. बर्थौज़ोज़, डब्ल्यू. ली, और एम. अग्रवाला। क्विकक्यूट: वर्णित वीडियो को संपादित करने के लिए एक इंटरैक्टिव टूल। यूजर इंटरफ़ेस सॉफ्टवेयर और प्रौद्योगिकी पर 29वीं वार्षिक संगोष्ठी की कार्यवाही में, पृष्ठ 497-507, अक्टूबर 2016।

[23] टी. डूओंग और एस. वेंकटेश। वीडियो अमर्तन: एक व्यवस्थित समीक्षा और वर्गीकरण। 3(1):37. एच.-वाई. वू, टी. सैंटारा, एम. लीस, आर. वर्गीस, और ए. झाली। स्वचालित वीडियो संपादन के लिए संयुक्त ध्यान। इंटरएक्टिव मीडिया, एक्सपीरियंस पर एसीएम अंतर्राष्ट्रीय सम्मेलन में, पृष्ठ 55-64, जून 2020।

[24] य. वू, टी. मेई, वाई.-क्यू. जू, एन. यू, और एस. ली. मवीअप: स्वचालित मोबाइल वीडियो मैशअप. वीडियो प्रौद्योगिकी के लिए सर्किट और सिस्टम पर आईईईई लेनदेन, 25(12):1941-1954, दिसंबर 2015।

[25] जेडू. वू, टी. याओ, वाई. फ़ू, और वाई.-जी. जियांग. वीडियो वर्गीकरण और कैप्शनिंग के लिए गहन शिक्षण। arXiv प्रीप्रिंट arXiv:1609.06782, 2016. सी. जू, एल. ली, एफ. यांग, पी..वांग, टी. वांग, और वाई. झांग। स्वचालित होम वीडियो संपादन: एक मल्टी-कोर समाधान। पेज 2।

प्रमाणन अभिकर्ता देखें