Visualization and Visual Analytics Approaches for Image and Video Datasets: A Survey

SHEHZAD AFZAL, SOHAIB GHANI, and MOHAMAD MAZEN HITTAWE, King Abdullah University of Science and Technology
SHEIKH FAISAL RASHID, German Research Center for Artificial Intelligence (DFKI)
OMAR M. KNIO, MARKUS HADWIGER, and IBRAHIM HOTEIT, King Abdullah University of Science and Technology

Image and video data analysis has become an increasingly important research area with applications in different domains such as security surveillance, healthcare, augmented and virtual reality, video and image editing, activity analysis and recognition, synthetic content generation, distance education, telepresence, remote sensing, sports analytics, art, non-photorealistic rendering, search engines, and social media. Recent advances in Artificial Intelligence (AI) and particularly deep learning have sparked new research challenges and led to significant advancements, especially in image and video analysis. These advancements have also resulted in significant research and development in other areas such as visualization and visual analytics, and have created new opportunities for future lines of research. In this survey article, we present the current state of the art at the intersection of visualization and visual analytics, and image and video data analysis. We categorize the visualization articles included in our survey based on different taxonomies used in visualization and visual analytics research. We review these articles in terms of task requirements, tools, datasets, and application areas. We also discuss insights based on our survey results, trends and patterns, the current focus of visualization research, and opportunities for future research.

CCS Concepts: • **Human-centered computing** → **Visualization application domains**; • **Computing methodologies** → *Computer vision;*

Additional Key Words and Phrases: Survey, image and video datasets, visual analytics, computer vision

5

1 INTRODUCTION

Image and video analysis research has significantly advanced in recent years due to massive growth in **artificial intelligence** (**AI**), especially in the field of deep learning. Deep learning-based frameworks have revolutionized the field of computer vision [74]. Considering the prevalence of image and video data in our daily lives, and the availability of high-performance computing infrastructure, intense research interest surrounds the field of computer vision and other related areas. Consequently, significant visualization research is being conducted in problem areas encountered when working with image and video datasets, such as medical datasets [39, 67, 141, 173], sports analytics [7, 22, 112, 114, 117, 118, 154, 169, 170], video and image editing and processing [14, 66, 72, 159, 171], video surveillance, activity and scene recognition, motion flow analysis, content analysis [3, 21, 24, 41, 60, 89, 95, 122, 160, 167], video and image search [11, 75], annotations, content summarization and synthesis [6, 42, 70, 104, 109], social media analytics [59], non-photorealistic rendering, art, video painting [45, 52, 61, 178, 178], and virtual and augmented reality [20, 106, 121, 151].

Advances in AI have significantly improved the state of the art in computer vision, but have simultaneously posed new challenges for visualization and visual analytics researchers to develop new visualization techniques and frameworks to address these challenges. This mandates a systematic review of the current state of the art at the intersection of image and video analysis and visualization research to identify the gaps that exist between the requirements of the two domains, and explore opportunities for future research.

For this purpose, we have gathered a multidisciplinary team of coauthors with visualization, machine learning, video processing, and computer vision backgrounds. To the best of our knowledge, no other recent relevant surveys have been published on the visualization and visual analytics approaches used for image and video datasets.

In the initial phase, we reviewed articles related to image and video analysis published in visualization conferences and journals. We categorized these research articles based on various taxonomies that focus on different aspects of visualization research, such as interaction, visualization, machine learning methods, data scale, and application areas. Categorizing and labeling the visualization research facilitated the study of current trends and patterns and helped identify gaps between the task requirements of the computer vision domain and the current focus of visualization research.

The major contributions of our survey article are

- — Categorization of articles in the visualization domain related to image and video datasets based on standard taxonomies. Extracted task requirements from these surveyed articles and grouped them into different application areas
- — Identification of tools, libraries, and datasets used in visualization research focused on image and video datasets and grouping them based on application areas
- — Identification of gaps, challenges, and opportunities for future collaborative research at the intersection of the computer vision and visualization domains

2 SURVEY DESIGN

In this section, we present details of the survey organization, methodology, and motivation for the work.

2.1 Survey Organization

In this survey report, we first provide the scope of the work. In Section 2.3, we introduce the topic and motivation for conducting this survey. We highlight the advances in AI and deep learning

that have triggered rapid advancements in computer vision, and how these advances necessitate this current survey on visualization and the visual analytics domain. We also discuss the relevant existing surveys, and how our state-of-the-art report is different from these existing surveys. We then describe our survey methodology (Section 2.4), including details of the keywords used to search for articles, the selection criteria for conferences and journals, the ranking of articles, the compilation of a final set of articles, and the scope of our survey.

In Section 3, we review the relevant articles in the visualization domain, and categorize those articles based on different taxonomies. These taxonomies include algorithms and techniques (Section 3.1), visualization techniques (Section 3.2), and application areas (Section 3.4). We also provide tables (Table 1a, 1b, and 1c) showing how surveyed articles in the visualization domain are labeled according to these taxonomies. We also provide additional insights into the high-level representative task requirements, tools, libraries, and datasets grouped based on application areas.

In Section 4, we provide details of discussions with computer vision domain experts concerning their domain specific visualization requirements for image and video analysis. In the final sections, we present the current trends and patterns, gaps in the current research, major challenges, and a discussion on future research directions. Finally, we outline some limitations of our work.

2.2 Scope

In this survey report, we aim at identifying overlaps and gaps in computer vision and visualization research focused on image and video datasets, and identify potential areas of collaborative research. We catalogue the current research occurring at the intersection of these domains in terms of techniques and algorithms, tools and libraries, datasets, application areas, and task requirements. The results and findings are summarized in the form of tables and discussions that aim at facilitating researchers working in relevant research areas in terms of guidance and identifying potential future research areas.

The tables presented throughout the article serve as a reference to highlight current trends, identify gaps in research, tools and techniques in use, and application areas that are less actively explored at the intersection of computer vision and visualization. These tables are also a reference guide for researchers to different visualization techniques for various computer vision problems.

This work can assist visualization researchers working on image and video datasets to obtain insights into the nature of task requirements of different application areas, existing tools and techniques, and research gaps. Discussions with computer vision domain experts provide additional insights based on their own experience.

This survey report can also aid computer vision researchers in understanding which type of interactive visualization solutions have already been designed for various computer vision problems and identify opportunities for collaborative research efforts.

2.3 Motivation

Visualization and visual analytics tools are used in image and video analysis in different application areas such as video surveillance, activity recognition, human motion analysis and recognition, scene interpretation, video and image editing, sports analytics, medical imaging analysis, and social media analytics. Due to advancements in the AI domain and the availability of high-performance computing infrastructure, computer vision techniques and algorithms have significantly evolved in the last few years. This, in turn, presents significant challenges, relating to different aspects of visualization design, for visualization and visual analytics researchers to address different task requirements for analyzing an image and video datasets. To identify the gaps between visualization research and the current focus of computer vision research across different dimensions, visualization literature should be explored using different visualization taxonomies. Existing

visual analytics models and frameworks [125, 126] must be evaluated in the context of task requirements for analyzing an image and video datasets in the wake of recent rapid growth in computer vision. Clearly, there is a need to evaluate the current state-of-the-art in visualization research.

There are some relevant surveys available in the literature [18, 19, 71, 155, 168, 179]. ML4VIS [155] focused on understanding the current practices of employing machine learning techniques to solve visualization problems. This survey explores the relevant research to determine which visualization processes can benefit from machine learning. Moreover, how machine learning techniques can be employed for visualization problems. AI4VIS [168] explored the vision of considering visualization as a new data format (visualization data) and reviewed recent advances in applying AI techniques to this data format. The ML4VIS and AI4VIS surveys probed different sets of questions with a different focus than our survey. Yuan et al. [179] categorized visual analytics techniques based on their usage before, during, and after model building in machine learning applications. Our survey, instead, focuses specifically on image and video datasets. Borgo et al. [18, 19] conducted a survey on video-based graphics and video visualization; however, this survey was conducted almost ten years ago and did not cover recent advancements in the computer vision and visualization domains. Kyprianidis et al. [71] conducted a survey in 2012 focusing on **non-photorealistic rendering** (**NPR**) techniques to transform images and videos into artistically styled renderings. Dudley and Kristensson [38] presented a review of user interface research to design effective interfaces for interactive machine learning algorithms. There are some other survey articles [28, 54, 85, 98, 132, 181] related to machine learning models analytics and visualization, but they have primarily not considered image and video datasets.

Some other related surveys also exist, but they focus on computer vision and are not relevant to visualization. Khurana and Kushwaha [65] surveyed literature related to human activity recognition in surveillance videos. Shih [137] reviewed research focused on content-aware video analysis for sports videos rather than from a spatiotemporal viewpoint. Wang and Ji [156] conducted a survey on effective video content analysis based on direct and implicit approaches. Other notable surveys, that are not directly relevant to our survey, include [17, 36, 79, 84, 99, 102, 111, 146, 150, 176]. In this work, we provide detailed coverage of visualization and visual analytics approaches for image and video datasets. We also identify the major challenges and future research directions.

2.4 Methodology

Figure 1 shows the flowchart of the methodology we employed in this work. We initiated our survey by searching for articles from major visualization and visual analytics conferences and journals with the keywords "image", "images", "video", or "videos" in their titles or abstracts. Our search included IEEE **Transactions on Visualization and Computer Graphics** (**TVCGs**), **EG & VGTC Conference on Visualization** (**EuroVis**), IEEE **Visual Analytics Science & Technology** (**VAST**), **Computer Graphics Forum** (**CGF**), IEEE **Scientific Visualization** (**SciVis**), IEEE **Symposium on Information Visualization** (**InfoVis**), **Computer Graphics Applications** (**CG&A**), IEEE **Large Scale Data Analysis & Visualization** (**LDAV**), and IEEE **Pacific Visualization Symposium** (**PacificVis**).

We collected around 150 articles and studied their titles and abstracts. Some non-relevant articles were filtered out after reading their full text. In total, we compiled 107 relevant articles. Figure 2 shows the temporal distribution of the surveyed articles according to the publishing year. It is clear from the distribution that image and video data-related problems are becoming more common due to advancements in AI and computer vision. We labeled the algorithms, visualization techniques, application areas, datasets, and so on, used in these articles, as shown in (Table 1a, 1b, and 1c). We present details of this coding along with a detailed discussion of the trends, lessons learned, and future research challenges.
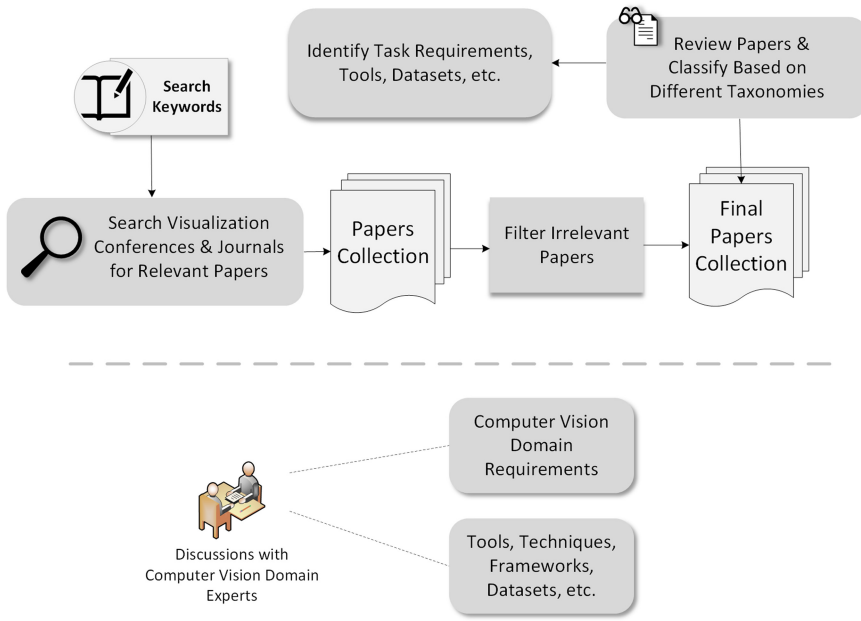
Fig. 1. Methodology of the survey.
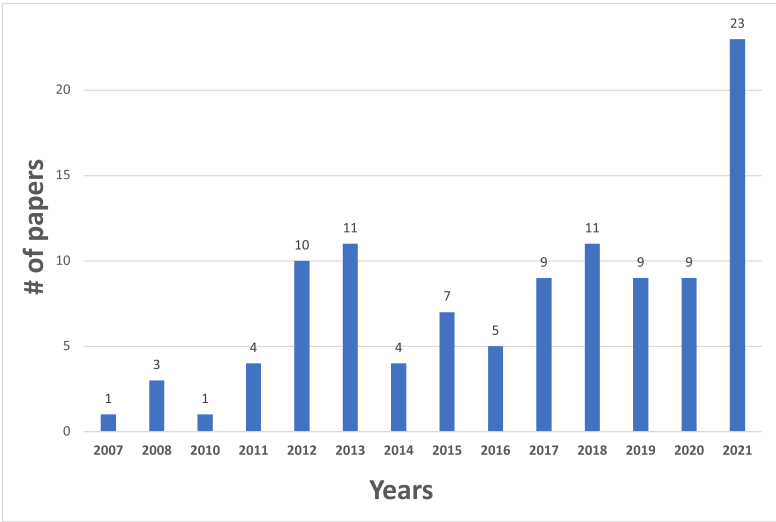
Fig. 2. The distribution of surveyed articles published per year since 2007.

The scope of this survey article was limited to articles published in major visualization conferences and journals. We did not search for any relevant visualization work for image and video data in other areas, such as big data, parallel computing, high-performance computing, and social media. There are also articles focused on interactive visualizations of deep learning networks [50, 62, 124, 144], but we considered visualization domain articles focusing only on image and video datasets.

Table 1a. Technique and algorithms, Visualization techniques and Application areas taxonomy and corresponding coded articles from the visualization domain related to image and video datasets

Years	References	Techniques And Algorithns									Visualization Techniques						Application Areas											
		Dimensionality Reduction	Regression	Clustering	Correlation	Machine Learning Methods	Optimization	Moving Average	Histograms and Aggregations	Other/Custom	2D Techniques	3D Techniques	Geometrically Transformed Displays	Iconic Displays	Dense Pixel Displays	Stacked Displays	Sports Analytics	Content Synthesis and Removal	Surveillance, Activity Analysis…	Video Editing, Stylization…	Video and Image Collection Analysis	Distance Education and MOOCs	Interaction Supported Learning	Video Storyboard and Summarization	Augmented and Virtual Reality…	Video Stereolization	Image Analysis, Editing…	Medical Applications
07	Bitter et al. [14]											X																X
2008	Botchen et al. [21]				X			X			X			X				X						X				
2008	Romero et al. [122]							X			X	X															X	
2008	Wang et al. [158]							X		X		X															X	
10	Liu et al. [86]									X	X															X		
2011	Kagaya et al. [61]									X	X										X							
2011	Pretorius et al. [119]			X							X								X									
2011	Parry et al. [109]									X	X							X										
2011	Xiao et al. [171]									X	X										X							
2012	Bowman et al. [22]										X				X	X					X							
2012	Ha et al. [47]									X	X							X										
2012	Höferlin et al. [58]										X								X					X				
2012	Höferlin et al. [57]										X								X					X				
2012	Itoh et al. [59]			X		X					X										X							
2012	Liao et al. [81]			X			X				X																X	
2012	Donovan and Hertzmann [105]							X			X												X					
2012	Sunkavalli et al. [148]									X	X							X										
2012	Wei et al. [163]																										X	
2012	Yoon et al. [178]									X	X							X								X		
2013	Debernardis et al. [35]																								X			
2013	Flagg and Rehg [42]	X																X						X				
2013	Kyprianidis et al. [71]	X			X	X					X									X								
2013	Lin et al. [83]			X							X							X										
2013	Legg et al. [75]				X						X			X			X											
2013	Li et al. [78]										X								X					X				
2013	Lu et al. [89]										X							X										
2013	Meghdadi and Irani [95]										X								X					X				
2013	Perin et al. [112]			X							X				X		X											
2013	Schmidt et al. [128]							X			X				X												X	
2013	Schultz and Kindlmann [130]											X															X	X
2014	Klemm et al. [67]			X								X																X
2014	Herling and Broll [52]									X	X							X										
2014	Polk et al. [118]										X			X			X											
2014	Semmo and Döllner [133]									X	X																X	

(years: 2007 to 2014).

3 IMAGE AND VIDEO DATA RESEARCH IN THE VISUALIZATION DOMAIN

In this section, we review articles in the visualization domain related to image and video datasets, and categorize them based on various taxonomies such as algorithms and techniques, visualization techniques, interaction methods, and application areas. The categorization of these articles is summarized in Table 1a, 1b, and 1c. In these tables, articles are grouped based on the year of publication, providing an overview of the work done over the years. However, within each year, the articles are not listed in any particular order as we do not think ordering is essential within a year. This table provides insights into the current coverage of these taxonomies in research relevant to image and video datasets in the visualization domain. We also discuss data types, scale, and dimensions of data used in research. Furthermore, we also identify higher-level representative

Table 1b. Technique and algorithms, Visualization techniques and Application areas taxonomy and corresponding coded articles from the visualization domain related to image and video datasets

Years	References	Techniques And Algorithms									Visualization Techniques						Application Areas											
		Dimensionality Reduction	Regression	Clustering	Correlation	Machine Learning Methods	Optimization	Moving Average	Histograms and Aggregations	Other/Custom	2D Techniques	3D Techniques	Geometrically Transformed Displays	Iconic Displays	Dense Pixel Displays	Stacked Displays	Sports Analytics	Content Synthesis and Removal	Surveillance, Activity Analysis…	Video Editing, Stylization…	Video and Image Collection Analysis	Distance Education and MOOCs	Interaction Supported Learning	Video Storyboard and Summarization	Augmented and Virtual Reality…	Video Stereolization	Image Analysis, Editing…	Medical Applications
2015	Duffy et al. [39]												▪											▪				
	Jang et al. [60]		▪													▪	▪											
	Kindlmann et al. [66]									▪	▪																▪	
	Kurzhals et al. [70]									▪	▪								▪									
	Lin et al. [82]					▪					▪																	
	Stengel et al. [143]							▪			▪								▪									
	Shi et al. [135]												▪		▪						▪							
2016	Amini et al. [6]												▪														▪	
	Bryan et al. [23]								▪				▪														▪	
	Chen et al. [25]	▪							▪				▪											▪			▪	
	Gutenko et al. [46]												▪															▪
	Hermann et al. [53]	▪											▪															▪
2017	Baldacci et al. [8]									▪	▪							▪									▪	
	Crissaff et al. [34]			▪						▪	▪																▪	
	Meka et al. [96]					▪				▪	▪														▪		▪	
	Poco et al. [116]					▪				▪	▪																▪	
	Pjanic et al. [115]									▪	▪														▪		▪	
	Rübel and Bowen [123]									▪	▪														▪		▪	
	Song et al. [141]			▪						▪	▪																▪	
	Turban et al. [151]	▪		▪						▪	▪																	
2018	Wu et al. [169]									▪					▪		▪										▪	
	Boorboor et al. [16]												▪		▪													
	He et al. [51]										▪				▪													
	Lu et al. [88]										▪										▪							
	Lobo et al. [87]										▪																▪	
	Lai et al. [72]						▪		▪		▪						▪											
	Stein et al. [142]					▪			▪		▪						▪											
	Wang et al. [160]										▪				▪													
	Wu and Qu [167]	▪								▪	▪						▪										▪	
	Wu et al. [170]	▪		▪							▪										▪							
	Xie et al. [172]					▪					▪										▪							
	Zhang et al. [183]					▪					▪													▪				
2019	Chan et al. [24]								▪		▪										▪							▪
	Fan et al. [41]	▪									▪										▪							
	polk et al. [117]									▪	▪																	
	Serrano et al. [134]			▪							▪								▪									
	Wang et al. [161]					▪				▪	▪																▪	

(years: 2015 to 2019).

task requirements based on application areas. Lastly, we provide details of the tools and libraries used in the visualization domain in the surveyed collection of articles.

3.1 Automated Techniques and Algorithms

In this section, we review the techniques and algorithms related to machine learning, statistics, and computer vision used in the surveyed articles. We grouped the algorithms and techniques into high-level categories, as shown in Table 1a, 1b, and 1c. We adopted an initial high-level categorization based on the taxonomy proposed by Patgiri [110] and adjusted these categories while reviewing the techniques and algorithms used in the surveyed articles and eventually merged similar ones.

Table 1c. Technique and algorithms, Visualization techniques and Application areas taxonomy and corresponding coded articles from the visualization domain related to image and video datasets

Years	References	Dimensionality Reduction	Regression	Clustering	Correlation	Machine Learning Methods	Optimization	Moving Average	Histograms and Aggregations	Other/Custom	2D Techniques	3D Techniques	Geometrically Transformed Displays	Iconic Displays	Dense Pixel Displays	Stacked Displays	Sports Analytics	Content Synthesis and Removal	Surveillance, Activity Analysis…	Video Editing, Stylization…	Video and Image Collection Analysis	Distance Education and MOOCs	Interaction Supported Learning	Video Storyboard and Summarization	Augmented and Virtual Reality…	Video Stereolization	Image Analysis, Editing…	Medical Applications
	Tan et al. [149]																											
	Zeng et al. [3]																											
	Halter et al. [49]																											
	Andrienko et al. [7]																											
2020	Zeng et al. [182]																											
	Wen et al. [165]																											
	Zhang et al. [184]																											
	Zahálka et al. [180]																											
	Bi et al. [13]																											
	Krueger et al. [69]																											
	Berson et al. [12]																											
	Chen et al. [27]																											
	Chen et al. [26]																											
2021	Weiss et al. [164]																											
	Xu et al. [173]																											
	Wang et al. [157]																											
	Seebacher et al. [131]																											
	Somarakis et al. [139]																											
	Nguyen et al. [103]																											
	Ma et al. [92]																											
	Shu et al. [138]																											
	Meuschke etal. [97]																											
	Dmitriev et al. [37]																											
	Ma et al. [91]																											
	Wang et al. [162]																											
	Pan et al. [107]																											
	Huang et al. [55]																											
	Fu et al. [43]																											
	Park et al. [107]																											
	Li et al. [80]																											
	Ghahremani et al. [44]																											
	Corvò et al. [33]																											
	Xu et al. [175]																											
	Shi et al. [136]																											
	Zhu and Kelly [185]																											
	Somarakis et al. [140]																											
	Total Entries	10	4	32	16	27	6	3	12	55	88	27	16	17	27	6	10	8	11	22	12	4	3	12	6	5	48	19

(years: 2019 to 2021).

3.1.1 Dimensionality Reduction. Visualization of high-dimensional data is often facilitated by applying different dimensionality reduction methods by converting high-dimensional data into lower meaningful dimensions [25, 94, 127, 140, 152, 153]. Turban et al. [151] used **principal component analysis** (**PCA**) to analyze the impact of criteria correlations on the distribution of the data, and to reduce the high-dimensional data in the video dataset. Herman et al. [53] used PCA spaces to make comparisons between different models, and support methods such as PCA navigation or browsing.

3.1.2 Regression. Identifying the multivariate relationship between data variables is one of the critical challenges in data analysis, especially when the number of variables considered is large.

Regression methods coupled with visual analytics can facilitate causal analysis and can be applied in areas such as sports analytics. ForVizor system [170] facilitates the analysis of dynamic changes in player formations during soccer matches under varying temporal and spatial scenarios. Utilizing the least-squares method, the system acquires player positions by mapping the tracking result from the real world to a 2D plane.

3.1.3   Clustering. Clustering is frequently used in visualization research to group items together based on their similarities [8, 60, 123, 151]. The effectiveness of results obtained by clustering methods varies depending on the intended usage and application, which has given rise to the development of different clustering algorithms. K-means is considered to be one of the most popular clustering algorithms used to find the centroid of clusters, where the number of clusters is represented by K [63]. In [151], an algorithm based on spatial prediction, pyramidal computation, and human vision characteristics was proposed for peripheral extension of existing video (movie) content. The authors used the K-means clustering algorithm to cluster each set of data scores (Enjoyment, Comfort, Consistency, Presence, and Emotion), where each score is assigned by a user to one video only.

Hierarchical clustering, which builds a tree-like clustering structure and maintains the relationships between different clusters, is used in many of the surveyed articles related to visualization [33, 67, 75, 119, 128, 130]. Schultz and Kindlmann [130] used adapting spectral clustering for specific image analysis tasks, which involved exploring hierarchical, spectral embeddings, and tuning parameters. Their work focused on 3D medical image analysis and they proposed a framework that maps a spectral clustering-based high-dimensional features space to a three-dimensional data space.

3.1.4   Correlation. Correlation is a task of association and quantification of the strength and direction of the relationship between numeric variables. Botchen et al. [21] proposed a technique to dynamically detect events and activities in video by converting it into a series of snapshots. A human figure is tracked in the input video to generate spatiotemporal movement data. An optical flow descriptor is then used to characterize the motions of different body parts. Spatiotemporal cross-correlations are computed to find the similarities between the motion descriptor and the stored database of action fragments.

3.1.5   Machine Learning Methods. With the advancements in machine learning, and especially deep learning, more advanced libraries and frameworks are being developed. Table 1a, 1b, and 1c show an increasing trend in the use of deep-learning methods in visualization research focused on image and video datasets. These results accord with discussions with the domain experts, who also emphasized that there will be an increasing focus on deep-learning methods due to advancements in computation resources.

Serrano et al. [134] proposed a technique to generate a 3D visualization from a 2D visualization, which is carried out by extracting per-frame depth map information from the video data using a **convolutional neural network** (**CNN**). VC-Net [162] utilized an end-to-end CNN-based framework to segment and visualize 3D sparse microvascular structures by leveraging information from **maximum intensity projection** (**MIP**).

Zeng et al. [182] implemented a visual analytics system to generate emotional summaries from classroom videos that uses an adapted CNN model to recognize facial expressions. Facetto [69] system integrates a CNN-based framework for cell classification that supports semi-automated analysis of high-dimensional multi-channel images in cancer studies. Zhu et al. [185] used a series of CNN-based networks to synthesize scale- and space-continuous satellite images conditioned on cartographic data.

Since processing image and video datasets is often intended to analyze content and summarize information [172], deep-learning techniques such as CNN are more commonly utilized than other deep-learning methods, as they are focused on analyzing the spatial content based on convolutional networks. In datasets and problems where analyzing temporal components of the dataset is desired, such as analyzing speech patterns, techniques such as **recurrent neural network** (**RNN**) are generally utilized. Different variants of deep-learning frameworks are used in practice. A combination of CNN and RNN was utilized by Bi et al. [13] to learn patterns of vehicle trajectories at intersections, which supports vehicle editing and generation of new simulations. Zhang et al. [184] used BASNet [120], a deep neural network-based architecture composed of encoder-decoder framework and a residual refinement module, to embed information into visualization images.

Gaining insights and understanding the internal decision process is important to trust the machine learning-based methods. Dmitriev et al. [37] utilized a visual analytics-based approach to explain the rationale behind a **computer-aided diagnosis** (**CAD**) for pancreatic lesions based on **random forest** (**RF**) and CNN components. This rationale is based on the visual analysis of these individual components.

3.1.6 Optimization. In the surveyed articles, certain problems are modeled as optimization problems. These include affine scaling and reduction methods for iterative non-linear optimization [149], unconstrained minimization [8], and contour optimization [82]. There are also some general non-convex quadratic programming problems. Liao et al. [81] used a combination of motion analysis with user interaction to convert videos into stereoscopic videos. The authors recover dense depth-maps information by analyzing the optical flow for all frames and utilizing a quadratic programming technique to recover both quantitative and qualitative depth information.

3.1.7 Moving Average. The moving average, also called the "rolling mean" or "moving mean", is a well-known statistical technique to analyze data points. Different variants of the moving average are commonly used. These variants include the exponential moving average [21], the weighted moving average [25], and the simple moving average [142]. Stein et al. [142] designed an interactive system for extracting player movements and visualizing their trajectories. If an individual player's position is incorrectly detected, it may lead to incorrect detection of the other players' movements. The authors used a moving average filter to overcome the scenario whenever a transformed player's position differs from the actual position.

3.1.8 Data Aggregations. Data aggregation is a data-processing technique that gathers and compiles data in a summarized format, which can then be used for further statistical analysis or in visualization representations [122, 128, 143]. Viz-A-Viz [122] is a video analytics tool to analyze activities in videos, generating spatial, temporal, and semantic aggregations based on an activities dataset. Statistical tools like histograms are often used to aggregate data by showing the distribution of the data. A color histogram is used to represent the distribution of the colors in image and video datasets [48, 49, 142, 170]. Vian [49], a visual film annotation system, uses a 3D space-filling curve to map a color histogram to 1D feature vectors.

3.1.9 Others. In addition to techniques and algorithms related to machine learning and statistics, the surveyed articles also employed other computer vision methods. We discuss some of these methods here.

Segmentation [42, 128] techniques are used to subdivide images into segments or parts based on different criteria, and those parts are then used in later stages of the image or video processing pipeline. There are different segmentation techniques, including color segmentation [128], graph-based segmentation [83], and semantic segmentation [72].
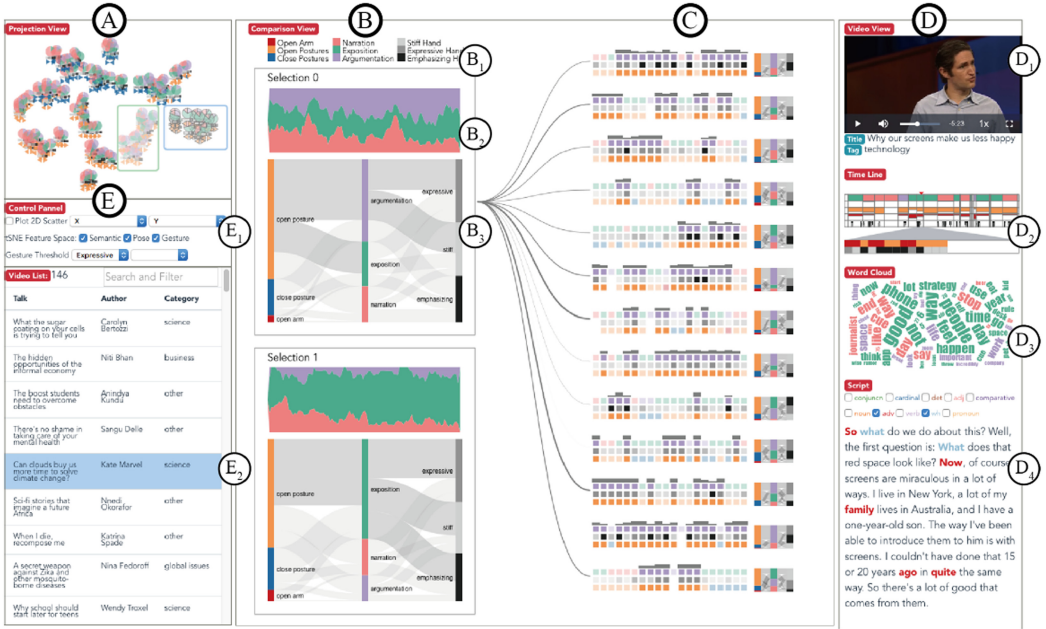
Fig. 3. Multimodal Analysis of Video Collections: Visual Exploration of Presentation Techniques in TED Talks visualizations [167].

Multiview stereo (MVS) enables reconstruction of 3D scenes from multiple calibrated static images captured from varying view points for the same scene. Liu et al. [86] adapted an MVS algorithm to utilize point cloud for 3D reconstruction. Optical Flow encodes the pattern of movement of different objects contained within a set of images caused by the movement of the objects or the observer, and its applications include finding similarities between videos [8] and improving segmentation results [83]. Motion Estimation calculates the motion vectors in videos or a sequence of images. Sunkavalli et al. [148] utilizes affine transformations to represent camera motion for videos, which is also a form of motion estimation.

3.2 Visualization Techniques

Various visualization techniques have been used by researchers to analyze image and video datasets. Since these datasets are often complex with a large number of features, multiple coordinated visualization techniques are often used for interactive visualization [23, 25, 33, 167].

In this work, we use the taxonomy introduced by Keim [64] and Ko [68] for standard visualization techniques. In this section, we briefly review each type with examples from our surveyed articles. Table 1a, 1b, and 1c also demonstrate the categorization of visualization techniques for our surveyed articles.

3.2.1 2D Techniques. 2D techniques are among the most common standard visualization techniques used to visualize various features of image and video data. These techniques include 2D charts such as pie charts, bar charts, and histograms. These charts are also often used to show various analytics associated with the data.

In the design of an interactive tool for analyzing multiple TED talks, Wu and Qu [167] used multiple 2D plots, such as stacked line charts, sankey diagrams, and word clouds (Figure 3). For the analysis of snooker game data, Parry et al. [109] also used bar charts and line charts to

display temporal information. The Motion Browser [24] tool extensively uses line charts to analyze patients' therapy data (Figure 11). Animation can be used to show temporal changes; for example, Lobo et al.[87] efficiently demonstrated the use of animation to show temporal changes in satellite images. Overall, 2D techniques are commonly used in designing tools for visualizing image and video data [25, 117, 118, 139, 161, 169].

3.2.2 3D Techniques. With the advancement of immersive technology, 3D visualization techniques are becoming more important. Researchers often visualize image and video datasets in 3D to show them in immersive environments, such as head mounted displays [134]. Various research works have also been proposed to convert 2D visualization into 3D visualization by extracting depth and motion information from the video data [81, 86, 134].

Parry et al. [109] introduced a 3D visualization template for summarizing events from a snooker game. Meghdadi and Irani [95] presented a 3D space-time cube for visualizing movement trajectory data. Semmo and Döllner [133] used texture mapping for rendering 3D scenes, also allowed users to interactively apply various image filters. Volume rendering is a common 3D visualization technique that is often used for visualizing medical images [16, 53, 141]. Weis et al. [164] used deep learning-based architecture to upscale the resolution of isosurfaces to higher resolution. Nguyen et al. [103] proposed a novel technique for construction of 3D mesoscale biological models by extracting statistical and spatial properties from 2D microscopy scans, and internal assembly through an interactive 3D rule specification.

Generally, visualization researchers are cautious to use 3D visualization techniques because of the inherent problems of perception and occlusion [100]. Nevertheless, 3D visualization techniques for imaging and video data are becoming more important due to the advancement and availability of immersive environments [93].

3.2.3 Geometrically Transformed Displays. Multidimensional datasets often use geometrically transformed display techniques to visualize multiple dimensions of data. These techniques are often based on dimensionality reduction principles [153] for meaningful representations of data. Common examples of this category are scatterplots and parallel coordinates.

Höferlin et al. [57] used a cascaded scatter plot to show the quality of trained classifiers used in interactive learning by involving human experts for video visual analytics (Figure 8). In the design of GazeDx [141], the authors used scatterplot matrices for the comparison of gaze analysis data for medical images.

Parallel coordinates uses multiple parallel axes to display multidimensional data on a 2D display. Legg et al. [75] utilized parallel coordinates along with other visualization views in the design of a visual analytics system that supports sketch-based search for rugby videos. They used parallel coordinates to show similarity metrics data for each frame of the video. The PeakVizor [25] tool, which analyzes clickstreams data in **Massive Open Online Courses** (**MOOCs**) to understand the behaviors of learners, also used parallel coordinates in their design for the correlation analysis of different learner groups (Figure 7).

3.2.4 Iconic Displays. Iconic displays are another useful technique that encodes features and attributes of data in terms of color, shape, or glyphs. Various custom glyphs-based visualizations have been designed by researchers while visualizing various features of image and video datasets [30, 39].

Chen et al. [25] introduced treemap-based glyphs to show peaks in the clickstream data of learners accessing open online courses (Figure 7). While designing a visual analytics tool to support multimodal analysis of TED talks, Wu and Qu also used a novel treemap, radar chart, and nested pie chart-based design for glyphs to encode various attributes of presentation techniques used by

speakers [167] (Figure 3). In the design of the TenniVis tool, the authors also presented various novel glyphs for displaying multiple attributes of tennis-match data [118].

3.2.5   Dense Pixel Displays. This technique encodes each pixel on the display with a particular data attribute, and can visualize the largest possible amount of data on the viewport because it can use one pixel for each data point. Researchers often use various configurations or stacking of pixels to display different information via this technique. Grid or radial layouts are the most common configurations of the pixels layout for this category.

Wu and Qu [167] presented a matrix-based grid layout color-coded view for comparing various clusters while analyzing various TED talks (Figure 3). Wu et al. [169] also used a matrix-based grid layout for the interactive analysis of table tennis game data. In the design of AnaFe [46], the authors also used a heatmap-style color-coded grid visualization for showing temporal changes in the feature progression, while performing image analysis. Overall, this visualization category is not extensively used for the visualization of image and video datasets.

3.2.6   Stacked Displays. This category of the visualization technique is used to display hierarchical data. Researchers often use various styles of partitioning to represent the hierarchical structure of data [64]. A common example of this category is treemaps, which often encode hierarchical information by using nested rectangles. Jang et al. [60] used treemap visualization to display various motions in their design of the tool for analyzing human-motion data. Pretorius et al. [119] presented a tree-based node-link visualization for depicting the clustering hierarchy of input parameter space while performing interactive image analysis. Like dense pixel displays, this category is also not frequently used in visualizing image and video datasets.

3.3   Interaction Methods

We reviewed the articles included in the survey based on the taxonomy of interaction methods proposed by Yi et al. [177], using the interaction categories "Select", "Explore", "Reconfigure", "Encode", "Abstract", "Filter", and "Connect". In the surveyed articles, basic interaction techniques such as "Select" and "Explore" were generally used. The experts mentioned, in the discussions, their desire to have an "overview, details-on-demand" (R22) functionality in deep-learning frameworks for computer vision applications. They explained that this type of functionality can improve the understanding of datasets, and can help in different preprocessing tasks. Querying and filtering was very common in articles in research areas such as sports analytics [118], activity analysis [60], and medical applications [24] (R23). "Connect" is often used in image and video data visualizations as these visualizations generally consist of multiple coordinated views.

A few visualization tools exist that include detailed interaction support. The ARIES [34] system is designed to explore interactive image processing, exploration, and manipulation. Xie et al. [172] designed a visual analytics system that has comprehensive interaction support for semantic-based image analysis. The DeHumor [157] system provides multiple linked views to support exploration of multimodal humor features at multiple levels of detail to facilitate analyzing human behavior. There are a few further works that have more limited interaction support [163].

Interaction methods can be useful in deep-learning applications as they can provide insights into the internal model structure, which is usually a black-box in computer vision applications. Shepherd is another interaction method [5, 90] that either implicitly (indirect shepherding) or explicitly (direct shepherding) helps the user to optimize the modeling process. Model selection or setting model parameters is an example of direct shepherding, whereas setting soft or hard constraints through a visual interface, such as defining distance thresholds, is an example of indirect shepherding. This method can be adapted for deep-learning techniques, we did not find any example of this interaction technique in the surveyed articles. The What-If Tool [166] is an open-source tool that

Fig. 4. ForVizor: Interactively analyzing various team formations in a soccer game with multiple coordinated visualizations [170].

enables interactive probing of machine learning models to understand their behavior. Users can evaluate the model's performance by creating hypothetical scenarios, performing intersectional analysis, supporting flexible visualization of input data, and easily switching views. Interaction methods can facilitate understanding transfer learning between deep-learning models when training and adapting models for new tasks. Ma et al. [92] implemented a visual analytics application that helps users understand transfer learning at multiple levels (data, model, and features) through a suite of linked visualizations.

3.4 Application Areas

In this section, we summarize and group the application areas of our surveyed articles, explaining each application area with examples. Application area sub-categories were derived based on a higher-level area categorization of each article and merging similar categories. Tables 1a, 1b, and 1c present the application area coding of each surveyed article.

3.4.1 Sports Analytics. The interactive analysis and visualization of sports data are rapidly gaining popularity [30, 113]. Sports experts use analytics and visualization charts to analyze the game data and to plan for future games accordingly. Various works have been carried out to interactively analyze the images and video data related to soccer [131], table tennis [169], tennis [118], rugby [29, 31], and other sports.

Stein et al. [142] proposed an interactive system that automatically extracts and visualizes object trajectories for visual analysis of team sports matches. They focused on soccer data and extracted the players' movements and analytics and put them into the original video. Their proposed method was implemented on GPUs for faster processing. That work helped domain experts to achieve a better analysis of such data. Wu et al. [170] presented an interactive visual analytics system called "ForVizor" to analyze dynamic changes in soccer team formation. Their front-end system consisted of multiple coordinated graphical views to show team formation flow and the changes happening (Figure 4). Perin et al. [112] proposed a tool to allow soccer analysts to analyze different phases and events of the game and communicate their insights in terms of visual stories. Seebacher et al. [131] supported the creation of spatio-temporal queries through the placement of magnets on a virtual tactic board to perform a similarity search in massive soccer datasets. Parry et al. [109] also presented a framework for video storyboarding, summarizing the main events of the game video. Although this work focused solely on snooker videos, in practice it could be applied to videos of other sports.
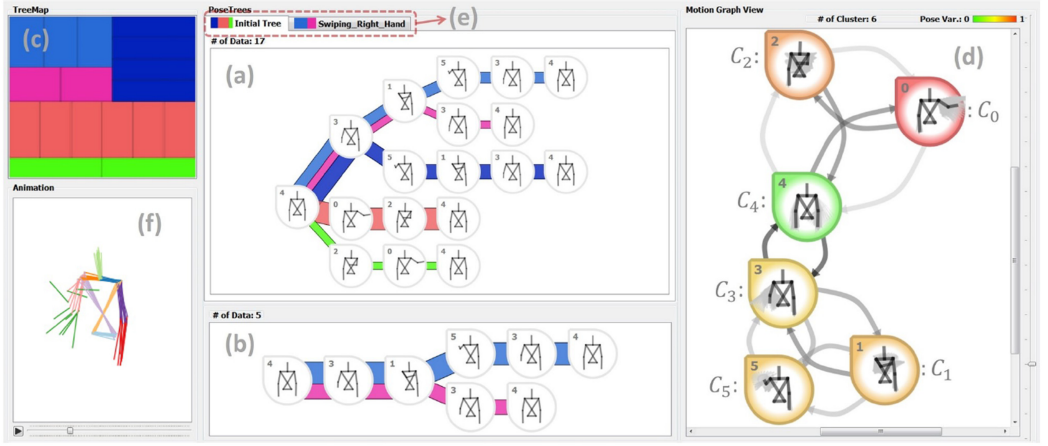
Fig. 5. MotionFlow for pattern analysis of human motion data visualization [60].

Legg et al. [75] designed a visual analytics system that supports a sketch-based search for rugby videos. Their system supports model visualization (based on parallel coordinates), search space visualization, search results visualization, and an interface to accept or reject results that, in turn, can improve the model by adjusting the parameter weights of an active learning model (supervised learning). Sketch-based search supports spatiotemporal attributes like motion, position, distance, trajectory, and spatial occupancy.

Wu et al. [169] proposed a visual analytics system "iTTVis" for the exploration and analysis of table tennis data. iTTVis presents visualizations that support three main perspectives: time-oriented, statistical, and tactical analysis, and also supports correlative analysis and identification of tactical patterns along with a score timeline. Similarly, various interactive visualization tools have been designed for the analysis of tennis match data [117, 118].

3.4.2 Content Synthesis and Removal. Content synthesis and removal have applications in areas such as gaming, entertainment, and architecture. This research facilitates designers in composing their desired content by utilizing existing natural examples [8, 42, 78, 165].

Flagg and Rehg [42] introduced crowd tubes for synthesizing video-based crowds, which are constraint-satisfying video objects placed by designers in a specific place and time in a video volume with an associated trajectory. Andrea et al. [8] utilized the similarity of optical flow between the reference video and the camera paths of a given 3D scene to generate a video of the scene resembling the reference video. The similarity in optical flow is due to the similarity in camera movement and scene geometry. Li et al. [80] proposed a framework to synthesize cartoon videos by using color information from the keyframes and animation information from the sketch. The correspondence between sketch and keyframes is used to create a blended image and then uses the estimated optical flow information from the user sketch to generate interpolated video.

3.4.3 Surveillance, Activity Analysis and Recognition, Motion Analysis and Tracking. Visual surveillance and activity analysis and recognition are active areas of research, and interactive visualization coupled with computer vision techniques can facilitate different task requirements in this field [58, 95]. MotionFlow [60] helps to understand the patterns of gestures through visualization of motion-data sequences, and also supports comparative analysis between different gestures (Figure 5). Wang et al. [160] used silhouette as a cue for 3D-pose estimation on handheld cameras
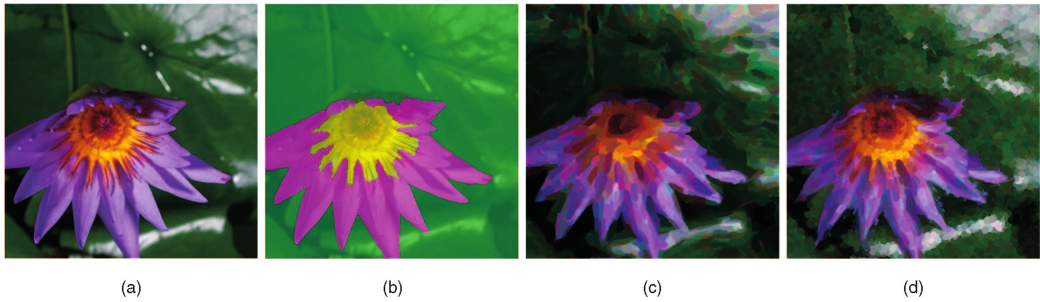
Fig. 6. Stylized frames of a blooming flower from a video. [61], (b) segmentation results, (c) same style applied to entire frame, and (d) different styles applied to petals, leaves, and stamens.

using motion capture systems. Romero et al. [122] proposed a system for activity analysis based on visualizing and analyzing overhead videos.

Video visual analytics are used to interpret data from surveillance cameras. Botchen et al. [21] investigated the inclusion of action-based details in video visualization by representing the video content in 3D to depict motion events and individual objects. In the context of video enhancement, many existing methods focus on manipulating visual content in the video. Stengel et al. [143] proposed a patch-based method to refine blurry frames of input video for eye-motion prediction when being watched.

Pattern analysis of human motions has broad applications, but some challenges remain. There are different motion pattern styles, and context information may be associated with such datasets. Visual analytics can facilitate such analysis of multidimensional spatiotemporal datasets through multiple linked visualizations.

3.4.4 Video Editing, Stylization, and Painterly Animation. Wei et al. [163] proposed a method for converting distorted fisheye videos into natural-looking video sequences that preserve temporal coherency. This method supports interactive annotations to guide the correction process, and utilizes six distinct correction criteria expressed as quadratic energy functions. Berson et al. [12] integrated a generative RNN-based framework for editing facial animations to generate facial motions to fill or replace missing segments.

Video stylization and painterly animations [61, 88, 105, 178] are emerging areas of research and have applications in areas such as movies, social media, and entertainment. Lu et al. [88] designed a real-time video stylization system that uses object-flow construction based on a novel learning-based technique, which is sufficiently robust to overcome partial occlusions, problems in optical flows, and unknown object transformations. The method also supports different painterly styles. Yoon et al. [178] preserved spatial and temporal coherence by constructing a stabilized 3D-feature flow field using a combination of a 3D Sobel operator and smoothness based on color similarities and saliency features. Kagaya et al. [61] provided a painterly rendering system that enables spatial and temporal variation of style parameters and brush stroke orientation, supporting features such as the ability to emphasize/de-emphasize certain objects, to modify contrast between neighboring objects, and to adjust the level of abstraction. Figure 6 shows an example of two stylized frames of a blooming flower.

3.4.5 Video and Image Collection Analysis. Keeping in view the scale and complexity of image and video collection datasets, visual analytics can facilitate analysis through multiple linked visualizations and support for interactive querying and filtering. Zahálka et al. [180] introduced II-20 (Image Insight, 2020), which dynamically models analytic categorization of image collections
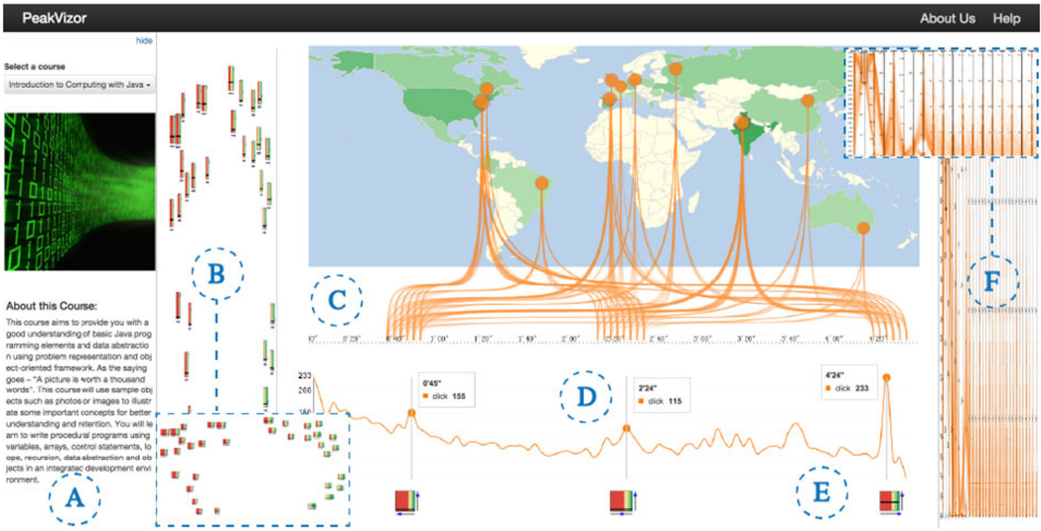
Fig. 7. Peakvizor: Visual exploration of peaks in the clickstream data relevant to a selected course [25].

based on user interactions. Pan et al. [107] generated visual summaries of image collections based on content diversity, conciseness, and visual aesthetics and applied a backpropagation algorithm to optimize the layout of the collage. Wu and Qu. [167] proposed an interactive visual analytics system to support multimodal analysis of TED talks focusing on presentation styles. Multimodal content consists of frame images, text and meta-data. There are three major views: projection view (for cluster analysis), comparison view (for intracluster analysis), and video view. The authors' analysis was focused on body postures, gestures, and rhetorical aspects of presentation (Figure 3). Recent research has focused on integrating machine intelligence with visualization to understand complex and large-scale data.

Image-Set Processing Streaming is an advanced technique in image processing that uses streams consisting of either image pixels or image sequences. Image population analysis is an essential method for understanding the evolution of a population that requires extensive computational power and memory. Ha et al. [47] presented a framework to solve this extensive computational problem with heterogeneous CPU or GPU-based systems. The authors presented an out-of-core solution that performed at the same level as that of an in-core solution, providing various examples to demonstrate the efficiency of the framework.

A comparative visualization was proposed by Schmidt et al. [128]. They proposed a multi-image view technique to visualize the similarities and differences in satellite image sets. LADV [91] used a deep-learning-based framework to learn design intentions from existing exemplars (dashboard images) or sketches to synthesize dashboard templates.

3.4.6   Distance Education and Massive Open Online Courses. Distance education programs and MOOCs platforms like Coursera, edX, and Udacity have gained significant popularity in recent years [135]. These platforms offer great flexibility in terms of timing, courses offered, and access methods. The MOOC platforms offering such courses are interested in the web access logs (click stream data) of these courses to analyze learner interactions and engagement with the course material [26]. Visual analytics tools like PeakVizor [25] (Figure 7) enable experts to gain insights that are otherwise difficult to discover from raw data. PeakVizor features include analyzing peaks (regions of interest) in clickstreams, extracting anomalies, identifying different learner groups and
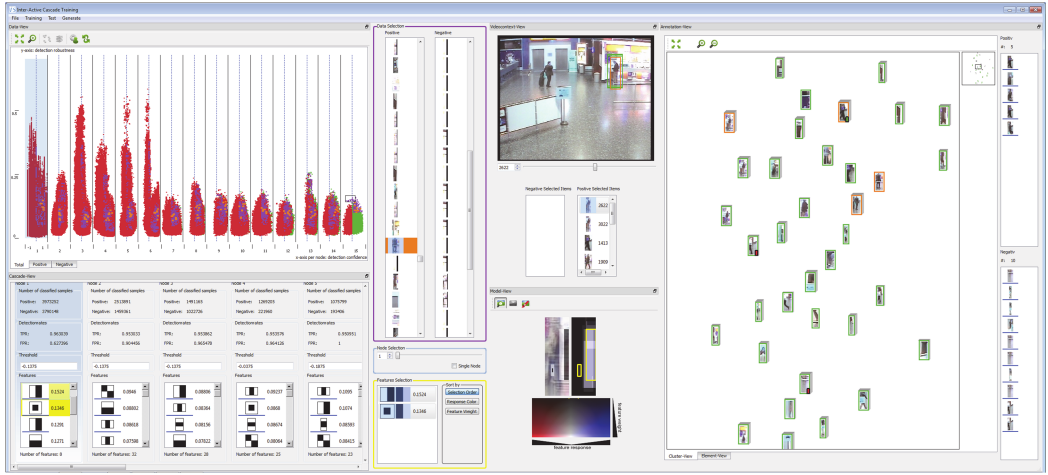
Fig. 8. Visual analytics-based interactive learning system [57].

their correlations in different peaks, and discovering patterns, the spatio-temporal distribution of clicks, and geographical and behavioral distribution of learners. He et al. [51] designed a visual analytics system "VUSphere" for the exploration and comparative analysis of video utilization in courses, students, and distance-learning centers.

3.4.7 Interaction Supported Learning. Visual analytics-based methods can provide insights into classifier performance and facilitate model manipulation by interactively adjusting data labels and retraining. Höferlin et al. [57] proposed an inter-active learning-based framework that supports interactive data querying and selection, annotating data instances, iterative classifier refinement, model visualization and direct manipulation, and visual analysis of classifier performance through cascaded scatterplots (Figure 8). Their results showed that, in certain instances, this form of inter-active learning can help achieve classifier performances comparable to other learning methods within a few cycles. Work by Huang et al. [55] supports the interactive analysis and understanding of multiple attributes learning models for x-ray scattering images, by visual exploration in embedding spaces defined on multiple criteria.

3.4.8 Video Storyboard and Summarization. Summarizing a large video to allow a user to quickly see the important events contained within it is another important application area. Researchers have also designed techniques to play non-important parts of a video at a faster rate [58, 73] or to add spatial context in the video for quick analysis. Wang et al. [158] conducted user studies to show that adding spatial context to a video helps participants to better understand it. Flagg and Rehg [42] presented a system to synthesize a crowd from the input video of natural crowds. AutoClips [136] generates videos automatically based on given data facts, utilizing a fact-driven clip library and an algorithm that selects clips, arranges them, and configures duration.

Sunkavalli et al. [148] presented a framework to generate a high-quality snapshot from a video clip along with a visual summary of the activities in the video (Figure 9). Meghdadi and Irani [95] proposed a video visual analytic system "sViSIT" to allow users to interactively search and track objects in a video. Their system automatically extracts all paths of an objects' movements and allows them to be visualized in different views and forms. Users can query and retrieve any data from the video. Botchen et al. [21] proposed a technique to detect events and activities in a video dynamically by converting it into a series of snapshots. Parry et al. [109] also proposed a video

(a) Example video frames

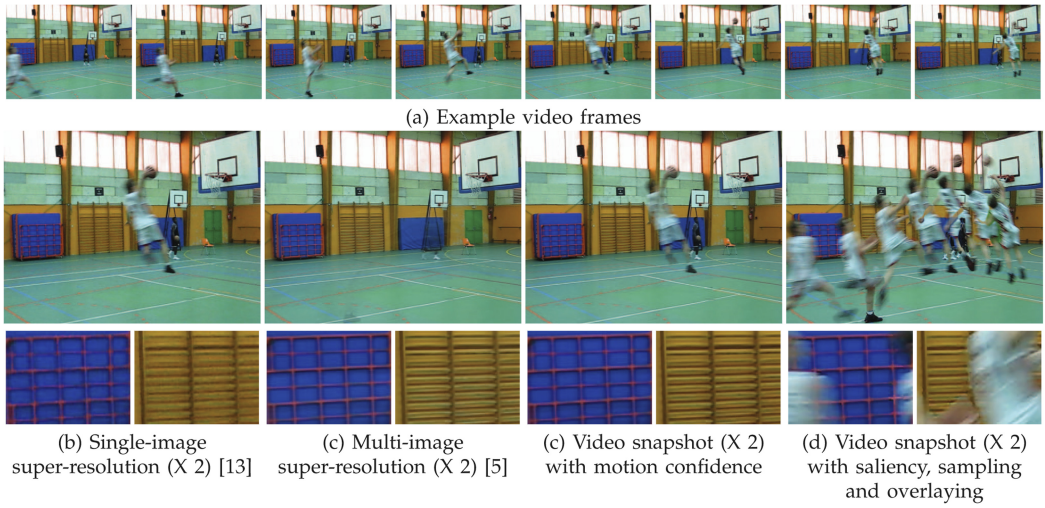| (b) Single-image super-resolution (X 2) [13] | (c) Multi-image super-resolution (X 2) [5] | (c) Video snapshot (X 2) with motion confidence | (d) Video snapshot (X 2) with saliency, sampling and overlaying |

Fig. 9. Video Summarization: Showing the main events of the video as the visual summary [148].

summarization system called video storyboard, which is a summarized video with important frames and activities enhanced by illustrative annotations. Perin et al. [112] demonstrated a tool called "SoccerStories" to allow soccer experts to interactively analyze quantitative game data with game context such as player positions, player actions, and player movements. Their tool also helps to effectively communicate the revealed insights. Shu et al. [138] conducted an exploratory user study to understand the impact of different data-GIF designs on storytelling and provide guidelines for effective designs.

3.4.9 Augmented and Virtual Reality and Telepresence. Recently, augmented and virtual reality environments have become more common for analyzing complex image and video visualizations. Head-mounted display devices are inexpensive and provide an immersive experience. Debarnardis et al. [35] evaluated various specifications of text visualizations on head mounted augmented reality displays. Serrano et al. [134] presented a technique for real-time playback of 360-degree videos in virtual reality headsets by adding parallax. The evaluation of their technique showed that the technique improves a users viewing experience. Turban et al. [151] proposed an algorithm (referred as Extrafoveal) based on spatial prediction, pyramidal computation, and human vision characteristics for peripheral extension of existing video (movie) content to improve the immersive experience.

Decomposing a video to augment the information contained within it has also been evaluated in the domain. Meka et al. [96] introduced a novel real-time method for the interactive intrinsic decomposition of scenes. Users can interactively improve the decomposition by using a mouse or through touch. The touch interaction also allows the user to place decomposition constraints directly in the 3D space. The authors' method supports a wide variety of interactive applications, such as photorealistic recoloring, material editing, and geometry-based relighting (Figure 10). The presented method is also the foundation for many augmented reality applications. Lin et al. [83] presented a video retargeting method where 3D space-time objects are transformed by as-rigid-as-possible warping, whereas non-significant objects undergo linear rescaling. This method results in better motion and shape preservation compared to other state-of-the-art methods.

Telepresence is a technique that allows people to remotely visit and interact with other people in distant locations. Zhang et al. [183] presented a 360-degree video camera-based redirected

Fig. 10. Augmented Reality Application: Dynamic relighting of an image [96].

walking robotic platform to support interaction and exploration of remote environments. The robotic platform was controlled remotely by the user wearing a head mounted display.

3.4.10 Video Stereolization. Video Stereolization is a technique that converts a video to stereoscopic video for 3D viewing. Various algorithms and techniques are designed to capture motion and depth information in the videos to convert them into stereoscopic videos. Liao et al. [81] presented a system that combines motion analysis of a video with user interactions for conversion to stereoscopic video. Liu et al. [86] proposed a three-stage (point cloud extraction, merging, and meshing) multiview stereo algorithm based on point-clouds to generate free-viewpoint videos. Their point-cloud extraction is resistant to occlusions, noise, and lack of texture. Serrano et al. [134] also introduced a method for displaying a 360-degree video in virtual reality head mounted displays.

Some research has also been carried out to combine various projection devices and displays/recordings. Pjanic et al. [115] demonstrated a calibration method to mix different projection display devices. Their work was focused on displaying content accurately on a 3D surface.

3.4.11 Image Analysis, Editing, Summarization, and Matching. Nearest patch matching techniques have recently emerged as a powerful tool for image and video matching, editing, and summarization, and are based on finding the most similar patch pairs between a source and target image [171]. Tan et al. [149] presented a technique to determine the distribution of paint pigment from an RGB image. This allowed users to make image editing operations in pigment space, and to perform operations such as edge enhancement, tonal adjustment, and recolor. Poco et al. [116] also introduced a technique for extracting color encodings from the bitmap images. Users can interactively verify the colors and correct them, if needed. Chartem [43] enables embedding additional information into chart images without having an impact on perception to facilitate reuse or repurposing. Chen et al. [27] proposed a framework based on R-CNN to recover 3D-shaped generalized cuboids and cylinders from a single photograph. Flyfusion [175] introduced a topology compactness strategy for the robust reconstruction of topology changes while reconstructing dynamic scenes using flying depth cameras.

Visualization of complex 3D models and scenes is difficult as the rendering of such scenes is often computationally expensive, and adding interactive navigation and analysis is challenging. Sunkavalli et al. [148] proposed a method for extracting a snapshot from a video, using an importance-based system to generate images using weighted values of image pixels. Presentation of 3D models using videos has also been proposed by Baldacci et al. [8]. This system provides many operations for a user, such as noise and blur reduction, super-resolution, and best focus selection, and also provides a visual summary of activities. Recent research has focused on optimizing input parameters of image analysis algorithms. Pretorius et al. [119] used a parameter optimization process coupled with key user requirements, and then developed a tool where users can examine the relationship between output and parameter values.
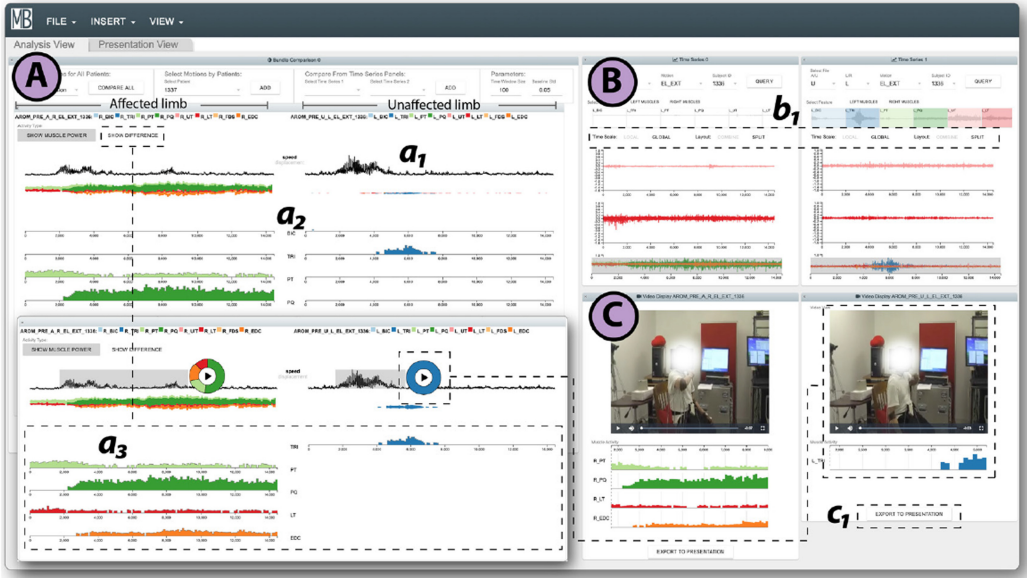
Fig. 11. Motion browser system to combine heterogeneous datasets (sensors, motion sensors, movement data, and videos) [24].

3.4.12 Medical Applications. 3D medical imaging has inspired the rapid development of visualization techniques for 3D medical image analysis [16, 44, 53, 130]. Chan et al. [24] presented a visual analytics system called "motion browser" that takes heterogeneous sensors and video-based patient hand therapy data as input, and allows users to explore the data interactively. The system comprises multiple coordinated views to allow physicians to compare and explore patient therapy data from multiple sources. The user can annotate the video data and compare it with sensors data (Figure 11). GUCCI [97] provides a suite of visualizations to compare and analyze blood flow data in the aorta of selected cohorts, helping to establish normal value ranges and derive guidelines.

In the field of medical research, many gaze analysis studies have been conducted to understand how radiologists read various types of medical images [77]. Song et al. proposed a tool called "**GazeDiagnosis (GazeDx)**", which is an interactive visual analytics framework to compare gaze datasets from multiple users working with image datasets [141]. The CMed [108] system supports interactive exploration of crowdsourced medical image data annotations with the support for interactive querying, and analysis from different aspects.

3.5 Data Types, Scale, and Dimensions

Image and video datasets used in practice contain not only the visual content that forms part of the raw images and videos, but also information derived from such datasets. These datasets have diverse characteristics; they are multi-dimensional, hypervariate, spatial and temporal, heterogeneous, hierarchical, augmented, network, and multi-resolution, among others. Table 3 shows different datasets used in computer vision and visualization research focused on image and video datasets, respectively. The nature and characteristics of these datasets vary based on the relevant application areas and underlying task requirements supported in the implementation. This variety in the nature of datasets poses unique visualization challenges.

The datasets may be extended further at different stages of the processing pipeline due to additional data generated by algorithms and techniques involved in the corresponding

implementations. Xie et al. [172] extracted semantic information from images and used a deep-learning framework based on CNN and LSTM to generate their descriptions. Bryan et al. [23] generated annotations to produce temporal summaries for time-varying datasets. DataClips [6] enabled interactive creation of data videos using existing data clips.

3.6 Task Requirements

We reviewed the visualization and visual analytics research related to image and video data to identify the major task requirements, and then grouped these based on the application areas. We also reviewed the recent major computer vision conferences and identified application areas relevant to image and video data analysis. This facilitated the comparison of task requirements and trends in the computer vision research and the visual analytics research related to image and video datasets.

Table 2 summarizes the major task requirements, organized into different application areas, based on the surveyed articles in the visualization domain. We identified representative higher-level task requirements (Column 3 : Task Requirements (Visualization)) aiming at explaining the needs of visualization researchers analyzing image and video datasets in the respective application areas. These task requirements not only provide an overview of the current research efforts in those application areas, but are also indicative of the volume of the research work conducted in those areas. We have also included references to selected articles relevant to different application areas in the visualization domain containing instances of these task requirements. The presentation of task requirements and application areas included in this table is not following any specific ordering strategy.

We have mentioned corresponding application areas in the computer vision domain (Column 4) by reviewing recent CVPR and ICCV program books, and those identified based on discussions with computer vision domain scientists (Column 5) (Section 4). The identification of these application areas provides us with an overview to draw comparisons between two domains, but it is certainly not an exhaustive list of application areas. This analysis helped us identify areas where there is an opportunity for collaborative future research efforts.

Application areas like "Video Editing, Stylization and Painterly Animation" have no direct overlap between the visualization and computer vision domains. In some areas, there is a partial overlap; for instance, the task requirements of visualization in "Medical Applications" include the need for analyzing heterogeneous datasets and interactive analysis of medical datasets, whereas there is more focus on annotation, localization, and segmentation in the task requirements of the computer vision domain.

The comparison of Computer Vision domain experts' application areas and the analysis of corresponding visualization and visual analytics task requirements shows that the data preprocessing task requirements, as mentioned by domain experts, are generally applicable to multiple areas. Also, as expected, the computer vision domain experts' application areas match more closely to the computer vision domain application areas (Table 2).

There are also multiple application areas that overlap in terms of similarity in task requirements in the visualization domain and those identified based on discussions with computer vision domain experts. However, areas like "Deep Learning", "Pattern Analysis, Anomalies", and "3D Modeling and Reconstruction" share more commonalities with the task requirements of the corresponding visualization application areas. Yet, the visualization domain task requirements are more centered toward interaction support, whereas the domain expert task requirements are more abstract and computation focused.

There are commonalities in task requirements of different application areas in the visualization domain. "Sports Analytics", "Surveillance, Activity Analysis and Recognition, Motion Analysis and Tracking", and "Interaction Supported Learning" have many overlaps, such as spatiotemporal

Table 2. Task Requirements Extracted from Surveyed Articles in the Visualization Domain (Related to Image and Video Datasets), and Grouped According to Application Areas in the Visualization Domain

Application Areas (Visualization)		Task Requirements (Visualization)	Application Areas (Computer Vision)	CV Domain Experts' Application Areas
Sports Analytics [75, 109, 142]	S1 S2 S3 S4 S5 S6 S7 S8	Sketch-based search interface User-guided optimizations in models Identify and visualize keyframes (events) in sports video datasets Event detection and classification in video datasets (sports) Support for event annotations Generate summary visualizations for sports video datasets Extract and Augment trajectory data with videos Analyze spatiotemporal patterns and correlations in sports videos	Motion and Tracking; Video Analysis and Understanding; Action and Behavior	Deep Learning; Pattern Analysis, Anomalies
Content Synthesis and Removal [8, 42, 80]	S9 S10	Generate synthetic content for images/videos Content generation by example	Image and Video Synthesis	3D Modeling and Reconstruction
Surveillance, Activity Analysis and Recognition, Motion Analysis and Tracking [60, 95, 122]	S11 S12 S13 S14 S15 S16 S17 S18	Timeline editing of objects and their motion characteristics Support for annotations Motion-data analysis to recognize activities Search and Track objects in videos Spatial context analysis and placement (contextualized videos) Intra- and Inter-cluster analysis based on activity analysis data Identify and search patterns in motion-data sequences Analyze spatiotemporal distributions of data	Surveillance, Person Re-identification; Motion and Tracking; Anomaly Detection in Videos; Action and Behavior	Scene Interpretation, Activity Recognition and Tracking; Deep Learning; Pattern Analysis, Anomalies
Video Editing, Stylization, and Painterly Animation [61, 88, 178]	S19 S20	Painterly animation styles and stroke synthesis (varying attributes) Spatiotemporal varying video painting styles		
Video and Image Collection Analysis [47, 107, 128]	S21 S22 S23 S24	Multimodal content analysis Verbalization and speech pattern analysis in videos Similarities and differences in images and videos Image-set processing	Image Retrieval, Datasets and Evaluation; Video Analysis and Understanding	Large-Scale Data Visualization; Segmentation and Classification
Distance Education and Massive Open Online Courses [25, 26, 51]	S25 S26 S27 S28 S29 S30	Identify regions of interest (peaks) in time-series (e.g., clickstream) Group people based on interaction behavior Correlative analysis among user groups in regions of interests Analyze patterns in time-series datasets Spatiotemporal distribution of user clicks, access locations Identify anomalous regions in time-series dataset	Vision Applications and Systems	
Interaction Supported Learning [55, 57, 91]	S31 S32 S33 S34 S35 S36	Summarization/aggregation of motion sequences, and activities Combining data from multiple heterogeneous sources Interactive exploration in data and feature space Overview + Details-on-Demand, different levels of detail Similarity-based semantic search in action-based videos Multifaceted and multilevel search	Human Machine; Action and Behavior; Surveillance, Person Re-identification; Segmentation, Grouping and Shape, Representation Learning	Scene Interpretation, Activity Recognition and Tracking; Deep Learning; Pattern Analysis, Anomalies; Large-Scale Data Visualization
Video Storyboard and Summarization [95, 109, 148]	S37 S38	Focus+context support for multifield visualization of videos Fast-forward video visualization	Video Analysis and Understanding	Large-Scale Data Visualization
Augmented and Virtual Reality, and Telepresence [83, 96, 183]	S39 S40 S41	Search videos based on similarity in 3D scenes Transformations of âĂŸobjects of interestâĂŹ in videos Mapping remote and locate translations and identify gains	Imaging Geometry, 3D from Single Image, 3D from Multiview and Sensors	3D Modeling and Reconstruction; Support for Semantic Analysis
Video Stereolization [81, 86, 134]	S42 S43	Interaction supported decomposition/modification Conventional videos to stereoscopic	Imaging Geometry, 3D from Single Image, 3D from Multiview and Sensors	3D Modeling and Reconstruction; Support for Semantic Analysis
Image Analysis, Editing, Summarization and Matching [43, 148, 171]	S44 S45	Nearest patch matching to facilitate images similarity Comparative analysis between multiple images	Image Retrieval; Datasets and Evaluation	Large-Scale Data Visualization; Segmentation and Classification
Medical Applications [24, 97, 108]	S46 S47 S48 S49 S50	Medical image analysis Video and Image Annotations Analyzing heterogeneous data Shape detection and analysis in medical datasets Dimensionality reduction and clustering for analysis	Medical, Biological and Cell Microscopy	Deep Learning; Pattern Analysis, Anomalies

We also matched them with the corresponding application areas of the computer vision domain and application areas extracted from domain expert discussions (Section 4). We have also included references to selected articles relevant to different application areas.

pattern analysis, correlative analysis, summarization/aggregation, events and activities analysis, support for annotations, and multilevel/multifaceted search. Similarly, task requirements in "Image Analysis, Editing, Summarization and Matching", "Video and Image Collection Analysis", and "Augmented and Virtual Reality, and Telepresence" areas also share similarities.

3.7 Tools and Libraries

Table 3 provides an overview of the tools and datasets commonly used for the visualization of image and video data in the visualization domain. In the table, we group the tools and datasets

Table 3. Tools and Datasets in the Surveyed Articles in the Visualization Domain, and Grouped based on Application Areas

Application Areas	Tools	Datasets
Sports Analytics	D3, jQuery [112], Palette [131], Node.js (back-end), AngularJS [169], GPU-based Implementation [142], OpenCV (back end), Vue.js (Frontend) [170].	Sports Videos (Rugby) [75], Match Data (Scores) [118], Table Tennis Data [169], Sport Video Dataset [142], Videos Dataset (Soccer Games Videos) [170], Spatio-temporal Tennis Data [117], Soccer dataset [131]
Content Synthesis and Removal	Adobe Effects CS5 [42], WaveNet [165] FLANN Library [8], HED [80], GPU-NVIDIA [175]	Video of Natural Crowd [42], RAVDESS dataset [3], Dyntex Dataset [78], RGB-D datasets [175], large-scale cartoon dataset [80]
Surveillance, Activity Analysis and Recognition, Motion Analysis and Tracking	Direct3D [21], Matlab (Image Processing Toolbox) [95]	Shopping Mall Datasets [21], BEHAVE, AVSS and i-LIDS Dataset [58], Gestures Dataset [60], Video Dataset [160]
Video Editing, Stylization and Painterly Animation	CPU Implementaion (OpenCV), GPU (BSGP) [178], SUMO [13], MontrealForced-Aligner [12], ImageJ [49], Open CV [88]	Video Dataset [178], [88], Traffic dataset [13], B3D(AC) dataset [12], Speech dataset [157], TubeTK dataset [162]
Video and Image Collection Analysis	GPU Implementation [47], JAVA, HTML5, Canvas, JavaScript [128], Face++ Landmarks API, OpenPose Flask, VueJS, D3 [167], VADER Library, Praat, Scikit-learn, Data-GIF [138]	Images Dataset [47], Gene Expression, Retina, Satellite Datasets [128], TED Talks Video, Text and Metadata [167], ImageNet, Places365[107]
Distance Education and Massive Open Online Courses	MongoDB [25], ECharts, Vue, Three JS [51]	Clickstream Video Data [25], ODE School Dataset [51]
Interaction Supported Learning	Custom [57], Flask [92]	Video Datasets, i-Lids Multi-camera Tracking Dataset (Two Video Sequences) [57], VOPE [180], Office-31[92]
Video Storyboard and Summarization	Direct3D [21], Adobe Effects CS5 [42], Matlab (Image Processing Toolbox) [95], D3, jQuery [112]	Shopping Mall Datasets [21], BEHAVE, AVSS and i-LIDS Dataset [58], Vimeo BCCP Dataset [148], Videos of Natural Crowd [42]
Augmented and Virtual Reality, and Telepresence	Iteratively Reweighted Leastsquares (IRLS) Solver on GPU, VoxelHashing [96]	Video Dataset [96], Streaming Videos [183]
Video Stereolization	C++, Matlab [115] and Custom Tools [81, 86]	Middleburry, dinoSparseRing, templeSparseRing Dataset [86], 2D Video Dataset [81]
Image Analysis, Editing, Summarization and Matching	CellProfiler [119], Matlab (GUI) [139], KNIME, SOMFlow, ASHLAR [69] FLANN Library [8], Custom (Python-based) [149], GUCCI [97], cloud computing environment [91], OpenCV, C++ [43], D3.js [33, 184], ImaCytE [140]	Image Datasets [119], Vimeo BCCP Dataset [148], CyCIF dataset [69] Image Dataset [149], Ejecta dataset [164], Cytometry datasets [140]
Medical Applications	Stationary Velocity Fields (SVF) [16], Flask + D3 (Backend: Pandas + MongoDB) [24], Novel 3D Segmentation [44]	Rodent Mandible Dataset [16, 53], Sensors and Videos Dataset [24], PPMI database [173], VC datasets [108], WFM image [44], breast cancer dataset[33]

The intensity of the shaded region corresponds to visualization support in that application area.

used in the surveyed articles (from the major visualization conferences and journals) according to application areas defined in Section 3.4. We have provided a general overview rather than an exhaustive list of tools and datasets in each application area. We also shaded each application area of the table based on the visualization support available for that area, and categorized them

into different classes based on the visualization support level. The support level represents the utilization and availability of interactive visualization tools in that area. We are not following any specific ordering strategy for application areas and tools in this table.

Based on the table, we observe that various standard libraries, such as D3, Angular, Node.js, Vue.js, and JQuery, are used for visualization and visual analytics requirements. Libraries such as Tensorflow and scikit-learn are used for machine-learning tasks. For image and video processing, OpenCV and Matlab are extensively used. For the GPU implementation of algorithms and techniques, OpenGL and Cuda are used. There are also some articles that use custom tools or have not mentioned the tools used; we labeled those articles as custom.

The color shading in Table 3 shows that the areas of Sports Analytics and Video and Image Collection Analysis have better interactive visualization tools support, followed by medical applications. On the other hand, areas like Content Synthesis and Removal, Surveillance, and Activity Analysis are not well supported.

We noticed that researchers mostly used application-oriented image and video datasets [42, 88, 148, 160, 170]. There were also a few datasets based on multiple sources [24].

4 DISCUSSION WITH DOMAIN EXPERTS

The analysis of the visualization research publications related to image and video datasets helped us understand the current research trends in this area. It provided insights into the data and task requirements relevant to different application areas based on different taxonomies. We conducted discussions with computer domain experts to enhance our previous findings based on reviewing visualization research and to obtain a perspective from computer vision domain researchers. This helped us understand how these domain experts use visualization tools in image and video data analysis tasks and provided insights into their needs and requirements, as computer vision domain experts widely use image and video datasets. It also allowed us to understand their current research workflows, identify requirements related to computer vision research problems, learn about the challenges and limitations in their research, and recognize areas in which visual analytics can help advance computer vision research.

As we conducted discussions with only five domain experts, we do not claim that these findings are comprehensive or exhaustive; however, these discussions augment our previous findings, based on a review of visualization research, and help us to collect valuable feedback. These discussions also provided us with an overview of the datasets, visualization tools, algorithms and techniques, APIs and libraries, and computing infrastructure utilized in their research. The computer vision domain experts have at least six years of experience and are scientists working on different professional levels. Their research problems relate to 3D computer vision, 3D modeling and reconstruction, camera calibration and localization, scene interpretation, visual surveillance, activity analysis and recognition, autonomous and self-driving cars, and image classification and manipulation.

These discussions were mainly qualitative, and we took notes of the major findings. Two of the authors of this survey article conducted the discussions, with one predominantly asking questions and the other taking notes. Each discussion was approximately one hour long. Most of the questions were open-ended, and the emphasis was on finding the pain points of using visualization tools for image and video datasets in their respective research workflows. We discussed details of the experts' application areas, their datasets, preprocessing tasks, major algorithms and techniques currently used in their research, the identification of visualization-related requirements, their current use of visualization tools in their workflows, and the APIs and frameworks utilized in their implementations.

In the initial phase, we discussed the details of the datasets currently used in the experts'
research and any general preprocessing steps involved when working with such datasets. The
sizes of these datasets were mainly on the scale of hundreds of **gigabytes** (**GBs**). Some example
datasets currently used by the experts include S3DIS, KITTI, 3D Point Cloud Datasets, Activity
Net, **Total Human Model for Safety** (**THUMS**), YouTube Videos, ImageNet, VarCity, and
Oxford RobotCar. Table 3 contains references to commonly used datasets in different application
areas of visualization domain. While discussing the preprocessing stage of their work, most in-
terviewees mentioned the difficulties involved in data normalization, cleaning, adjusting features,
and exploring the overall characteristics of such datasets. They stated a need for interactive
visualization tools that can both summarize and aggregate these datasets, but also enable them
to make adjustments. One of the domains experts mentioned in the interview: "A visual analytics
tool that can interactively provide insights about the data to facilitate preprocessing, cleaning,
anomalies detection, and so on would be useful for the analysis." Based on these discussions, we
identified the following major requirements related to the preprocessing stage, which are focused
on preparing the data for analysis in the later stages:

Data Preprocessing.

R1 Provide overview of the dataset, with support for denoising, data cleaning, and outliers
detection
R2 Support for data normalization, transformation, standardization, and encoding
R3 Adjustment in features that target all items in the dataset, such as image resizing, cropping,
and rotations
R4 Support for data augmentation and sampling methods

Almost all of the experts mentioned the prevalence and utilization of deep-learning methods
in their research. One of the domain experts stated that more than 90 percent of their current
research work in the domain of computer vision uses some form of deep learning. Major algorithms
or techniques used by these domain experts include CNN, GCN, LSTM, GAN, PCA, and pooling
methods (reducing data-dimensions). Based on our discussions, we identified the following major
requirements grouped into different areas:

Support for Semantic Analysis.

R5 Support for semantic augmentation
R6 Semantic understanding in 3D utilizing semantic classes

Segmentation and Classification.

R7 Annotation support in object segmentation and classification
R8 Labeling points in the point clouds to represent their relationship with the objects

Scene Interpretation, Activity Recognition, and Tracking.

R9 Activity analysis and recognition
R10 Trajectory analysis of moving objects
R11 Illumination changes in images and video datasets

3D Modeling and Reconstruction.

R12 Conversion of video sequences to 3D models
R13 Camera localization and 3D reconstruction

Deep Learning.

R14 Modification of internal structure and architecture of deep learning models
R15 Learn use of algorithms for certain static and dynamic scene scenarios
R16 Increase training samples

In the last phase of the discussion, we mainly focused on how interactive visualizations can support the experts' analysis tasks. They mentioned that using interactive visualizations could help them understand the characteristics of the datasets they utilize in their work and can provide valuable insights that may help fine-tune the deep-learning models used. They expressed the desire to have interactive visualizations that can provide them with an overview of the datasets, with the support to explore details of interesting subsets of the entire dataset (overview and details-on-demand). One of the experts mentioned: "Instead of having a black-box approach where the internals of the algorithms and methods are not visible to the user of the system, a visual analytics system that opens up the black-box and provides insights into the internals of the model could help build trust and confidence in the results of the system." Opening the black-box strategy can help understand the internals of the models and changes in patterns at different stages of learning [92, 101].

They currently use **Visualization Toolkit** (**VTK**) [129], PyVista [147], Meshlab [32], Matplotlib [56], Bokeh [15], TensorBoard [4], TensorFlow Lucid [2], LSTMVis [145], DarkSight [174], Facets [1], GANDissect [10], and NN-SVG [76] to address their visualization needs. Based on our discussion, we identified the following major requirements related to visualization support in their research workflows:

Large-Scale Data Visualization.

R17 Support for visualizing large-scale datasets
R18 Support for data aggregation and summarization
R19 Visualization framework that supports providing overview and details on demand
R20 Support for data navigation, multidimensional querying, spatial and temporal filtering

Pattern Analysis, Anomalies.

R21 Explore and highlight patterns and anomalies in the data
R22 Separation of classes and support for projection
R23 Insights into the deep network architecture, loss, activation filters, and convergence
R24 Support to avoid the manual task of labeling the datasets, noise removal, and adjustments in data items

Multiple Scenarios, Uncertainty.

R25 Uncertainty quantification in visualizations
R26 Visualization of an ensemble of images (e.g., multiple reconstruction scenes)

5   DISCUSSION AND FUTURE DIRECTIONS

Collaborative research efforts at the intersection of visual analytics and computer vision can open new avenues of research and further advance the state of the art in both of these areas. In this survey, we reviewed the research in visual analytics and visualization conferences and journals related to image and video datasets. We also obtained feedback from computer vision domain experts. We found that dealing with image and video data presents a multifaceted and unique set of challenges. These include adapting and scaling the visualization libraries to process large data; supporting interactive designs with support for multiple view visualizations; querying and filtering; sampling, summarization, and aggregation; adapting deep-learning libraries for visualization tasks;

and visualization-specific benchmarks. Below, some of our findings and the existing challenges identified based on this survey are discussed. We also highlight potential future research directions.

5.1 Scalability Issues

In the computer vision domain, there is a large focus on using deep-learning techniques; in fact, more than 90 percent of current research uses these techniques, and domain experts also highlighted this point in our discussions. There are certain challenges associated with advancing the field of visual analytics to bring it in line with computer vision, when it comes to incorporating deep-learning. The tools and libraries in the machine-learning and visual analytics domains are also strongly focused on different domain-specific tasks [40], and there is a need for more collaborative efforts to design libraries and tools so that they can address the needs of domain experts from both fields.

In computer vision, deep learning is often used as a black-box, with little focus on providing insights into the internals of this black-box. On the other hand, visual analytics research intends to open up this box. To be able to trust the outcomes based on deep-learning techniques, users need insights into the decision-making process within this black box, and the rationale behind the outcomes. This builds trust on the outcomes, which is especially important if deep-learning techniques are used in applications of critical nature such as medical applications, public policy making, and law enforcement. There are also limitations associated with the availability of visualization libraries and frameworks that can enable this type of access to deep-learning techniques.

The scale of datasets used in computer vision research is extremely large (usually Gigabytes or Terabytes), and enabling visual analytics for such datasets is challenging as it may also involve integration of big-data frameworks. To provide access to data at multiple levels of detail (e.g., data, model, and features) there is an even greater need for big data frameworks to support interactive exploration.

In visualization research, there is a need to design advanced data processing frameworks that can facilitate visualization at multiple scales and granularity levels. The libraries and frameworks need to adapt and scale up to handle the exponential growth in the size of datasets. This also gives rise to unique issues of data sampling, summarization, querying, interactivity, transformation, and so on. Also, with advancements in high-performance computing technology, the use of big data frameworks and the availability of better visualization libraries for deep learning will bring more focus to the use of these techniques in visual analytics.

5.2 Insights from Surveyed Articles

Tables 1a, 1b, and 1c show the coding results based on the taxonomies of techniques and algorithms, visualization techniques, and application areas. This table can help domain researchers to understand the current trends and focus of research efforts relevant to image and video datasets. It also helps identify areas for future collaborative research. For example, areas relevant to medical applications, image and video analysis, and editing have relatively greater coverage. Below, we discuss more findings and insights based on these tables.

Application Areas. Focusing on the coding tables of visualization and visual analytics articles in Tables 1a, 1b, and 1c, we observe that there has been a greater focus on "Image Analysis, Editing, Summarization, and Matching" compared to other areas. In the "Surveillance, Activity Analysis and Recognition, Motion Analysis and Tracking" application area, there are a very limited number of articles in comparison with the number of articles on current trends in computer vision research, where this application area is one of the actively explored research areas.

Furthermore, there was rapid growth in "Medical Applications" publications in 2021. If we look at the "techniques and algorithms" utilized in these articles, there is an increasing use of deep-learning-based techniques. In medical domain research publications, providing support for visual analysis of individual models or components at different levels of detail is important in order to build confidence in the analysis results. The support for comprehensive interactive visual querying, filtering, and visualizaton at different abstraction levels is a growing need in this area. This helps users explore and understand the rationale behind certain outcomes useful in CAD [37] and medical studies [69, 97]. In "Video Storyboard and Summarization", there have been very few articles published since 2013.

Automated Techniques and Algorithms. In the techniques and algorithms taxonomy (Table 1a, 1b, and 1c), we observe that there was less focus on machine-learning techniques before 2017, even though, by that time, deep learning was one of the most active research areas in computer vision. This trend may be due to the limited availability of visualization libraries related to deep learning, although general purpose libraries for deep learning were available. There is a strong separation in terms of tools and libraries used in both domains [40], but in the future there will be a pressing need for interactive visualization libraries with strong support for deep-learning frameworks due to the increasing popularity of machine learning. In recent years, the number of machine-learning related articles in visualization research has increased.

In Table 1a, 1b, and 1c, the "Other/Custom" category is prevalent in algorithms and techniques taxonomy. In the "Other/Custom" category, the techniques are mostly computer vision related; this category has the most publications, because the articles are related to image and video datasets. There is also an increasing trend in the "Clustering" and "Correlation" category. Deep learning-based frameworks may not be the only solution for image and video data analysis. As seen from our coding table, other/custom techniques are also prevalent, which shows that researchers design many custom or hybrid solutions to tackle image and video datasets. Although the recent trend shows extensive use of standard deep-learning techniques, opening the black-box is one of the pressing needs, as it not only helps explain the decisions but also enables domain experts to integrate domain knowledge during the analysis phase. This also explains the use of semi-automated approaches, extended basic models or hybrid architectures in certain cases.

Visualization Techniques. In visualization techniques taxonomy (Table 1a, 1b, and 1c), 2D techniques are the most commonly used because image and video datasets are often 2D; thus, the visualizations designed for visualizing such datasets, and associated analytics, are also mostly 2D. The video and image datasets and information derived from such datasets are multivariate and complex in nature, so multiple linked views are often used to support their analysis. 3D is consistently being used as a visualization technique; however, its future use is projected to be more a result of advancements in immersive technologies [93]. In the surveyed articles, there is generally less focus on provided interactive support on model explorations and understanding. In the future, there is scope for providing better support for these kinds of interactions.

There is an also increasing trend of using more linked visualizations in a single dashboard. The datasets utilized in such scenarios are multi-dimensional or multi-attribute in nature. The nature of interactive explorations and interactions supported are also dependant on the underlying modeling frameworks utilized in the implementations. For example, deep-learning-based architectures may provide information at different levels (data, model, and features) and the supporting interactive exploration, analysis, and visualization of such information has unique challenges.

Immersive Analytics. There is an increasing trend in the use of immersive environments for analytics [9, 93]. Immersive environments use 3D visualizations and have applications in areas like

medical imaging and gaming. Due to advancements in computer vision in areas such as stereo-scopic vision, 3D scene modeling and reconstruction, and imaging geometry, the role of immersive environments in supporting analytics will become more eminent, thus presenting an opportunity for future work in visual analytics. During our discussions with domain experts, an expert in 3D reconstruction emphasized the need for virtual reality and immersive environments (providing six degrees of freedom) in analytics applications.

Tools and Libraries. The tools and libraries that are used in computer vision research are mostly general purpose libraries that address computer vision task requirements. Their configurations are often complicated and require deep-learning domain expertise. On the other hand, there is limited support for visualization libraries that provide the desired functionality and interfaces to access the deep-learning frameworks. There is a need for collaborative efforts at the intersection of these two areas for the development of tools and libraries that would provide support for visualization tools to be integrated within deep-learning frameworks. With the advancements in the deep-learning domain, high-level usage tools for deep learning, such as extraction of geometry, path, and object, are becoming more common; this will also facilitate advancements in visualization research focused on image and video datasets.

If we compare the tools and libraries in computer vision tools and visualization research (Table 3), the computer vision tools are mostly focused on deep learning, whereas visualization uses a combination of the tools from both domains (e.g., TensorFlow, PyTorch, D3, and Node.js).

Most visualization tools are designed to solve a particular problem or set of tasks, resulting in tools that are highly customized and not inter-operable. There is a need for visualization frameworks where multiple tools can be used to solve challenging problems. There are many application areas, and each has a unique set of requirements and challenges. Algorithms and techniques utilized in these areas also vary considerably. Studies should be conducted that provide clear guidelines for problem or application area-focused general frameworks that can also leverage the developments in other areas, such as technological advancements in deep-learning frameworks in computer vision.

Datasets. In computer vision research, there is a focus on creating benchmark datasets. In our discussion with domain experts, they also mentioned that their focus is more on the preprocessing stage or the stage after the model training is complete in their typical research workflows dealing with dataset; their focus is not on providing visual analytics support. We also observed this when we assessed the computer vision articles included in this survey. In the visualization and visual analytics research, the datasets are more application oriented (Table 3), and they are usually in a form that enables interactive linked visualizations and utilizes the derived data, in addition to the image and video datasets.

The datasets usually used in the surveyed articles in the visualization domain are mostly derived from image and video in a semi-supervised manner. In the future, due to advancements in machine-learning techniques, we foresee that there will be more focus on the unsupervised generation of these datasets, resulting in an increase in the data scale, and an even higher demand to visualize this data.

Most datasets in our surveyed articles were either collected or generated by the authors, and are not standard datasets publicly available (demonstrated clearly in Table 3). In the computer vision domain, it is common to use standard data sets, which are publicly available; in the future, some work must be done to make various standard datasets publicly available for bench-marking and comparison.

Evaluation. We reviewed the evaluation methods used in the surveyed articles. One common method was to evaluate the performance of the algorithms and techniques used [75, 88, 183]. In

some articles, studies with domain experts were conducted to evaluate the visualization techniques [67, 142, 170]. Quantitative and qualitative evaluations were also conducted with real users [58, 151, 167]. Overall, there was less focus on evaluating the design tool in terms of perception and cognition of computer vision domain scientists. In the future, more studies are needed to evaluate how interactive visualization can better cater to the needs of computer vision domain experts.

Benchmark Tasks and Studies. Benchmark tasks are quite common in the computer vision domain for comparing any model with the state-of-the-art model on standard tasks such as data processing, querying, and inference. There are no benchmark tasks available in the visualization domain for the use of machine learning or deep learning models. This presents an opportunity to design benchmark tasks for the benchmarking of machine learning models used in the visualization domain.

Furthermore, no studies exist to provide guidelines for the designer about users' perception and cognition at the intersection of the computer vision and visualization domains. There is an opportunity to design such guidelines, that will help in designing better tools for image and video data analysis to improve ease of use and understanding for users.

5.3   ML Models not Designed for Visualization Domain

Machine learning models used in visualization applications are usually designed for computer vision problems. They are not tailored toward visualization use. More work is needed to effectively tailor these models to be used for visualization and visual analytics applications. In usual integration scenarios, machine learning models are considered as a black-box. On the other hand, visual analytics models are interactive with a focus on the human in the loop. The "expert in the loop" systems are designed to utilize human domain knowledge in the interactive analysis phase [69]. The routine tasks or automated approaches can be configured to trigger as a result of specific selections or interactions. Future research should evaluate how supervised and unsupervised models can be optimized by adding "human in the loop" and to interactively steer the automated algorithms.

Understanding the model behavior and performance across a wide range of input data and scenarios is important before integration into visual analytics systems. Tools like What-if [166] enable model understanding through interactive probing with minimal coding. These are open challenges and present opportunities for collaborative efforts to address these concerns. Understanding the transfer learning process while adapting models for new inputs [92] is particularly important, especially in the context of "expert in the loop" in visual analytics systems. There has been limited work in this area of visual analytics research focused on image and video datasets.

Semi-supervised and Reinforcement learning models were rare in our surveyed articles. Again, these models are not tailored toward visualization use, and there is scope for collaborative efforts to optimize their design taking into consideration the requirements of visual analytics. Other more complex and advanced machine learning techniques, such as federated learning and transform learning, are rarely used for image and video data, but we foresee this to change in the future.

5.4   Insights from Discussions with Domain Experts

During our discussions with computer vision domain experts, almost all of the experts emphasized that there is a need for visualization tools that can support interactive exploration of datasets with features like "Overview + details-on-demand". This would help them better understand their data and apply any normalization or standardization techniques to improve the data quality before they move on to the training phase (data preprocessing). They also mentioned the limited availability of visualization tools that provide information about the internals of deep-learning networks.

Currently, they mostly use it as a black-box, without gaining insights about what is going on inside the box. They also mentioned the need for interactive tools that provide information about the convergence of models during the training phase, support for interactively tuning parameters, and interactive visualization guided optimizations.

6   LIMITATIONS

Here, we discuss some of the limitations of our work. Discussions with computer vision domain experts enabled us to gather valuable insights and more practical details involved in deep-learning based implementations relevant to image and video datasets. These discussions were not exhaustive, as they were restricted to only five domain experts in computer vision. Information on the datasets, processes, algorithms and techniques, computing infrastructure, and details about the implementation and technical difficulties were occasionally not directly available from the articles. However, the discussions helped us gather this information from the domain experts.

While searching for visualization and visual analytics articles, we only focused on major visualization related conferences. There may be further relevant articles in some other related areas such as big data, high-performance computing, machine learning, and parallel computing, but each of these areas have technical issues of their own, and their inclusion would have excessively expanded the scope of our work.

In the future, we plan to build an interactive recommendation tool to explore articles, and related details included in this survey, and provide recommendations in terms of visualization techniques, automated algorithms, tools and datasets, while keeping in view the task requirements.

7   CONCLUSION

We presented a comprehensive survey of the state of the art in visualization and visual analytics research related to image and video datasets. We described recent advances at the intersection of computer vision and visualization to facilitate areas of visualization research, including different visualization tools, techniques, and solutions across different application areas. Overall, this also helped us to identify gaps and opportunities for future collaborative research. Discussions with experts working in the computer vision domain allowed us to identify the requirements for interactive visualizations of image and video datasets. After collection of relevant articles from the major visualization conferences and journals, we categorized the algorithms, visualization techniques, application areas, and task requirements used in these articles based on standard taxonomies. We summarized information gathered from publications in the visualization domain along with discussions with computer vision domain scientists into a systematic tabular format that facilitates comparisons and finding opportunities for future research. Lastly, we provided a detailed discussion about the insights found based on our survey results. We also presented current gaps in research, and associated limitations and challenges.

ACKNOWLEDGMENTS

REFERENCES

[1] 2020. Facets: An Open Source Visualization Tool for Machine Learning Training Data. Retrieved September 17, 2020 from https://github.com/pair-code/facets.

[2] 2020. TensorFlow Lucid. Retrieved September 17, 2020 from https://github.com/tensorflow/lucid.

[3] H. Zeng, X. Wang, A. Wu, Y. Wang, Q. Li, A. Endert, and H. Qu. 2019. EmoCo: Visual analysis of emotion coherence in presentation videos. *IEEE Transactions on Visualization and Computer Graphics* 1 (2019), 1–1. DOI : https://doi.org/10.1109/TVCG.2019.2934656

[4] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dandelion Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. 2015. TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems. Retrieved from https://www.tensorflow.org/. Software available from tensorflow.org. Accessed December 1, 2022.

[5] S. Afzal, M. M. Hittawe, S. Ghani, T. Jamil, O. Knio, M. Hadwiger, and I. Hoteit. 2019. The state of the art in visual analysis approaches for ocean and atmospheric datasets. *Computer Graphics Forum* 38, 3 (2019), 881–907. DOI : https://doi.org/10.1111/cgf.13731 Accessed December 1, 2022.

[6] Fereshteh Amini, Nathalie Henry Riche, Bongshin Lee, Andres Monroy-Hernandez, and Pourang Irani. 2016. Authoring data-driven videos with dataclips. *IEEE Transactions on Visualization and Computer Graphics* 23, 1 (2016), 501–510.

[7] Gennady Andrienko, Natalia Andrienko, Gabriel Anzer, Pascal Bauer, Guido Budziak, Georg Fuchs, Dirk Hecker, Hendrik Weber, and Stefan Wrobel. 2019. Constructing spaces and times for tactical analysis in football. *IEEE Transactions on Visualization and Computer Graphics* 27, 4 (2019), 2280–2297.

[8] A. Baldacci, F. Ganovelli, M. Corsini, and R. Scopigno. 2017. Presentation of 3D scenes through video example. *IEEE Transactions on Visualization and Computer Graphics* 23, 9 (2017), 2096–2107. DOI : https://doi.org/10.1109/TVCG.2016.2608828

[9] A. Batch, A. Cunningham, M. Cordeil, N. Elmqvist, T. Dwyer, B. H. Thomas, and K. Marriott. 2020. There Is no spoon: Evaluating performance, space use, and presence with expert domain users in immersive analytics. *IEEE Transactions on Visualization and Computer Graphics* 26, 1 (2020), 536–546. DOI : https://doi.org/10.1109/TVCG.2019.2934803

[10] David Bau, Jun-Yan Zhu, Hendrik Strobelt, Bolei Zhou, Joshua B. Tenenbaum, William T. Freeman, and Antonio Torralba. 2019. GAN dissection: Visualizing and understanding generative adversarial networks. In *Proceedings of the International Conference on Learning Representations*.

[11] M. Behrisch, B. Bach, M. Hund, M. Delz, L. Von Rüden, J. Fekete, and T. Schreck. 2017. Magnostics: Image-based search of interesting matrix views for guided network exploration. *IEEE Transactions on Visualization and Computer Graphics* 23, 1 (2017), 31–40. DOI : https://doi.org/10.1109/TVCG.2016.2598467

[12] Eloïse Berson, Catherine Soladié, and Nicolas Stoiber. 2020. Intuitive facial animation editing based on a generative RNN framework. 39, 8 (2020), 241–251.

[13] Huikun Bi, Tianlu Mao, Zhaoqi Wang, and Zhigang Deng. 2020. A deep learning-based framework for intersectional traffic simulation and editing. *IEEE Transactions on Visualization and Computer Graphics* 26, 7 (2020), 2335–2348. DOI : https://doi.org/10.1109/TVCG.2018.2889834

[14] Ingmar Bitter, Robert Van Uitert, Ivo Wolf, Luis Ibanez, and Jan-Martin Kuhnigk. 2007. Comparison of four freely available frameworks for image processing and visualization that use ITK. *IEEE Transactions on Visualization and Computer Graphics* 13, 3 (2007), 483–493.

[15] Bokeh Development Team. 2020. *Bokeh: Python Library for Interactive Visualization*. Retrieved from https://bokeh.org/. Accessed December 1, 2022.

[16] Saeed Boorboor, Shreeraj Jadhav, Mala Ananth, David Talmage, Lorna Role, and Arie Kaufman. 2018. Visualization of neuronal structures in wide-field microscopy brain images. *IEEE Transactions on Visualization and Computer Graphics* 25, 1 (2018), 1018–1028.

[17] Paulo Vinicius Koerich Borges, Nicola Conci, and Andrea Cavallaro. 2013. Video-based human behavior understanding: A survey. *IEEE Transactions on Circuits and Systems for Video Technology* 23, 11 (2013), 1993–2008.

[18] Rita Borgo, Min Chen, Ben Daubney, Edward Grundy, Gunther Heidemann, Benjamin Höferlin, Markus Höferlin, Heike Jänicke, Daniel Weiskopf, and Xianghua Xie. 2011. A survey on video-based graphics and video visualization. In *Proceedings of the Eurographics 2011 - State of the Art Reports*. N. John and B. Wyvill (Eds.), The Eurographics Association. DOI : https://doi.org/10.2312/EG2011/stars/001-023

[19] Rita Borgo, Min Chen, Ben Daubney, Edward Grundy, Gunther Heidemann, Benjamin Höferlin, Markus Höferlin, Heike Leitte, Daniel Weiskopf, and Xianghua Xie. 2012. State of the art report on video-based graphics and video visualization. In *Proceedings of the Computer Graphics Forum*. Wiley Online Library, 2450–2477.

[20] R. A. Borsoi and G. H. Costa. 2018. On the performance and implementation of parallax free video see-through displays. *IEEE Transactions on Visualization and Computer Graphics* 24, 6 (2018), 2011–2022. DOI : https://doi.org/10.1109/TVCG.2017.2705184

[21] R. P. Botchen, S. Bachthaler, F. Schick, M. Chen, G. Mori, D. Weiskopf, and T. Ertl. 2008. Action-based multifield video visualization. *IEEE Transactions on Visualization and Computer Graphics* 14, 4 (2008), 885–899. DOI : https://doi.org/10.1109/TVCG.2008.40

[22] Brian Bowman, Niklas Elmqvist, and T. J. Jankun-Kelly. 2012. Toward visualization for games: Theory, design space, and patterns. *IEEE Transactions on Visualization and Computer Graphics* 18, 11 (2012), 1956–1968.

[23] Chris Bryan, Kwan-Liu Ma, and Jonathan Woodring. 2016. Temporal summary images: An approach to narrative visualization via interactive annotation generation and placement. *IEEE Transactions on Visualization and Computer Graphics* 23, 1 (2016), 511–520.

[24] G. Y. Chan, L. G. Nonato, A. Chu, P. Raghavan, V. Aluru, and C. T. Silva. 2019. Motion browser: Visualizing and understanding complex upper limb movement under obstetrical brachial plexus injuries. *IEEE Transactions on Visualization and Computer Graphics* 26, 1 (2019), 1–1. DOI : https://doi.org/10.1109/TVCG.2019.2934280

[25] Q. Chen, Y. Chen, D. Liu, C. Shi, Y. Wu, and H. Qu. 2016. PeakVizor: Visual analytics of peaks in video clickstreams from massive open online courses. *IEEE Transactions on Visualization and Computer Graphics* 22, 10 (2016), 2315–2330. DOI : https://doi.org/10.1109/TVCG.2015.2505305

[26] Qing Chen, Xuanwu Yue, Xavier Plantaz, Yuanzhe Chen, Conglei Shi, Ting-Chuen Pong, and Huamin Qu. 2020. Viseq: Visual analytics of learning sequence in massive open online courses. *IEEE Transactions on Visualization and Computer Graphics* 26, 3 (2020), 1622–1636.

[27] Xin Chen, Yuwei Li, Xi Luo, Tianjia Shao, Jingyi Yu, Kun Zhou, and Youyi Zheng. 2020. Recovering 3D editable objects from a single photograph. *IEEE Transactions on Visualization and Computer Graphics* 26, 3 (2020), 1466–1475. DOI : https://doi.org/10.1109/TVCG.2018.2871190

[28] J. Choo and S. Liu. 2018. Visual analytics for explainable deep learning. *IEEE Computer Graphics and Applications* 38, 4 (2018), 84–92.

[29] David H. S. Chung, Philip A. Legg, Matthew L. Parry, Rhodri Bown, Iwan W. Griffiths, Robert S. Laramee, and Min Chen. 2015. Glyph sorting: Interactive visualization for multi-dimensional data. *Information Visualization* 14, 1 (2015), 76–90. https://doi.org/10.1177/1473871613511959

[30] David H.S. Chung, Matthew L. Parry, Iwan W. Griffiths, Robert S. Laramee, Rhodri Bown, Philip A. Legg, and Min Chen. 2016. Knowledge-assisted ranking: A visual analytic application for sports event data. *IEEE Computer Graphics and Applications* 36, 3 (2016), 72–82. DOI : https://doi.org/10.1109/MCG.2015.25

[31] David H. Chung, Matthew L. Parry, Philip A. Legg, Iwan W. Griffiths, Robert S. Laramee, and Min Chen. 2012. Visualizing multiple error-sensitivity fields for single camera positioning. *Computing and Visualization in Science* 15, 6 (2012), 303–317. DOI : https://doi.org/10.1007/s00791-014-0216-9

[32] Paolo Cignoni, Marco Callieri, Massimiliano Corsini, Matteo Dellepiane, Fabio Ganovelli, and Guido Ranzuglia. 2008. MeshLab: An open-source mesh processing tool. In *Proceedings of the Eurographics Italian Chapter Conference*. Vittorio Scarano, Rosario De Chiara, and Ugo Erra (Eds.), The Eurographics Association. DOI : https://doi.org/10.2312/LocalChapterEvents/ItalChap/ItalianChapConf2008/129-136

[33] A. Corvò, H. S. Garcia Caballero, M. A. Westenberg, M. A. van Driel, and J. J. van Wijk. 2021. Visual analytics for hypothesis-driven exploration in computational pathology. *IEEE Transactions on Visualization and Computer Graphics* 27, 10 (2021), 3851–3866. DOI : https://doi.org/10.1109/TVCG.2020.2990336

[34] Lhaylla Crissaff, Louisa Wood Ruby, Samantha Deutch, R. Luke DuBois, Jean-Daniel Fekete, Juliana Freire, and Claudio Silva. 2017. ARIES: Enabling visual exploration and organization of art image collections. *IEEE Computer Graphics and Applications* 38, 1 (2017), 91–108.

[35] Saverio Debernardis, Michele Fiorentino, Michele Gattullo, Giuseppe Monno, and Antonio Emmanuele Uva. 2013. Text readability in head-worn displays: Color and style optimization in video versus optical see-through devices. *IEEE Transactions on Visualization and Computer Graphics* 20, 1 (2013), 125–139.

[36] Chandni J. Dhamsania and Tushar V. Ratanpara. 2016. A survey on human action recognition from videos. In *Proceedings of the 2016 Online International Conference on Green Engineering and Technologies*. IEEE, 1–5.

[37] Konstantin Dmitriev, Joseph Marino, Kevin Baker, and Arie E. Kaufman. 2021. Visual analytics of a computer-aided diagnosis system for pancreatic lesions. *IEEE Transactions on Visualization and Computer Graphics* 27, 3 (2021), 2174–2185. DOI : https://doi.org/10.1109/TVCG.2019.2947037

[38] John J. Dudley and Per Ola Kristensson. 2018. A review of user interface design for interactive machine learning. *ACM Transactions on Interactive Intelligent Systems* 8, 2 (2018), 37 pages. DOI : https://doi.org/10.1145/3185517

[39] B. Duffy, J. Thiyagalingam, S. Walton, D. J. Smith, A. Trefethen, J. C. Kirkman-Brown, E. A. Gaffney, and M. Chen. 2015. Glyph-based video visualization for semen analysis. *IEEE Transactions on Visualization and Computer Graphics* 21, 8 (2015), 980–993. DOI : https://doi.org/10.1109/TVCG.2013.265

[40] Alex Endert, William Ribarsky, Cagatay Turkay, B. L. William Wong, Ian Nabney, I. Díaz Blanco, and Fabrice Rossi. 2017. The state of the art in integrating machine learning into visual analytics. In *Proceedings of the Computer Graphics Forum*. Wiley Online Library, 458–486.

[41] Mingming Fan, Ke Wu, Jian Zhao, Yue Li, Winter Wei, and Khai N. Truong. 2019. VisTA: Integrating machine intelligence with visualization to support the investigation of think-aloud sessions. *IEEE Transactions on Visualization and Computer Graphics* 26, 1 (2019), 343–352.

[42] M. Flagg and J. M. Rehg. 2013. Video-based crowd synthesis. *IEEE Transactions on Visualization and Computer Graphics* 19, 11 (2013), 1935–1947. DOI:https://doi.org/10.1109/TVCG.2012.317

[43] Jiayun Fu, Bin Zhu, Weiwei Cui, Song Ge, Yun Wang, Haidong Zhang, He Huang, Yuanyuan Tang, Dongmei Zhang, and Xiaojing Ma. 2021. Chartem: Reviving chart images with data embedding. *IEEE Transactions on Visualization and Computer Graphics* 27, 2 (2021), 337–346. DOI:https://doi.org/10.1109/TVCG.2020.3030351

[44] Parmida Ghahremani, Saeed Boorboor, Pooya Mirhosseini, Chetan Gudisagar, Mala Ananth, David Talmage, Lorna W. Role, and Arie E. Kaufman. 2021. NeuroConstruct: 3D reconstruction and visualization of neurites in optical microscopy brain images. *IEEE Transactions on Visualization and Computer Graphics* 28, 12 (2021), 1–1. DOI:https://doi.org/10.1109/TVCG.2021.3109460

[45] A. Gilbert, M. Trumble, A. Hilton, and J. Collomosse. 2018. Inpainting of wide-baseline multiple viewpoint video. *IEEE Transactions on Visualization and Computer Graphics* 26, 7 (2018), 1–1. DOI:https://doi.org/10.1109/TVCG.2018.2889297

[46] Ievgeniia Gutenko, Konstantin Dmitriev, Arie E. Kaufman, and Matthew A. Barish. 2016. AnaFe: Visual analytics of image-derived temporal features-focusing on the spleen. *IEEE Transactions on Visualization and Computer Graphics* 23, 1 (2016), 171–180.

[47] L. K. Ha, J. Kruger, J. L. D. Comba, C. T. Silva, and S. Joshi. 2012. ISP: An optimal out-of-core image-set processing streaming architecture for parallel heterogeneous systems. *IEEE Transactions on Visualization and Computer Graphics* 18, 6 (2012), 838–851. DOI:https://doi.org/10.1109/TVCG.2012.32

[48] Markus Hadwiger, Ronell Sicat, Johanna Beyer, Jens Krüger, and Torsten Möller. 2012. Sparse PDF maps for nonlinear multi-resolution image operations. *ACM Transactions on Graphics* 31, 6 (2012), 12 pages. DOI:https://doi.org/10.1145/2366145.2366152

[49] Gaudenz Halter, Rafael Ballester-Ripoll, Barbara Flueckiger, and Renato Pajarola. 2019. VIAN: A visual annotation tool for film analysis. In *Proceedings of the Computer Graphics Forum*. Wiley Online Library, 119–129.

[50] Adam W. Harley. 2015. An interactive node-link visualization of convolutional neural networks. In *Proceedings of the Advances in Visual Computing*. George Bebis, Richard Boyle, Bahram Parvin, Darko Koracin, Ioannis Pavlidis, Rogerio Feris, Tim McGraw, Mark Elendt, Regis Kopper, Eric Ragan, Zhao Ye, and Gunther Weber (Eds.), Springer International Publishing, Cham, 867–877.

[51] H. He, O. Zheng, and B. Dong. 2018. VUSphere: Visual analysis of video utilization in online distance education. In *Proceedings of the 2018 IEEE Conference on Visual Analytics Science and Technology*. 25–35. DOI:https://doi.org/10.1109/VAST.2018.8802383

[52] J. Herling and W. Broll. 2014. High-quality real-time video inpaintingwith PixMix. *IEEE Transactions on Visualization and Computer Graphics* 20, 6 (2014), 866–879. DOI:https://doi.org/10.1109/TVCG.2014.2298016

[53] M. Hermann, A. C. Schunke, T. Schultz, and R. Klein. 2016. Accurate interactive visualization of large deformations and variability in biomedical image ensembles. *IEEE Transactions on Visualization and Computer Graphics* 22, 1 (2016), 708–717. DOI:https://doi.org/10.1109/TVCG.2015.2467198

[54] F. Hohman, M. Kahng, R. Pienta, and D. H. Chau. 2019. Visual analytics in deep learning: An interrogative survey for the next frontiers. *IEEE Transactions on Visualization and Computer Graphics* 25, 8 (2019), 2674–2693.

[55] Xinyi Huang, Suphanut Jamonnak, Ye Zhao, Boyu Wang, Minh Hoai, Kevin Yager, and Wei Xu. 2021. Interactive visual study of multiple attributes learning model of x-ray scattering images. *IEEE Transactions on Visualization and Computer Graphics* 27, 2 (2021), 1312–1321. DOI:https://doi.org/10.1109/TVCG.2020.3030384

[56] J. D. Hunter. 2007. Matplotlib: A 2D graphics environment. *Computing in Science & Engineering* 9, 3 (2007), 90–95. DOI:https://doi.org/10.1109/MCSE.2007.55

[57] B. Höferlin, R. Netzel, M. Höferlin, D. Weiskopf, and G. Heidemann. 2012. Inter-active learning of ad-hoc classifiers for video visual analytics. In *Proceedings of the 2012 IEEE Conference on Visual Analytics Science and Technology*. 23–32. DOI:https://doi.org/10.1109/VAST.2012.6400492

[58] M. Höferlin, K. Kurzhals, B. Höferlin, G. Heidemann, and D. Weiskopf. 2012. Evaluation of fast-forward video visualization. *IEEE Transactions on Visualization and Computer Graphics* 18, 12 (2012), 2095–2103. DOI:https://doi.org/10.1109/TVCG.2012.222

[59] M. Itoh, M. Toyoda, T. Kamijo, and M. Kitsuregawa. 2012. Visualizing flows of images in social media. In *Proceedings of the 2012 IEEE Conference on Visual Analytics Science and Technology*. 229–230. DOI:https://doi.org/10.1109/VAST.2012.6400539

[60] Sujin Jang, Niklas Elmqvist, and Karthik Ramani. 2015. Motionflow: Visual abstraction and aggregation of sequential patterns in human motion tracking data. *IEEE Transactions on Visualization and Computer Graphics* 22, 1 (2015), 21–30.

[61] M. Kagaya, W. Brendel, Q. Deng, T. Kesterson, S. Todorovic, P. J. Neill, and E. Zhang. 2011. Video painting with space-time-varying style parameters. *IEEE Transactions on Visualization and Computer Graphics* 17, 1 (2011), 74–87. DOI:https://doi.org/10.1109/TVCG.2010.25

[62] M. Kahng, P. Y. Andrews, A. Kalro, and D. H. Chau. 2018. ActiVis: Visual exploration of industry-scale deep neural network models. *IEEE Transactions on Visualization and Computer Graphics* 24, 1 (2018), 88–97.

[63] Tapas Kanungo, David M. Mount, Nathan S. Netanyahu, Christine D. Piatko, Ruth Silverman, and Angela Y. Wu. 2002. An efficient k-means clustering algorithm: Analysis and implementation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24, 7 (2002), 881–892.

[64] Daniel Keim. 2002. Information visualization and visual data mining. *IEEE Transactions on Visualization and Computer Graphics* 8, 1 (2002), 1–8.

[65] Rajat Khurana and Alok KumarSingh Kushwaha. 2018. A deep survey on human activity recognition in video surveillance. In *Proceedings of the 2018 International Conference on Research in Intelligent and Computing in Engineering* . IEEE, 1–5.

[66] Gordon Kindlmann, Charisee Chiw, Nicholas Seltzer, Lamont Samuels, and John Reppy. 2015. Diderot: A domain-specific language for portable parallel scientific visualization and image analysis. *IEEE Transactions on Visualization and Computer Graphics* 22, 1 (2015), 867–876.

[67] P. Klemm, S. Oeltze-Jafra, K. Lawonn, K. Hegenscheid, H. Völzke, and B. Preim. 2014. Interactive visual analysis of image-centric cohort study data. *IEEE Transactions on Visualization and Computer Graphics* 20, 12 (2014), 1673–1682. DOI : https://doi.org/10.1109/TVCG.2014.2346591

[68] S. Ko, I. Cho, S. Afzal, C. Yau, J. Chae, A. Malik, K. Beck, Y. Jang, W. Ribarsky, and D. S. Ebert. 2016. A survey on visual analysis approaches for financial data. *Computer Graphics Forum* 35, 3 (2016), 599–617.

[69] Robert Krueger, Johanna Beyer, Won-Dong Jang, Nam Wook Kim, Artem Sokolov, Peter K. Sorger, and Hanspeter Pfister. 2020. Facetto: Combining unsupervised and supervised learning for hierarchical phenotype analysis in multi-channel image data. *IEEE Transactions on Visualization and Computer Graphics* 26, 1 (2020), 227–237. DOI : https://doi.org/10.1109/TVCG.2019.2934547

[70] Kuno Kurzhals, Marcel Hlawatsch, Florian Heimerl, Michael Burch, Thomas Ertl, and Daniel Weiskopf. 2015. Gaze stripes: Image-based visualization of eye tracking data. *IEEE Transactions on Visualization and Computer Graphics* 22, 1 (2015), 1005–1014.

[71] J. E. Kyprianidis, J. Collomosse, T. Wang, and T. Isenberg. 2013. State of the art: A taxonomy of artistic stylization techniques for images and video. *IEEE Transactions on Visualization and Computer Graphics* 19, 5 (2013), 866–885. DOI : https://doi.org/10.1109/TVCG.2012.160

[72] W. Lai, Y. Huang, N. Joshi, C. Buehler, M. Yang, and S. B. Kang. 2018. Semantic-driven generation of hyperlapse from 360 degree video. *IEEE Transactions on Visualization and Computer Graphics* 24, 9 (2018), 2610–2621. DOI : https://doi.org/10.1109/TVCG.2017.2750671

[73] Shuyue Lan, Rameswar Panda, Qi Zhu, and Amit K. Roy-Chowdhury. 2018. FFNet: Video fast-forwarding via reinforcement learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.* 6771–6780.

[74] Yann LeCun, Y. Bengio, and Geoffrey Hinton. 2015. Deep learning. *Nature* 521, 05 (2015), 436–44. DOI : https://doi.org/10.1038/nature14539

[75] P. A. Legg, D. H. S. Chung, M. L. Parry, R. Bown, M. W. Jones, I. W. Griffiths, and M. Chen. 2013. Transformation of an uncertain video search pipeline to a sketch-based visual analytics loop. *IEEE Transactions on Visualization and Computer Graphics* 19, 12 (2013), 2109–2118. DOI : https://doi.org/10.1109/TVCG.2013.207

[76] Alexander LeNail. 2019. NN-SVG: Publication-ready neural network architecture schematics. *Journal of Open Source Software* 4, 33 (2019), 747. DOI : https://doi.org/10.21105/joss.00747

[77] Lucie Lévêque, Hilde Bosmans, Lesley Cockmartin, and Hantao Liu. 2018. State of the art: Eye-tracking studies in medical imaging. *IEEE Access* 6 (2018), 37023–37034.

[78] C. Li, D. Pickup, T. Saunders, D. Cosker, D. Marshall, P. Hall, and P. Willis. 2013. Water surface modeling from a single viewpoint video. *IEEE Transactions on Visualization and Computer Graphics* 19, 7 (2013), 1242–1251. DOI : https://doi.org/10.1109/TVCG.2012.302

[79] Sheng Li, Zhiqiang Tao, Kang Li, and Yun Fu. 2019. Visual to text: Survey of image and video captioning. *IEEE Transactions on Emerging Topics in Computational Intelligence* 3, 4 (2019), 297–312.

[80] Xiaoyu Li, Bo Zhang, Jing Liao, and Pedro Sander. 2021. Deep sketch-guided cartoon video inbetweening. *IEEE Transactions on Visualization and Computer Graphics* 28, 8 (2021), 1–1. DOI : https://doi.org/10.1109/TVCG.2021.3049419

[81] M. Liao, J. Gao, R. Yang, and M. Gong. 2012. Video stereolization: Combining motion analysis with user interaction. *IEEE Transactions on Visualization and Computer Graphics* 18, 7 (2012), 1079–1088. DOI : https://doi.org/10.1109/TVCG.2011.114

[82] I-Chen Lin, Yu-Chien Lan, and Po-Wen Cheng. 2015. SI-Cut: Structural inconsistency analysis for image foreground extraction. *IEEE Transactions on Visualization and Computer Graphics* 21, 7 (2015), 860–872.

[83] S. Lin, C. Lin, I. Yeh, S. Chang, C. Yeh, and T. Lee. 2013. Content-aware video retargeting using object-preserving warping. *IEEE Transactions on Visualization and Computer Graphics* 19, 10 (2013), 1677–1686. DOI : https://doi.org/10.1109/TVCG.2013.75

[84] Honghai Liu, Shengyong Chen, and Naoyuki Kubota. 2013. Intelligent video systems and analytics: A survey. *IEEE Transactions on Industrial Informatics* 9, 3 (2013), 1222–1233.

[85] Shixia Liu, Xiting Wang, Mengchen Liu, and Jun Zhu. 2017. Towards better analysis of machine learning models: A visual analytics perspective. *Visual Informatics* 1, 1 (2017), 48 – 56. DOI : https://doi.org/10.1016/j.visinf.2017.01.006

[86] Y. Liu, Q. Dai, and W. Xu. 2010. A point-cloud-based multiview stereo algorithm for free-viewpoint video. *IEEE Transactions on Visualization and Computer Graphics* 16, 3 (2010), 407–418. DOI : https://doi.org/10.1109/TVCG.2009.88

[87] María-Jesús Lobo, Caroline Appert, and Emmanuel Pietriga. 2018. Animation plans for before-and-after satellite images. *IEEE Transactions on Visualization and Computer Graphics* 25, 2 (2018), 1347–1360.

[88] C. Lu, Y. Xiao, and C. Tang. 2018. Real-time video stylization using object flows. *IEEE Transactions on Visualization and Computer Graphics* 24, 6 (2018), 2051–2063. DOI : https://doi.org/10.1109/TVCG.2017.2700470

[89] S. Lu, S. Zhang, J. Wei, S. Hu, and R. R. Martin. 2013. Timeline editing of objects in video. *IEEE Transactions on Visualization and Computer Graphics* 19, 7 (2013), 1218–1227. DOI : https://doi.org/10.1109/TVCG.2012.145

[90] Yafeng Lu, Rolando Garcia, Brett Hansen, Michael Gleicher, and Ross Maciejewski. 2017. The state-of-the-art in predictive visual analytics. *Computer Graphics Forum* 36, 3 (2017), 539–562. DOI : https://doi.org/10.1111/cgf.13210

[91] Ruixian Ma, Honghui Mei, Huihua Guan, Wei Huang, Fan Zhang, Chengye Xin, Wenzhuo Dai, Xiao Wen, and Wei Chen. 2021. LADV: Deep learning assisted authoring of dashboard visualizations from images and sketches. *IEEE Transactions on Visualization and Computer Graphics* 27, 9 (2021), 3717–3732. DOI : https://doi.org/10.1109/TVCG.2020.2980227

[92] Yuxin Ma, Arlen Fan, Jingrui He, Arun Reddy Nelakurthi, and Ross Maciejewski. 2021. A visual analytics framework for explaining and diagnosing transfer learning processes. *IEEE Transactions on Visualization and Computer Graphics* 27, 2 (2021), 1385–1395. DOI : https://doi.org/10.1109/TVCG.2020.3028888

[93] Kim Marriott, Jian Chen, Marcel Hlawatsch, Takayuki Itoh, Miguel A. Nacenta, Guido Reina, and Wolfgang Stuerzlinger. 2018. *Immersive Analytics: Time to Reconsider the Value of 3D for Information Visualisation.* Springer International Publishing, Cham, 25–55.

[94] Leland McInnes, John Healy, Nathaniel Saul, and Lukas Großberger. 2018. UMAP: Uniform Manifold approximation and projection. *Journal of Open Source Software* 3, 29 (2018), 861. DOI : https://doi.org/10.21105/joss.00861

[95] A. H. Meghdadi and P. Irani. 2013. Interactive exploration of surveillance video through action shot summarization and trajectory visualization. *IEEE Transactions on Visualization and Computer Graphics* 19, 12 (2013), 2119–2128. DOI : https://doi.org/10.1109/TVCG.2013.168

[96] A. Meka, G. Fox, M. Zollhöfer, C. Richardt, and C. Theobalt. 2017. Live user-guided intrinsic video for static scenes. *IEEE Transactions on Visualization and Computer Graphics* 23, 11 (2017), 2447–2454. DOI : https://doi.org/10.1109/TVCG.2017.2734425

[97] Monique Meuschke, Uli Niemann, Benjamin Behrendt, Matthias Gutberlet, Bernhard Preim, and Kai Lawonn. 2021. GUCCI - guided cardiac cohort investigation of blood flow data. *IEEE Transactions on Visualization and Computer Graphics* (2021), 1–1. DOI : https://doi.org/10.1109/TVCG.2021.3134083

[98] Sina Mohseni, Niloofar Zarei, and Eric D. Ragan. 2021. A multidisciplinary survey and framework for design and evaluation of explainable AI systems. *ACM Transactions on Interactive Intelligent Systems* 11, 3–4 (2021), 45 pages. DOI : https://doi.org/10.1145/3387166

[99] Arthur G. Money and Harry Agius. 2008. Video summarisation: A conceptual framework and survey of the state of the art. *Journal of Visual Communication and Image Representation* 19, 2 (2008), 121 – 143. DOI : https://doi.org/10.1016/j.jvcir.2007.04.002

[100] T. Munzner. 2015. *Visualization Analysis and Design.* CRC Press.

[101] Thomas Mühlbacher, Harald Piringer, Samuel Gratzl, Michael Sedlmair, and Marc Streit. 2014. Opening the black box: Strategies for increased user involvement in existing algorithm implementations. *IEEE Transactions on Visualization and Computer Graphics* 20, 12 (2014), 1643–1652. DOI : https://doi.org/10.1109/TVCG.2014.2346578

[102] Neeta A. Nemade and V. V. Gohokar. 2016. A survey of video datasets for crowd density estimation. In *Proceedings of the 2016 International Conference on Global Trends in Signal Processing, Information Computing and Communication .* IEEE, 389–395.

[103] Ngan Nguyen, Ondřej Strnad, Tobias Klein, Deng Luo, Ruwayda Alharbi, Peter Wonka, Martina Maritan, Peter Mindek, Ludovic Autin, David S. Goodsell, and Ivan Viola. 2021. Modeling in the time of COVID-19: Statistical and rule-based mesoscale models. *IEEE Transactions on Visualization and Computer Graphics* 27, 2 (2021), 722–732. DOI : https://doi.org/10.1109/TVCG.2020.3030415

[104] Y. Nie, C. Xiao, H. Sun, and P. Li. 2013. Compact video synopsis via global spatiotemporal optimization. *IEEE Transactions on Visualization and Computer Graphics* 19, 10 (2013), 1664–1676. DOI : https://doi.org/10.1109/TVCG.2012.176

[105] P. O'Donovan and A. Hertzmann. 2012. AniPaint: Interactive painterly animation from video. *IEEE Transactions on Visualization and Computer Graphics* 18, 3 (2012), 475–487. DOI : https://doi.org/10.1109/TVCG.2011.51

[106] N. Padmanaban, T. Ruban, V. Sitzmann, A. M. Norcia, and G. Wetzstein. 2018. Towards a machine-learning approach for sickness prediction in 360° stereoscopic videos. *IEEE Transactions on Visualization and Computer Graphics* 24, 4 (2018), 1594–1603. DOI : https://doi.org/10.1109/TVCG.2018.2793560

[107] Xingjia Pan, Fan Tang, Weiming Dong, Chongyang Ma, Yiping Meng, Feiyue Huang, Tong-Yee Lee, and Changsheng Xu. 2021. Content-based visual summarization for image collections. *IEEE Transactions on Visualization and Computer Graphics* 27, 4 (2021), 2298–2312. DOI: https://doi.org/10.1109/TVCG.2019.2948611

[108] Ji Hwan Park, Saad Nadeem, Saeed Boorboor, Joseph Marino, and Arie Kaufman. 2021. CMed: Crowd analytics for medical imaging data. *IEEE Transactions on Visualization and Computer Graphics* 27, 6 (2021), 2869–2880. DOI : https://doi.org/10.1109/TVCG.2019.2953026

[109] M. L. Parry, P. A. Legg, D. H. S. Chung, I. W. Griffiths, and M. Chen. 2011. Hierarchical event selection for video story-boards with a case study on snooker video visualization. *IEEE Transactions on Visualization and Computer Graphics* 17, 12 (2011), 1747–1756. DOI : https://doi.org/10.1109/TVCG.2011.208

[110] Ripon Patgiri. 2018. A taxonomy on big data: Survey. arXiv:1808.08474. Retrieved from https://arxiv.org/abs/1808.08474.

[111] N. Patil and Prabir Kumar Biswas. 2016. A survey of video datasets for anomaly detection in automated surveillance. In *Proceedings of the 2016 6th International Symposium on Embedded Computing and System Design.* IEEE, 43–48.

[112] Charles Perin, Romain Vuillemot, and Jean-Daniel Fekete. 2013. SoccerStories: A kick-off for visual soccer analysis. *IEEE Transactions on Visualization and Computer Graphics* 19, 12 (2013), 2506–2515.

[113] Charles Perin, Romain Vuillemot, Charles D. Stolper, John T. Stasko, Jo Wood, and Sheelagh Carpendale. 2018. State of the art of sports data visualization. In *Proceedings of the Computer Graphics Forum.* Wiley Online Library, 663–686.

[114] H. Pileggi, C. D. Stolper, J. M. Boyle, and J. T. Stasko. 2012. SnapShot: Visualization to propel Ice hockey analytics. *IEEE Transactions on Visualization and Computer Graphics* 18, 12 (2012), 2819–2828. DOI : https://doi.org/10.1109/TVCG.2012.263

[115] P. Pjanic, S. Willi, and A. Grundhöfer. 2017. Geometric and photometric consistency in a mixed video and gal-vanoscopic scanning laser projection mapping system. *IEEE Transactions on Visualization and Computer Graphics* 23, 11 (2017), 2430–2439. DOI : https://doi.org/10.1109/TVCG.2017.2734598

[116] Jorge Poco, Angela Mayhua, and Jeffrey Heer. 2017. Extracting and retargeting color mappings from bitmap images of visualizations. *IEEE Transactions on Visualization and Computer Graphics* 24, 1 (2017), 637–646.

[117] Tom Polk, Dominik Jäckle, Johannes Häußler, and Jing Yang. 2019. CourtTime: Generating actionable insights into tennis matches using visual analytics. *IEEE Transactions on Visualization and Computer Graphics* 26, 1 (2019), 397–406.

[118] Tom Polk, Jing Yang, Yueqi Hu, and Ye Zhao. 2014. Tennivis: Visualization for tennis match analysis. *IEEE Transactions on Visualization and Computer Graphics* 20, 12 (2014), 2339–2348.

[119] A. J. Pretorius, M. Bray, A. E. Carpenter, and R. A. Ruddle. 2011. Visualization of parameter space for image analy-sis. *IEEE Transactions on Visualization and Computer Graphics* 17, 12 (2011), 2402–2411. DOI : https://doi.org/10.1109/TVCG.2011.253

[120] Xuebin Qin, Zichen Zhang, Chenyang Huang, Chao Gao, Masood Dehghan, and Martin Jagersand. 2019. BASNet: Boundary-aware salient object detection. In *Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition.* 7471–7481. DOI : https://doi.org/10.1109/CVPR.2019.00766

[121] T. Rhee, L. Petikam, B. Allen, and A. Chalmers. 2017. MR360: Mixed reality rendering for 360° panoramic videos. *IEEE Transactions on Visualization and Computer Graphics* 23, 4 (2017), 1379–1388. DOI : https://doi.org/10.1109/TVCG.2017.2657178

[122] M. Romero, J. Summet, J. Stasko, and G. Abowd. 2008. Viz-A-Vis: Toward visualizing video through computer vision. *IEEE Transactions on Visualization and Computer Graphics* 14, 6 (2008), 1261–1268. DOI : https://doi.org/10.1109/TVCG.2008.185

[123] Oliver Rübel and Benjamin P. Bowen. 2017. BASTet: Shareable and reproducible analysis and visualization of mass spectrometry imaging data via OpenMSI. *IEEE Transactions on Visualization and Computer Graphics* 24, 1 (2017), 1025–1035.

[124] Dominik Sacha, Michael Sedlmair, Leishi Zhang, John Aldo Lee, Daniel Weiskopf, Stephen North, and Daniel Keim. 2016. Human-centered machine learning through interactive visualization. *European Symposium on Artificial Neural Networks*, ESANN.

[125] D. Sacha, A. Stoffel, F. Stoffel, B. C. Kwon, G. Ellis, and D. A. Keim. 2014. Knowledge generation model for visual analytics. *IEEE Transactions on Visualization and Computer Graphics* 20, 12 (2014), 1604–1613. DOI : https://doi.org/10.1109/TVCG.2014.2346481

[126] D. Sacha, L. Zhang, M. Sedlmair, J. A. Lee, J. Peltonen, D. Weiskopf, S. C. North, and D. A. Keim. 2017. Visual interac-tion with dimensionality reduction: A structured literature analysis. *IEEE Transactions on Visualization and Computer Graphics* 23, 1 (2017), 241–250. DOI : https://doi.org/10.1109/TVCG.2016.2598495

[127] Lawrence K. Saul. 2020. A tractable latent variable model for nonlinear dimensionality reduction. *Proceedings of the National Academy of Sciences* 117, 27 (2020), 15403–15408.

[128] J. Schmidt, M. E. Gröller, and S. Bruckner. 2013. VAICo: Visual analysis for image comparison. *IEEE Transactions on Visualization and Computer Graphics* 19, 12 (2013), 2090–2099. DOI:https://doi.org/10.1109/TVCG.2013.213

[129] Will J. Schroeder, Bill Lorensen, and Ken Martin. 2006. *The Visualization Toolkit: An Object-oriented Approach to 3D Graphics.* Kitware.

[130] T. Schultz and G. L. Kindlmann. 2013. Open-box spectral clustering: Applications to medical image analysis. *IEEE Transactions on Visualization and Computer Graphics* 19, 12 (2013), 2100–2108. DOI:https://doi.org/10.1109/TVCG.2013.181

[131] Daniel Seebacher, Thomas Polk, Halldor Janetzko, Daniel Keim, Tobias Schreck, and Manuel Stein. 2021. Investigating the sketchplan: A novel way of identifying tactical behavior in massive soccer datasets. *IEEE Transactions on Visualization and Computer Graphics* (2021), 1–1. DOI:https://doi.org/10.1109/TVCG.2021.3134814

[132] Christin Seifert, Aisha Aamir, Aparna Balagopalan, Dhruv Jain, Abhinav Sharma, Sebastian Grottel, and Stefan Gumhold. 2017. *Visualizations of Deep Neural Networks in Computer Vision: A Survey.* Springer International Publishing, Cham, 123–144. DOI:https://doi.org/10.1007/978-3-319-54024-5_6

[133] Amir Semmo and Jürgen Döllner. 2014. Image filtering for interactive level-of-abstraction visualization of 3D scenes. In *Proceedings of the Workshop on Computational Aesthetics.* 5–14.

[134] A. Serrano, I. Kim, Z. Chen, S. DiVerdi, D. Gutierrez, A. Hertzmann, and B. Masia. 2019. Motion parallax for 360° RGBD video. *IEEE Transactions on Visualization and Computer Graphics* 25, 5 (2019), 1817–1827. DOI:https://doi.org/10.1109/TVCG.2019.2898757

[135] Conglei Shi, Siwei Fu, Qing Chen, and Huamin Qu. 2015. VisMOOC: Visualizing video clickstream data from massive open online courses. In *Proceedings of the IEEE Pacific Visualization Symposium.* IEEE, 159–166.

[136] Danqing Shi, Fuling Sun, Xinyue Xu, Xingyu Lan, David Gotz, and Nan Cao. 2021. AutoClips: An automatic approach to video generation from data facts. In *Proceedings of the Computer Graphics Forum.* Wiley Online Library, 495–505.

[137] Huang-Chia Shih. 2017. A survey of content-aware video analysis for sports. *IEEE Transactions on Circuits and Systems for Video Technology* 28, 5 (2017), 1212–1231.

[138] Xinhuan Shu, Aoyu Wu, Junxiu Tang, Benjamin Bach, Yingcai Wu, and Huamin Qu. 2021. What makes a data-GIF understandable? *IEEE Transactions on Visualization and Computer Graphics* 27, 2 (2021), 1492–1502. DOI:https://doi.org/10.1109/TVCG.2020.3030396

[139] Antonios Somarakis, Marieke E. Ijsselsteijn, Sietse J. Luk, Boyd Kenkhuis, Noel F. C. C. de Miranda, Boudewijn P.F. Lelieveldt, and Thomas Höllt. 2021. Visual cohort comparison for spatial single-cell omics-data. *IEEE Transactions on Visualization and Computer Graphics* 27, 2 (2021), 733–743. DOI:https://doi.org/10.1109/TVCG.2020.3030336

[140] Antonios Somarakis, Vincent Van Unen, Frits Koning, Boudewijn Lelieveldt, and Thomas Höllt. 2021. ImaCytE: Visual exploration of cellular micro-environments for imaging mass cytometry data. *IEEE Transactions on Visualization and Computer Graphics* 27, 1 (2021), 98–110. DOI:https://doi.org/10.1109/TVCG.2019.2931299

[141] H. Song, J. Lee, T. J. Kim, K. H. Lee, B. Kim, and J. Seo. 2017. GazeDx: Interactive visual analytics framework for comparative gaze analysis with volumetric medical images. *IEEE Transactions on Visualization and Computer Graphics* 23, 1 (2017), 311–320. DOI:https://doi.org/10.1109/TVCG.2016.2598796

[142] M. Stein, H. Janetzko, A. Lamprecht, T. Breitkreutz, P. Zimmermann, B. Goldlücke, T. Schreck, G. Andrienko, M. Grossniklaus, and D. A. Keim. 2018. Bring it to the pitch: Combining video and movement data to enhance team sport analysis. *IEEE Transactions on Visualization and Computer Graphics* 24, 1 (2018), 13–22. DOI:https://doi.org/10.1109/TVCG.2017.2745161

[143] M. Stengel, P. Bauszat, M. Eisemann, E. Eisemann, and M. Magnor. 2015. Temporal video filtering and exposure control for perceptual motion blur. *IEEE Transactions on Visualization and Computer Graphics* 21, 5 (2015), 663–671. DOI:https://doi.org/10.1109/TVCG.2014.2377753

[144] Hendrik Strobelt, Sebastian Gehrmann, Michael Behrisch, Adam Perer, Hanspeter Pfister, and Alexander M. Rush. 2018. Seq2Seq-Vis: A visual debugging tool for sequence-to-sequence models. *IEEE Transactions on Visualization and Computer Graphics* 25, 1 (2018), 353–363. DOI:https://doi.org/10.1109/TVCG.2018.2865044

[145] Hendrik Strobelt, Sebastian Gehrmann, Hanspeter Pfister, and Alexander M. Rush. 2018. LSTMVis: A tool for visual analysis of hidden state dynamics in recurrent neural networks. In *Proceeding of the IEEE Transactions on Visualization and Computer Graphics*, Vol. 24, 667–676. DOI:10.1109/TVCG.2017.2744158

[146] T. Subetha and S. Chitrakala. 2016. A survey on human activity recognition from videos. In *Proceedings of the 2016 International Conference on Information Communication and Embedded Systems.* IEEE, 1–7.

[147] C. Bane Sullivan and Alexander Kaszynski. 2019. PyVista: 3D plotting and mesh analysis through a streamlined interface for the visualization toolkit (VTK). *Journal of Open Source Software* 4, 37 (2019), 1450. DOI:https://doi.org/10.21105/joss.01450

[148] K. Sunkavalli, N. Joshi, S. B. Kang, M. F. Cohen, and H. Pfister. 2012. Video snapshots: Creating high-quality images from video clips. *IEEE Transactions on Visualization and Computer Graphics* 18, 11 (2012), 1868–1879. DOI : https://doi.org/10.1109/TVCG.2012.72

[149] J. Tan, S. DiVerdi, J. Lu, and Y. Gingold. 2019. Pigmento: Pigment-based image analysis and editing. *IEEE Transactions on Visualization and Computer Graphics* 25, 9 (2019), 2791–2803. DOI : https://doi.org/10.1109/TVCG.2018.2858238

[150] Bin Tian, Qingming Yao, Yuan Gu, Kunfeng Wang, and Ye Li. 2011. Video processing techniques for traffic flow monitoring: A survey. In *Proceedings of the 2011 14th International IEEE Conference on Intelligent Transportation Systems.* IEEE, 1103–1108.

[151] L. Turban, F. Urban, and P. Guillotel. 2017. Extrafoveal video extension for an immersive viewing experience. *IEEE Transactions on Visualization and Computer Graphics* 23, 5 (2017), 1520–1533. DOI : https://doi.org/10.1109/TVCG.2016.2527649

[152] Laurens Van der Maaten and Geoffrey Hinton. 2008. Visualizing data using t-SNE. *Journal of Machine Learning Research* 9, 11 (2008), 2579–2605.

[153] Laurens Van Der Maaten, Eric Postma, and Jaap Van den Herik. 2009. Dimensionality reduction: A comparative. *Journal of Machine Learning Research* 10, 66–71 (2009), 13.

[154] Jiachen Wang, Kejian Zhao, Dazhen Deng, Anqi Cao, Xiao Xie, Zheng Zhou, Hui Zhang, and Yingcai Wu. 2019. Tac-simur: Tactic-based simulative visual analytics of table tennis. *IEEE Transactions on Visualization and Computer Graphics* 26, 1 (2019), 407–417.

[155] Qianwen Wang, Zhutian Chen, Yong Wang, and Huamin Qu. 2021. A survey on ML4VIS: Applying machinelearning advances to data visualization. *IEEE Transactions on Visualization and Computer Graphics* 28, 12 (2021), 1–1. DOI : https://doi.org/10.1109/TVCG.2021.3106142

[156] Shangfei Wang and Qiang Ji. 2015. Video affective content analysis: A survey of state-of-the-art methods. *IEEE Transactions on Affective Computing* 6, 4 (2015), 410–430.

[157] Xingbo Wang, Yao Ming, Tongshuang Wu, Haipeng Zeng, Yong Wang, and Huamin Qu. 2021. DeHumor: Visual analytics for decomposing humor. *IEEE Transactions on Visualization and Computer Graphics* 28, 12 (2021), 1–1. DOI : https://doi.org/10.1109/TVCG.2021.3097709

[158] Y. Wang, D. Bowman, D. Krum, E. Coalho, T. Smith-Jackson, D. Bailey, S. Peck, S. Anand, T. Kennedy, and Y. Abdrazakov. 2008. Effects of video placement and spatial context presentation on path reconstruction tasks with contextualized videos. *IEEE Transactions on Visualization and Computer Graphics* 14, 6 (2008), 1755–1762. DOI : https://doi.org/10.1109/TVCG.2008.126

[159] Y. Wang, F. Liu, P. Hsu, and T. Lee. 2013. Spatially and temporally optimized video stabilization. *IEEE Transactions on Visualization and Computer Graphics* 19, 8 (2013), 1354–1361. DOI : https://doi.org/10.1109/TVCG.2013.11

[160] Y. Wang, Y. Liu, X. Tong, Q. Dai, and P. Tan. 2018. Outdoor markerless motion capture with sparse handheld video cameras. *IEEE Transactions on Visualization and Computer Graphics* 24, 5 (2018), 1856–1866. DOI : https://doi.org/10.1109/TVCG.2017.2693151

[161] Y. Wang, Z. Wang, C. Fu, H. Schmauder, O. Deussen, and D. Weiskopf. 2019. Image-based aspect ratio selection. *IEEE Transactions on Visualization and Computer Graphics* 25, 1 (2019), 840–849.

[162] Yifan Wang, Guoli Yan, Haikuan Zhu, Sagar Buch, Ying Wang, Ewart Mark Haacke, Jing Hua, and Zichun Zhong. 2021. VC-Net: Deep volume-composition networks for segmentation and visualization of highly sparse and noisy image data. *IEEE Transactions on Visualization and Computer Graphics* 27, 2 (2021), 1301–1311. DOI : https://doi.org/10.1109/TVCG.2020.3030374

[163] J. Wei, C. Li, S. Hu, R. R. Martin, and C. Tai. 2012. Fisheye video correction. *IEEE Transactions on Visualization and Computer Graphics* 18, 10 (2012), 1771–1783. DOI : https://doi.org/10.1109/TVCG.2011.130

[164] Sebastian Weiss, Mengyu Chu, Nils Thuerey, and Rüdiger Westermann. 2021. Volumetric Isosurface rendering with deep learning-based super-resolution. *IEEE Transactions on Visualization and Computer Graphics* 27, 6 (2021), 3064–3078. DOI : https://doi.org/10.1109/TVCG.2019.2956697

[165] Xin Wen, Miao Wang, Christian Richardt, Ze-Yin Chen, and Shi-Min Hu. 2020. Photorealistic audio-driven video portraits. *IEEE Transactions on Visualization and Computer Graphics* 26, 12 (2020), 3457–3466.

[166] James Wexler, Mahima Pushkarna, Tolga Bolukbasi, Martin Wattenberg, Fernanda Viégas, and Jimbo Wilson. 2020. The what-if tool: Interactive probing of machine learning models. *IEEE Transactions on Visualization and Computer Graphics* 26, 1 (2020), 56–65. DOI : https://doi.org/10.1109/TVCG.2019.2934619

[167] Aoyu Wu and Huamin Qu. 2018. Multimodal analysis of video collections: Visual exploration of presentation techniques in ted talks. *IEEE Transactions on Visualization and Computer Graphics* 26, 7 (2018), 2429–2442.

[168] Aoyu Wu, Yun Wang, Xinhuan Shu, Dominik Moritz, Weiwei Cui, Haidong Zhang, Dongmei Zhang, and Huamin Qu. 2021. AI4VIS: Survey on artificial intelligence approaches for data visualization. *IEEE Transactions on Visualization and Computer Graphics* 28, 12 (2021), 1–1. DOI : https://doi.org/10.1109/TVCG.2021.3099002

[169] Yingcai Wu, Ji Lan, Xinhuan Shu, Chenyang Ji, Kejian Zhao, Jiachen Wang, and Hui Zhang. 2017. Ittvis: Interactive visualization of table tennis data. *IEEE Transactions on Visualization and Computer Graphics* 24, 1 (2017), 709–718.

[170] Yingcai Wu, Xiao Xie, Jiachen Wang, Dazhen Deng, Hongye Liang, Hui Zhang, Shoubin Cheng, and Wei Chen. 2018. Forvizor: Visualizing spatio-temporal team formations in soccer. *IEEE Transactions on Visualization and Computer Graphics* 25, 1 (2018), 65–75.

[171] C. Xiao, M. Liu, N. Yongwei, and Z. Dong. 2011. Fast exact nearest patch matching for patch-based image editing and processing. *IEEE Transactions on Visualization and Computer Graphics* 17, 8 (2011), 1122–1134. DOI: https://doi.org/10.1109/TVCG.2010.226

[172] Xiao Xie, Xiwen Cai, Junpei Zhou, Nan Cao, and Yingcai Wu. 2018. A semantic-based method for visualizing large image collections. *IEEE Transactions on Visualization and Computer Graphics* 25, 7 (2018), 2362–2377.

[173] Chaoqing Xu, Tyson Allan Neuroth, Takanori Fujiwara, Ronghua Liang, and Kwan-Liu Ma. 2021. A predictive visual analytics system for studying neurodegenerative disease based on DTI fiber tracts. *IEEE Transactions on Visualization and Computer Graphics* (2021), 1–1. DOI: https://doi.org/10.1109/TVCG.2021.3137174

[174] Kai Xu, Dae Hoon Park, Chang Yi, and Charles A. Sutton. 2018. Interpreting deep classifier by visual distillation of dark knowledge. arXiv:1803.04042. Retrieved from http://arxiv.org/abs/1803.04042.

[175] Lan Xu, Wei Cheng, Kaiwen Guo, Lei Han, Yebin Liu, and Lu Fang. 2021. FlyFusion: Realtime dynamic scene reconstruction using a flying depth camera. *IEEE Transactions on Visualization and Computer Graphics* 27, 1 (2021), 68–82. DOI: https://doi.org/10.1109/TVCG.2019.2930691

[176] Mai Xu, Yilin Liang, and Zulin Wang. 2015. State-of-the-art video coding approaches: A survey. In *Proceedings of the 2015 IEEE 14th International Conference on Cognitive Informatics & Cognitive Computing*. IEEE, 284–290.

[177] Ji Soo Yi, Youn ah Kang, and John Stasko. 2007. Toward a deeper understanding of the role of interaction in information visualization. *IEEE Transactions on Visualization and Computer Graphics* 13, 6 (2007), 1224–1231.

[178] J. Yoon, I. Lee, and H. Kang. 2012. Video painting based on a stabilized time-varying flow field. *IEEE Transactions on Visualization and Computer Graphics* 18, 1 (2012), 58–67. DOI: https://doi.org/10.1109/TVCG.2011.47

[179] Jun Yuan, Changjian Chen, Weikai Yang, Mengchen Liu, Jiazhi Xia, and Shixia Liu. 2021. A survey of visual analytics techniques for machine learning. *Computational Visual Media* 7, 1 (2021), 3–36.

[180] Jan Zahálka, Marcel Worring, and Jarke J. Van Wijk. 2020. II-20: Intelligent and pragmatic analytic categorization of image collections. *IEEE Transactions on Visualization and Computer Graphics* 27, 2 (2020), 422–431.

[181] Haipeng Zeng. 2016. Towards better understanding of deep learning with visualization. *The Hong Kong University of Science and Technology* (2016).

[182] Haipeng Zeng, Xinhuan Shu, Yanbang Wang, Yong Wang, Liguo Zhang, Ting-Chuen Pong, and Huamin Qu. 2020. Emotioncues: Emotion-oriented visual summarization of classroom videos. *IEEE Transactions on Visualization and Computer Graphics* 27, 7 (2020), 3168–3181.

[183] J. Zhang, E. Langbehn, D. Krupke, N. Katzakis, and F. Steinicke. 2018. Detection thresholds for rotation and translation gains in 360° video-based telepresence systems. *IEEE Transactions on Visualization and Computer Graphics* 24, 4 (2018), 1671–1680. DOI: https://doi.org/10.1109/TVCG.2018.2793679

[184] Peiying Zhang, Chenhui Li, and Changbo Wang. 2020. Viscode: Embedding information in visualization images using encoder-decoder network. *IEEE Transactions on Visualization and Computer Graphics* 27, 2 (2020), 326–336.

[185] Jialin Zhu and Tom Kelly. 2021. Seamless satellite-image synthesis. In *Proceedings of the Computer Graphics Forum*. Wiley Online Library, 193–204.