AR Record&Replay: Situated Compositing of Video Content in Mobile Augmented Reality

Tobias Langlotz[#], Mathäus Zingerle[*], Hannes Kaufmann[*], Gerhard Reitmayr[#]

[#]Graz University of Technology
Inffeldgasse 16, 8010 Graz, Austria
{langlotz, reitmayr}@icg.tugraz.at

[*]Vienna University of Technology
Favoritenstrasse 9-11, 1040 Vienna, Austria
{zingerle, kaufmann}@ims.tuwien.ac.at

Figure 1. Situated video augmentations. (Left) Original video footage recorded using a mobile phone. (Right) Augmented Video application. The foreground video object – in this case the skateboarder – is augmented in the users view.

ABSTRACT

In this paper we present a novel approach to record and replay video content composited in-situ with a live view of the real environment. Our real-time technique works on mobile phone, and made usage of an augmented reality panorama-based tracker to create visually seamless and spatially registered mixing of video content. For this purpose, we apply a temporal foreground-background segmentation of a video footage and show how the segmented information can be precisely registered in real-time in the camera view of a mobile phone. We present the user interface and the video posts effects implemented in our prototype as well as demonstrating our approach for a skateboard training application. Our technique can also be used with online video material or support the creation of augmented situated documentaries.

Author Keywords

Augmented Video, Augmented Reality, Mobile phone

ACM Classification Keywords

H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems – Artificial, augmented, and virtual realities

OZCHI'12, November 26–30, 2012, Melbourne, Victoria, Australia.

INTRODUCTION

The availability of cheap mobile video recorders and the integration of high quality video recording capabilities into smartphones has rapidly increased the amount of videos being created and shared with other people using platforms such as YouTube[1] or Vimeo[2]. With more than 50 hours of video uploaded every minute on YouTube and billions of video viewed each day[3], new way to search, browse and experience video content is highly needed.

Yet, current user interfaces of online tools for video content have mostly been replicating the existing photo interfaces. Features such as geo-tagging or browsing geo-referenced content in Google Earth[4] (or other map-based applications) have been mainly reproduced for video content.

More recently, efforts have been made to explore further the spatio-temporal aspect of videos. Application such as Photo tourism (Snavely et al., 2006) have inspired work of Ballan et al., 2010, allowing end-users to experience multi-viewpoint events recorded by multiple cameras. Their system allows smooth transition between camera

[1] http://www.youtube.com

[2] http://www.vimeo.com

[3] http://www.youtube.com/t/press_statistics

[4] http://www.earth.google.com

viewpoints and offers a flexible way to browse and create video montage captured from multiple perspectives.

However, these systems limit themselves to produce and explore video content on desktop user interface (e.g. web, virtual globe), out of a real context. *Augmented Reality* (AR) technology can overcome this issue, as providing a way to place geo-referenced video content on a live, spatially registered view of the real world.

For example, Höllerer et al., 1999 investigate situated documentaries and show how to incorporate video information into an outdoor wearable AR system to realize complex narratives in an outdoor environment. Recent commercial AR browsers such as Layar[5] or Wikitude[6] are now integrating this feature, supporting video files or image stacks, but with limited spatial registration due to the fact that the video is always screen aligned and the error prone registration based on sensors.

Video augmentation has also been explored for publishing media. RedBull[7] presented an AR application that augmented pages of their Red Bulletin magazine with video material using *Natural Feature Tracking* (NFT). The application was running within a webpage as an Adobe Flash application to detect the shown magazine page and played the video content spatially overlaid on top of that page.

As these projects generally present the video on a 2D billboard type of representation, other work have been exploring how to provide more seamless mixing between the video content and a live video view. MacIntyre et al. researched within their *Three Angry Men* project the use of video information as an element for exploiting narratives in augmented reality (MacIntyre et al, 2003). They proposed a system where a user wearing a *Head Mounted Display* (HMD) can see overlay video actors virtually seated and discussing around a real table. The augmented video actors were prerecorded and foreground-background segmentation applied to guarantee a seamless integration into the environment, created with their desktop authoring tool (MacIntyre et al, 2001, MacIntyre et al, 2002).

Whereas MacIntyre et al. used static camera recording of actors, the 3D Live (Prince et al, 2002) system extended this concept to 3D video. Prince et al. used a cylindrical multi-camera capture system, allowing the capture and the real-time replay a 3D model of a person using a shape-from-silhouette approach. Their system was supporting remote viewing, by transmitting the 3D via a network and displaying the generated 3D video onto an AR setup at a remote location as part of a teleconference system.

As these applications were proposed for indoor scenarios, Farrago[8], an application for mobile phones, proposed video mixing with 3D graphical content for outdoor configurations. The tool records videos that can be edited afterwards by manually adjusting the position of the virtual 3D objects overlay on the video image, but required the usage of 2D marker or face tracking. Once the video is re-rendered with the overlay it can be shared with other users.

In this work, we wanted to investigate how we can offer a new user experience to a mobile user to create real-time compositing between his view of the real world with prerecorded geo-referenced content. Similar to MacIntyre et al., we were interested to extract the information of the video (e.g. moving person or objects) and offer the possibilities to spatially navigate the video (by rotating the phone) mixed with the view of the real world. Differently, we focused on mobile platform, outdoor environment but also looked at offering simple way to record and capture this type of video content with only minimal input. We also have fewer restrictions during the recording as we allow a free rotating camera and do not rely on uni-colored background for recording the video augmentations.

In this paper, we present our AR technique offering accurate spatial registration between recorded video content (e.g. person, motorized vehicles) and real-time of the real world with a seamless visual integration (e.g. extracted break dancer recorded the day before overlay on your camera video). Our system allows to replay past geo-referenced video sequences, to re-enact past captured event for a broad range of application covering sports, history, cultural heritage or tutorials. We offer tools for the user control video playback but also video effects, proposing a first view of what can be a future real-time AR video montage tool on mobile platform.

The presented system operates in three steps. The first step is the recording of the video. After finishing the recording the video is GPS tagged and uploaded to a server for further processing and to make it accessible to other users.

Secondly, a pre-processing step is performed on the cloud server hosting the video or on a desktop PC. In this pre-processing step the information about what parts of the video should be later augmented are extracted. We compute this by applying a segmentation requiring only that the user coarsely outlines the object of interest in the first frame. We also extract the background information of the video and assemble it into a panoramic representation of the background. We use this panorama later for the precise registration of the video content into the environment.

The last step of the system is performed once another user moves to the vicinity of the position the original movie was recorded – identified via GPS. The system downloads the pre-processed video to a smartphone. While the user explores the environment the video is registered and augmented into the users view. For the registration we apply a vision-based matching that allows a precise registration of the video content into the

[5] http://www.layar.com

[6] http://www.wikitude.com

[7] http://www.redbullusa.com

[8] http://www.farragoapp.com

Figure 2. Workflow TBD

environment. Figure 2 presents the typical workflow of our approach.

The proposed system contributes to the field of Augmented Reality by introducing a solution how to incorporate video information into outdoor AR application as well as by demonstrating an interface allowing end users to participate in the content creation process for this kind of application. We hope that this shows the potential of video content in AR application as it also demonstrates how to incorporate more dynamic content in mobile AR applications.

SITUATED VIDEO COMPOSITING FOR AR

In the following we give a detailed description of our approach and algorithms performed in three main steps for video compositing in AR

Video recording

In the first step we record the video material, which we later want to replay in place. The recording of the video is performed as usual but requires the camera position to be not changed while recording – rotational movements of the camera are possible though. This recording can be done with a digital camera or a smartphone with integrated camera. Once the video is completed it is tagged with the GPS location for indexing and to roughly re-identify the position. The tagged video is then uploaded to a cloud-based service or transferred to the personal PC.

Offline video processing

In the next step we need to process the video. The main challenge is to separate the object of interest in the video (foreground) from the remaining information such as the background or other moving objects that are not of interest. We later used the object of interest as overlay but we also want to keep the background information as it is needed to register the video overlay into the new scene. This preprocessing can be done on the personal PC or on the cloud server hosting the uploaded video.

We start segmenting the video by applying a variation of the GraphCut algorithm, namely GrabCut, presented by Rother et al. (Rother et al, 2004). To initiate the algorithm, the user has to roughly sketch the object of interest (the foreground object) and mark some background pixel (see left Figure 3). Once processed the GrabCut delivers a segmentation of the foreground object based on an iterative GraphCut.

This approach previously demonstrated for static images needs to be done for each video frame with temporal content. To avoid this cumbersome task, we extended the method in a similar way to the approach presented by Mooser et al. (Mooser et al., 2007). The idea is to use the segmentation output of the GrabCut algorithm of the previous video frame, to initialize the GrabCut algorithm computing the segmentation for the current frame.

As there is likely a movement between the two frames we cannot simply apply the result of the segmentation from the previous frame to the current one. We overcome this by estimating the position of the segmented foreground object current frame by computing the optical flow of pixels between the previous and the current frame using Lucas-Kanade (Lucas and Kanade, 1981). This gives us an approximation of the foreground objects position in the current frame. Dilating the estimated foreground objects footprint compensates for tracking inaccuracies. We compute the boundary of the estimated foreground object and select pixels within (pixels of the foreground object) and outside (background pixels) and use them as input for GrabCut. Applying this approach for each frame yields the foreground objects for all consecutive frames of the video. We apply a dilate and erosion operation on the segmented foreground objects to remove the noisy borders and only keep the largest connected component as foreground object in case the segmentation computed different segments. We also kept an option to

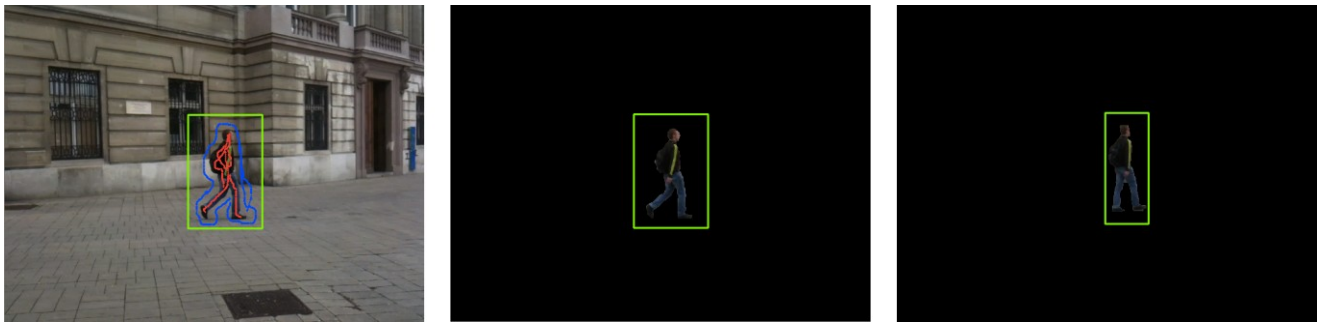

Figure 3. (Left) Manual initialization of the segmentation step. User sketches the foreground object (red) and outlines the background (blue). (Middle) Result of applying GrabCut segmentation to subsequent video frames: Segmented foreground object and size-optimized texture (green outline). (Right) Tracking the segment using Lukas-Kanade Tracker allows segmentation of later frames even in cases the appearance changes.

manually initialize the GrabCut for specific frames in case the object of interest is (partially) not segmented properly.

The segmented foreground object is often only a fraction of the size of the full video frame (see Figure 3). To reduce the data we store the foreground object by only saving the bounding rectangle around it and the offset within the video frame.

Using this approach we can segment the object of interest from the background information by only requiring the user to outline the object of interest in the first frame of the video.

Once the foreground object is extracted we need also the background information, which we later use for registering the object into the view of the user. We therefore take the segmented frames and focus in the following only on the background pixels.

Due to the rotating camera the recorded frames hold different portions of the scene background. Furthermore, the foreground object also occludes parts of the background, reducing the amount of visual features that are later available for vision-based registration. We still want to reconstruct as much background information as possible. We therefore do not only take into account the background information from one video frame but from all frames and integrate them into one a bigger panoramic image.

We create this panoramic image holding the background pixels by using modified version of the panoramic mapping and tracking approach presented by Wagner et al. (Wagner et al, 2010). This approach uses features in the incoming video frames to stitch them to a panoramic image. Wagner et al. also demonstrated in their approach how to track the camera motion R_S of the recording camera besides constructing the panoramic image. Following Wagner et al. we assume that the camera only performed rotational movements.

We changed the panoramic mapping and tracking application in a way that it can handle alpha channels and only maps pixels into the panoramic image that are considered to be background pixels. The holes in the panorama caused by the occluding foreground object that are not mapped are closed over time as the foreground object moves within the camera frames (see Figure 4). We store the resulting panoramic image that contains the background information contained in the video. We also store the camera rotation R_S for each video frame.

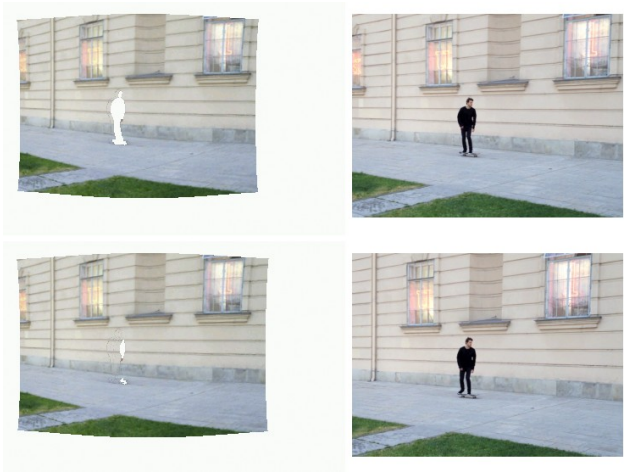

Figure 4. Creating a panoramic image containing background information. (Top) Holes that are caused by occlusion are closed over time (Bottom). The right side always shows the latest camera image that is used as input for the panorama computation.

The resulting processed video information – the segmented video, the panoramic image holding the background information and the camera rotation for each frame – are packaged in a dataset and tagged with the GPS location of the video. The packaged dataset can be shared with friends or made available via the cloud.

Online video processing

The final part of our approach is the online video processing on the smartphone with the goal of augmenting the users view with the object of interest from the video.

We assume that the user's phone is equipped with GPS. The video augmentations are preinstalled on the phone or the current GPS position is used to query a cloud service

for close by video augmentations. The available video augmentations are downloaded as a packaged dataset to the device.

Once the download finishes, we start registering the video into the users view. We therefore start to build a new panoramic image from the current camera feed, while we are also track the camera rotations R_T using again the approach from Wagner et al., 2010.

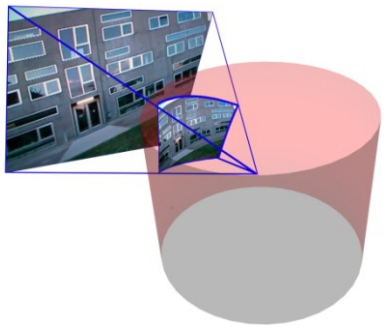

Figure 5. Projection of the camera image into the cylindrical-mapped panoramic image allowing only rotational movements of the camera.

The use of the panorama-based tracking allows for a higher precision of the registration and the tracking, as we do not rely on noisy sensor values. This constrains the camera movement to be only rotational (see Figure 5). However, most users only perform rotational movements while using outdoor AR applications (Grubert et al, 2011) making this constraint acceptable in most scenarios.

While building the new panorama of the environment we try to match the loaded panorama holding the background pixels against this newly built panorama. The matching is performed using PhonySIFT (Wagner et al., 2008) point features, which we compute and match for the panoramas. As soon the overlapping area between both panoramas is big enough, the matching should succeed and give us the transformation T_{ST} describing the transformations between the camera used to record the video (the source camera) and the camera where the video information should be registered in (the target camera). By assuming that the user of the system is roughly at the same position (identified via GPS) we can constrain the transformation T_{ST} to be pure rotational (see Figure 6).

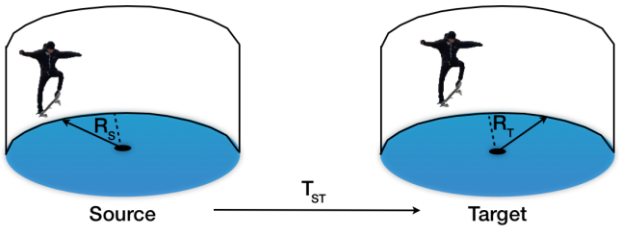

Figure 6. Illustration of the applied transformation between the source camera and the target camera used for replaying the augmented video

Once we computed the transformation T_{ST} we can transform each pixel from the source panorama into the target panorama and vice versa. This allows us to play the video information by overlaying the current environment with the object of interest from the video frame. We therefore load the video frames and by applying the known transformation R_S (the orientation of the source camera computed in the offline video processing step), the transformation T_{ST} (the transformation between the source and the target camera gained from the registration) and the transformation R_T (the orientation of the target camera computed using the panorama-based tracking) we can precisely augment the video content into the users view (see Figure 6). By using the panorama-based tracker we compute an update of the transformation R_T at each frame. This allows us to rotate the target camera completely independent from the source camera and maintaining the precise registration of the video in the current view.

PROTOTYPE

Integrated into AR browser system video augmentations would allow for a wide range of possible applications in the domain of Entertainment and Edutainment.

We implemented our technique in a prototype of a mobile video editing AR application. Inspired by the current tools proposed in desktop video editing applications, we focused on some of their major features: video layers, video playback control and video effects.

In the following we present an overview about the implemented interface and post-effects that are within video augmentations as implemented in our system and further present two implemented applications that make use of our approach and the post effects as well as they show the potential of our approach.

User Interface

The interface of our prototype is inspired by video editing tools and the controls are grouped into 3 main groups that can all be operated video the touchscreen of the phone. The control groups are: the video control group, the video layer group and the video effects group (see Figure 7).

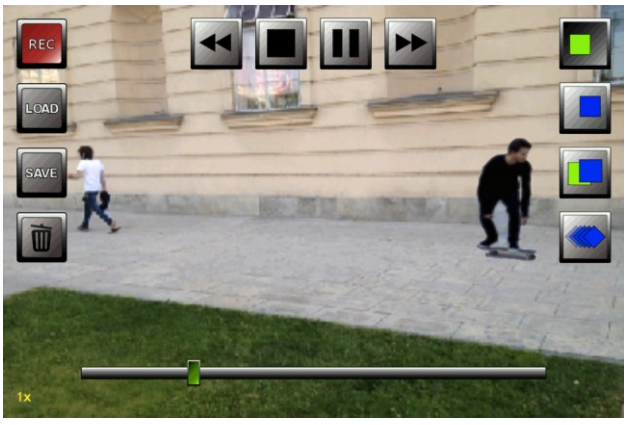

Figure 7. Screenshot of our prototype showing a video augmentation together with the implemented interface with the control groups for video control, video layers and video effects. (TBD update image)

The controls in the video control group consist of the default elements provided in most video players such as

Figure 8. Examples of realized post effects as used in our Skatboard Tutor application. (Left) Playing back two video augmentations allows the comparison of the riders' performance. (Right) Flash-trail effects to visualize the path and the motion within the video.

the play control buttons (e.g. play, pause, double video speed). We further provide a slider at the bottom of the screen that allows moving forward and backward in the video or even jump to the asked position.

The video layer group provides access to video layers. In our approach a video layer is one video that can be augmented. As our approach allows the playback of multiple video augmentations at the same time the video layer group can be used two switch certain videos on or off or even allows playing all of them simultaneously to combine the recorded actions.

The video effects group allows to switch between the set of implemented post effects that are applied in real-time to the video augmentations. Each effect can be switched on or off and the effects can be combined with each other.

Post effects

Applying visual effects is an important part of video post-productions as they can be used to highlight certain actions as well as they can be used to show things that are impossible in the real world. Normally these effects are applied to the video material in a rendering step that is carried out in an offline manner (Linz et al, 2010).

Because of the nature of our approach we are able to perform a wide variety of these video effects in real-time on a mobile device without the need of pre-rendering the video.

Our approach allows us to play back more than one video at the same time by still allowing a seamless integration into the environment (see Figure 8 left). This allows it to compare certain action that were performed at the same place but at different time by integrate them into one view thus bridging time constrains. Each video corresponds in our system to a video layer and the user can switch between these layers or play them simultaneously.

We have further investigated the use of space-time visual effects such as multi-exposure effects, open flash and flash-trail effects. Multi-exposure effects simulate the behavior of a multi exposure film where several images are visible at the same time. We can easily simulate this behavior for cameras with a fixed viewpoint by augmenting several frames of our videos at the same time. This results in having the subject appearing several times within the current view, such as in a multi exposure.

An extension of this effect is the Flash trail effects. Flash trail effects also allows seeing multiple instances of the same subject but the visibility depends on the time passed by (see Figure 8 right). This effect supports a better understanding of the motion in the recorded video. We implemented the Flash trail effect by blending in past frames of the augmented video with increasing amount of transparency. Thereby the strength of the transparency and the time between the frames can be freely adjusted.

All these presented effects do not require any preprocessing but are carried out on the device while playing back the video. They can therefor be combined or switched off on users demand.

Skateboard Tutor Application

We decided to demonstrate our system as part of a Skateboard Tutor application. Tutorial/How-To videos take a big share in todays YouTube videos (Sharma and Elidrisi, 2008) showing the interest in this video genre.

The Skateboard Tutor application allows recording skateboard tricks that can be shared with other users for demonstration and learning purposes. The application can be used to overlay the prerecorded videos in place to experience the skateboard tricks and actions in the correct context. We think that this supports the learning process as normal skateboard videos can give the wrong perception of the environment by using camera lenses with specific characteristics.

The videos are recorded with normal digital cameras or smartphones and are processed using our approach of situated video compositing for AR. We also make use of the proposed post effects and layers. The layer approach allows recording own skateboard tricks, which can later be played in parallel with downloaded tricks for

Figure 8. Scenario used during the user study. (Left) Skateboarder was recorded with a mobile phone while performing his actions. (Middle) Frame of the recorded video sequence. (Right) TBD The same action as augmented within our Skateboard Tutor application.

comparison (e.g. speed, height of jumps). The flash-trail effect can be used to highlight the motion and the path of the rider.

We implemented the offline video processing using OpenCV[9]. We mobile application is implemented on the iOS platform using the Studierstube ES framework (Schmalstieg and Wagner, 2008). We tested the application successfully on the iPhone 3GS, iPhone 4S and an iPad2 where it runs in realtime between 17fps (iPhone 3GS) and 28 fps (iPhone 4S/iPad2).

EVALUATION

Because of the novelty of the proposed approach we decided to evaluate our approach of situated compositing of video content in mobile augmented reality within the scope of the created Skateboard Tutor application.

Scenario and Setting

As outlined earlier there are many possible application scenarios for in-situ video augmentations on mobile phones. However, we decided to conduct our evaluation using the created Skateboard Tutor application as we agreed that this is a very relevant use case scenario and further allows us to exploit several unique characteristics of our system. Firstly, skateboard videos are usually falling into the genre of how-to videos. Therefore they resemble one of the most common genres at YouTube or other video platforms. Secondly, producers of this kind of videos are commonly also consumers, allowing us to receive feedback for the creation of video augmentations as well as the experiencing of video augmentations. Finally, tricks performed with a skateboard are heavily tied to the environment and location through obstacles and ramps, which makes it interesting to experience them in the real environment.

We therefore decided to demonstrate our Skateboard Tutor application to invited participants. We only invited skateboarders that have also experience in creating skateboard videos or tutorials that are shared video were shared via online video platforms. We therefor assume all participants to be domain experts.

We conducted this preliminary user study for gathering first user feedback on our technique as well as to identify flaws and get additional ideas for further improvements.

Our main interest was on the usefulness of our approach as well as the usability of our created prototype.

In total we had 5 expert users (>7 years of skateboarding experience, all of them involved in producing skateboard videos, some produced videos for marketing). The participants were all male and between 25 and 28 years old. Two did previously hear about augmented reality but all participants never used any kind of AR application beforehand. All except one participant stated that they are very familiar with the usage of mobile devices, while one participant said that he only uses his old mobile phone to place calls or write messages.

We gave all participants the chance to get hands-on experience with our prototype demonstrated on an iPhone 3GS and an iPad2. Because of time constrains only 2 were able to create their own skateboard video that was later augmented while all users had the chance to experience the videos. After the participants finished the demonstration we asked them a series of questions as part of a semi-structured interview.

Results

In the interview all participants agreed that our presented system is easy to use. They said that they think this is valid not only for them but also for other users. All except one participant agreed that they felt comfortable using our application. The remaining participant gave an average rating but justified it with that fact that he is not used to the form factor of the used devices as he is using an older and smaller phone. All users agreed that our system is intuitive and easy to learn to use and they also left positive feedback as asked about the interface.

Asked about the applicability and the usefulness of experiencing video in place the user agreed that they were convinced about what they saw and that they think it can be quite useful. However, two of them pointed out during the interview that the users have to visit the place, which makes more sense in certain specific cases. They stated therefor that they see in general more as gadget.

[9] http://sourceforge.net/projects/opencvlibrary/

Asked about the implemented Skateboard Tutor scenario they all agreed on the usefulness. They stated that based on their perception and knowledge many kids have a wrong impression over the context the shot was made in making them believe certain tricks have a different difficulty. Three of the five participants said that they enjoyed the freedom of having full control of the perspective during playback, as they did not rely on the original camera perspective. This is something not possible with normal video playback. They explicitly highlighted that they enjoyed the possibility of playing several videos/layers at the same time that are overlaid in parallel. They said that this is very useful tool as it allows the users to compare their own run (after they recorded it), with the tutorial video to detect differences. They also liked the flash-trail effect saying that this effect seem to be useful for studying "the line" a rider skates.

The last part of the interview contained questions regarding the visual quality of the presentation. Three participants had the feeling that the whole rendering was that the scene and the rider were 3-dimenensional and very authentic. One participant said that he had the feeling the scene and the registration of the skateboarder was 3-dimensional but the skateboarder itself was flat/2D. The remaining two agreed that it was all 2D. They all agreed that the movement of the skateboarders within the scene was very realistic and even after being explicitly asked they could remember to have seen any drifting between the augmentation and the background. However, asked about how seamless the integration was they gave mixed answers. They reasoned that by stating that sometimes the skateboarder did not have the same appearance as the background as it appeared it bit too dark or lightened incorrectly. Two participants also noticed small segmentation errors (a wheel of the skateboard was disappearing in a couple of video frames).

The two participants that also tried the creation of an augmented video using our approach stated that it was easy to use and the additional overhead is justified by the possible result, even though one of the participants mentioned that he had to run the segmentation twice to achieve a good result. Asked about constrains it movement (only rotational movements of the camera are permitted) they said that is likely to be acceptable in most cases as in their opinion a huge majority of the people is making short videos with smartphone devices from a single point of view. Thus they will meet all criteria. One participant said that the given constrains fit the medium, as he thinks the majority of the short online videos were shot in this [constrained] way. Finally, during the open questions one participant proposed the possible use of our application as a mobile blue screen, which allows users to capture objects and scenes and assemble them together using the layer view.

DISCUSSION

Overall the evaluation using domain experts showed that our approach poses advantages over existing video applications especially within our realized scenario. However the final outcome and the usefulness strongly depends on the use case. Our application is easy to handle and intuitive to use.

The biggest problem was the visual quality of the overlay. Even though the ratings were above average the user complained about the effect know as visual coherence: The augmentation looked different than the current environment. In our case this was mostly caused by cloudy weather during recording time resulting in low contrast actors while it was sunny during the playback of the augmentations. However, the problem of visual coherence is an active research area in augmented reality and needs to be treaded outside the scope of this project. Another problem was the segmentation that sometimes was not accurate enough, especially if applied to a well structure background as needed for vision-based registration. However more sophisticated segmentation algorithms and better algorithms for tracking the segmented objects exist but need to be investigated in the context of this work.

Despite these drawbacks that need further investigations our application showed that augmented video can be an interesting element and especially as video content is often easier to create than 3D content, making our approach interesting for many applications

Professional applications can benefit from video augmentations as realized in our approach. Augmented reality-based tourist guides could display more interactive content e.g. by capturing the guide for later replay. Furthermore authoring that content is less demanding than creating dynamic 3D content. This allows it to easily created in-situ narratives similar to the concept of situated documentaries presented by Höllerer et al., 1999.

Many augmented reality application can benefit from the ease needed to create video augmentations using our approach allowing laypersons to create content and share it with friends. This made it possible to create videos of certain events (e.g. parades, street artists etc.) and play them back in place at a different time.

We see a further application area in mobile video assembling and editing. This means that one subject is doing certain actions that are recorded in a video. When this augmented video is played another subject can appear in the camera image, thus both of their actions are combined in one video that again can be replayed in place.

CONCLUSION

We presented an approach for in-situ compositing of video content in mobile augmented reality. We showed how to create and process video files for the use in mobile AR as well as how to register them precisely in the users environment using a panorama-based tracking approach. Even though the approach is constrained to rotational movements of the cameras due to the usage of an panoramic representation of the environment it could be applied to many existing outdoor AR applications as this movement pattern is the most common for using AR browsers and as well as for doing small video shots.

We demonstrated the whole application within a Skateboard Tutor prototype. Our prototype allows experiencing Skateboard tricks and actions recorded by other people that are augmented in-place and displayed at interactive frame rates on mobile phones.

Future work targets better segmentation algorithms and an improved visual coherence between the overlay and the augmented environment. Porting the offline video processing to the mobile could be another future step.

Overall we hope that this work demonstrates possible usage of video footage in future mobile AR applications as well as it shows the advantages of interfaces that allow experiencing videos in place.

ACKNOWLEDGMENTS

We would like to thank all users participating in the experiments. We especially thank Raphael Grasset and Holger Regenbrecht for their input and discussion. This work was partially supported by the Christian Doppler Laboratory for Handheld Augmented Reality.

REFERENCES

Ballan, L., Brostow, G.J., Puwein, J., and Pollefeys, M. Unstructured video-based rendering: Interactive Exploration of Casually Captured Videos. ACM SIGGRAPH 2010 papers on – SIGGRAPH'10, ACM Press (2010).

Grubert, J., Langlotz, T., and Grasset, R. Augmented Reality Browser Survey. Technical Report 2012, http://www.icg.tugraz.at/publications/augmented-reality-browser-survey, 2012.

Guven, S. and Feiner, S. Visualizing and navigating complex situated hypermedia in augmented and virtual reality. 2006 IEEE/ACM International Symposium on Mixed and Augmented Reality, IEEE (2006), 155-158.

Höllerer, T., Feiner, S., and Pavlik, J. Situated Documentaries: Embedding Multimedia Presentations in the Real World. In Proceedings of the 3rd IEEE International Symposium on Wearable Computers (ISWC'99), (1999), 79-86.

Linz, C., Lipski, C., Rogge, L., Theobalt, C., and Magnor, M. Space-time visual effects as a post-production process. Proceedings of the 1st international workshop on 3D video processing - 3DVP '10, ACM Press (2010).

Lucas, B.D. and Kanade, T. An iterative image registration technique with an application to stereo vision. IJCAI'81 Proceedings of the 7th international joint conference on Artificial intelligence - Volume 2, (1981), 674-679.

Macintyre, B., Lohse, M., Bolter, J.D., and Moreno, E. Ghosts in the Machine : Integrating 2D Video Actors into a 3D AR System Georgia Institute of Technology. 2nd International Symposium on Mixed Reality, (2001).

MacIntyre, B., Lohse, M., Bolter, J.D., and Moreno, E. Integrating 2-D video actors into 3-D augmented-reality systems. Presence Teleoperators, (2002), 189-202.

MacIntyre, B., Bolter, J.D., Vaughn, J., et al. Three Angry Men: An Augmented-Reality Experiment In Point-Of-View Drama. IN PROCEEDINGS OF TIDSE 2003, (2003), 24 - 26.

Mooser, J., You, S., and Neumann, U. Real-Time Object Tracking for Augmented Reality Combining Graph Cuts and Optical Flow. 2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality, IEEE (2007), 1-8.

Prince, S., Cheok, A.D., Farbiz, F., et al. 3D Live: Real Time Captured Content for Mixed Reality. ISMAR'02 Proceedings of the 1st International Symposium on Mixed and Augmented Reality, (2002).

Rother, C., Kolmogorov, V., and Blake, A. „GrabCut": interactive foreground extraction using iterated graph cuts. ACM Transactions on Graphics 23, 3 (2004), 309.

Sharma, A.S., & Elidrisi, M. Classification of multi-media content (videos on YouTube) using tags and focal points. Unpublished manuscript, (2008).

Schmalstieg, D. and Wagner, D. Mobile Phones as a Platform for Augmented Reality. Proceedings of the IEEE VR 2008 Workshop on Software Engineering and Architectures for Realtime Interactive Systems, (2008), 43-44.

Snavely, N., Seitz, S.M., and Szeliski, R. Photo tourism: Exploring photo collections in 3D. ACM Transactions on Graphics 25, 3 (2006).

Wagner, D., Mulloni, A., Langlotz, T., and Schmalstieg, D. Real-time panoramic mapping and tracking on mobile phones. 2010 IEEE Virtual Reality Conference (VR), (2010), 211-218.

Wagner, D., Reitmayr, G., Mulloni, A., Drummond, T., and Schmalstieg, D. Pose tracking from natural features on mobile phones. 2008 7th IEEE/ACM International Symposium on Mixed and Augmented Reality, (2008), 125-134.

The columns on the last page should be of approximately equal length.