Teleseminar HWS 2014: Topics

Topic 1: Video Recording and Stabilization with Smartphones

Contact Person

Stephan Kopf (kopf@informatik.uni-mannheim.de)

Description

Nowadays, smartphones are widely used for video capture. A disadvantage of these small devices compared to camcorders is the large amount of shaking in captured video. Several video stabilization algorithms have been proposed to improve the quality of shaky videos. These algorithms typically require the analysis of the pixels in each video frame and to track the objects' positions over time. The computational effort of object tracking is very high and not applicable in smartphones.

Smartphones may offer additional information about the captured video. Motion sensors can be used to collect data about the orientation and viewing direction. The task is to analyze the shakiness of a video and compare it to motion sensor data. The students should evaluate if motion sensor data can be used to identify problematic video segments and also apply shake-reduction algorithms using this data to stabilize them.

Task

1. Use smartphones to acquire a reasonable set of videos and capture/store additional motion data with the build-in motion sensors.
2. Analyze available video stabilization algorithms.
3. Identify shaky video segment from the sensor data. It is not necessary to process and analyze the data on a smartphone, e.g., the algorithms may run on a PC.
4. Evaluate whether the precision of the sensor information is sufficient to improve the video quality by applying a video stabilization algorithm.
5. Write a 10 page report (IEEE style) which discusses the implemented application and the evaluation results.

Skills

JAVA or C++ programming skills

References

[1] First-person Hyperlapse Videos: http://research.microsoft.com/hyperlapse, 2014

[2] Matsushita, Y., Ofek, E., Ge, W., Tang, X., and Shum, H.-Y. 2006. Full-frame video stabilization with motion inpainting. IEEE Transactions on Pattern Analysis and Machine Intelligence 28, 7, 1150–1163.

[3] Puglisi, G.; Battiato, S. "A Robust Image Alignment Algorithm for Video Stabilization Purposes", Circuits and Systems for Video Technology, IEEE Transactions on, page(s): 1390 - 1400 Volume: 21, Issue: 10, Oct. 2011

[4] Chongwu Tang; Xiaokang Yang; Li Chen; Guangtao Zhai "A fast video stabilization algorithm based on block matching and edge completion", Multimedia Signal Processing (MMSP), 2011 IEEE 13th International Workshop on, page(s): 1 – 5

[5] Jinhai Cai; Walker, R. "Robust Motion Estimation for Camcorders Mounted in Mobile Platforms", Digital Image Computing: Techniques and Applications (DICTA), 2008, page(s): 491 – 497

[6] Wei Jiang, Dongqing Zhang, Heather Yu, “Sensor-assisted image deblurring of consumer photos on smartphones, ” IEEE International Conference on Multimedia & Expo (ICME) (to appear; we can send an email with the PDF)

[7] Hanning, Gustav, et al. Stabilizing cell phone video using inertial measurement sensors. In: Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on. IEEE, 2011. S. 1-8.

[8] Raimbault, Félix, and Yalcin Incesu. "Adaptive video stabilisation with dominant motion layer estimation for home video and TV broadcast." ICIP. 2013.

Topic 2: Visualizing Sensor Information of Mobile Video

Contact Person

Philip Mildner (mildner@informatik.uni-mannheim.de)

Description

When trying to retrieve a video from a large public event in near real-time it is crucial to make a useful selection before starting the actual upload process. In order to come up with such a selection, the first step is to gather as many information as possible on a video. When dealing with devices like smartphones or tablets, this is not limited to the video stream itself, but it also includes various sensor data. For example, data from an accelerometer can indicate whether a video involves a lot of unwanted shaking. The position in combination with the viewing angle can also provide useful information, e.g., to select the video that covers the best overview of a specific event. As all these data sets can change quickly, not only a single value should be recorded but a stream that covers the whole video.

The goal of this project is to develop an application for the Android platform that allows to capture a video along with all available metadata (sensor data, position, screenshot, etc.). The metadata should be stored in a well-defined format (e.g., XML along with image files) so that it can be used by other services. The gathered data should then be transferred to a server to be further processed. The uploading process should happen continuously, to that a low bandwidth video stream (e.g., only key frames) enhanced with available metadata is transmitted. Parts of the application may be derived from an existing application that has been developed during last year's seminar.

As mobile devices have limited computing power and limited battery life, the application should also take care of not gathering too much or redundant data. Therefore, an evaluation should be performed to find suitable values (interval, amount of data items, etc.) for gathering data from the various sources and how to transmit video and metadata to the server in a bandwidth-friendly way.

Task

1. Develop a well-defined format for providing metadata that describes the quality of videos (sensor data, position, screenshot, etc.) over time.
2. Implement an Android application that produces videos along with such metadata files.
3. Perform an evaluation to find suitable values for gathering values from the different data sources.
4. Tweak parameters of data gathering according to performed evaluation.

5. Implement a mechanism to transfer the gathered metadata to a server.
6. Write a 10 page report (IEEE style) which discusses the implemented application.

Skills

Java programming skills, interest in the Android platform

References

[1] Moo-Ryong Ra, Jeongyeup Paek, Abhishek B. Sharma, Ramesh Govindan, Martin H. Krieger, and Michael J. Neely. 2010. Energy-delay tradeoffs in smartphone applications. In Proceedings of the 8th international conference on Mobile systems, applications, and services (MobiSys '10). ACM, New York, NY, USA, 255-270.

[2] Zhang, B. et al. 2010. Annotating and navigating tourist videos. Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems - GIS ’10 (New York, New York, USA, 2010), 260–269.

[3] Lew, M.S. et al. 2006. Content-based multimedia information retrieval. ACM Transactions on Multimedia Computing, Communications, and Applications. 2, 1 (Feb. 2006), 1–19.

[4] Klaus Schoeffmann, Mario Taschwer, and Laszlo Boeszoermenyi. 2010. The video explorer: a tool for navigation and searching within a single video based on fast content analysis. In *Proceedings of the first annual ACM SIGMM conference on Multimedia systems* (MMSys '10). ACM, New York, NY, USA, 247-258.

[5] Beomjoo Seo, Jia Hao, and Guanfeng Wang. 2011. Sensor-rich video exploration on a map interface. InProceedings of the 19th ACM international conference on Multimedia (MM '11).

[6] Guanfeng Wang, Beomjoo Seo, and Roger Zimmermann. 2012. Motch: an automatic motion type characterization system for sensor-rich videos. In Proceedings of the 20th ACM international conference on Multimedia (MM '12)

[7] Jia Hao, Guanfeng Wang, Beomjoo Seo, and Roger Zimmermann. 2011. Keyframe presentation for browsing of user-generated videos on map interfaces. In Proceedings of the 19th ACM international conference on Multimedia (MM '11)

[8] Gonçalo Noronha, Carlos Álvares, and Teresa Chambel. 2012. Sight surfers: 360º videos and maps navigation. In Proceedings of the ACM multimedia 2012 workshop on Geotagging and its applications in multimedia(GeoMM '12).

Topic 3: Analyzing sensor data for video bitrate reduction

Contact Person

Philip Mildner (mildner@informatik.uni-mannheim.de)

Description

Video quality can be measured in different ways. Looking directly at the video data may produce very good results when examining brightness, shaking or composition, but this approach also comes with a lot of computational effort. A lot of data can also be retrieved from available metadata like sensor data, position or screenshots. By defining a metric for the different data sources, the video quality can be measured based on different targets like “most stable video”, “video with best view on certain location” or “three videos with best overview of a certain area”.

The goal of this project is to develop a web-based tool that displays a set of videos identified by low quality video data and available metadata. This tool then can act as the server component described in Topic 2. The metadata should be displayed in an appropriate format, e.g., a graph for sensor data, a map for positions and a slide show for screenshots/key frames. Along with displaying the videos, a quality metric should be developed that takes into account the various data sources of the metadata. The tool should support displaying data from fully uploaded videos as well as data from incoming continuous low bandwidth streams, so that a director can choose from available videos in real-time. Based on this metric, a mechanism should be implemented that allows users to select one or more videos that matches certain presets or user-defined optimization goals. The output of the algorithm should be either one video, a set of videos or even a set of video segments that then can be combined to an optimal video stream.

Task

1. Implement a web-based framework that displays a set of video streams by just using metadata like sensor data, position, screenshot, etc. as an interactive timeline.
2. Develop a quality metric derived from the video metadata.
3. Implement presets a user can choose from to interactively select the best combination of video streams for the given preset.
4. Write a 10 page report (IEEE style) which discusses the implemented application.

Skills

Programming skills, basic knowledge in web development

References

[1] Zhang, Y., Wang, G., Seo, B., and Zimmermann, R. Multi-video summary and skim generation of sensor- rich videos in geo-space. Proceedings of the 3rd Multimedia Systems Conference on - MMSys ’12, ACM Press (2012), 53–64.

[2] Zsombori, V., Frantzis, M., Guimaraes, R. L., Ursu, M. F., Cesar, P., Kegel, I., Bulterman, D. C. a. (2011). Automatic Generation of Video Narratives from Shared UGC. In Proceedings of the 22nd ACM conference on Hypertext and hypermedia - HT ’11 (pp. 325–334). New York, New York, USA: ACM Press.

[3] Junyong You, Andrew Perkis, Miska M. Hannuksela, and Moncef Gabbouj. 2009. Perceptual quality assessment based on visual attention analysis. In *Proceedings of the 17th ACM international conference on Multimedia* (MM '09). ACM, New York, NY, USA, 561-564.

[4] Philip Mildner, Frederik Claus, Stephan Kopf, and Wolfgang Effelsberg. 2013. Navigating videos by location. In *Proceedings of the 5th Workshop on Mobile Video* (MoVid '13). ACM, New York, NY, USA, 43-48.

Topic 4: Automatic Mobile Video Director

Contact Person

Stephan Kopf (kopf@informatik.uni-mannheim.de)

Description

Consider an event like a festival or concert with several thousand people. Many people may capture parts of the event with their smartphones. Due to the limited overall bandwidth, the upload of all videos is not possible during the event. This topic should consider the question which streams should be selected and uploaded to a central server in such a scenario.

The students' task is to develop an algorithm that rates the quality of a video on the fly. The algorithm should be very fast so that decisions can be made nearly in real-time. Without a decision, each device should transfer a low bandwidth stream to the server, so that a director can decide on which video to transmit in higher quality. Because video analysis algorithms take too much time on a smartphone additional motion sensor data should be used to make this decision. High quality videos (based on motion information) should then be transferred to a server as soon as the (automatic) director on the server has chosen a suitable stream.

Task

1. Use motion sensor data captured by a smartphone to classify the quality of a video.
2. Implement the data exchange between a server and several smartphones. The smartphones compute the video quality from sensor information and send it to the server. The server makes a decision about suitable video streams and sends notifications to the clients. Notified clients upload their videos to the server.
3. Write a 10 page report (IEEE style) which discusses the implemented application.

Skills

JAVA programming skills

References

[1] Seshadri, P.; Chan, M.; Ooi, W.; Chiam, J., "On Demand Retrieval of CrowdSourced Mobile Video," Sensors Journal, IEEE (available online).

[2] Xuan Bao and Romit Roy Choudhury. 2010. MoVi: mobile phone based video highlights via collaborative sensing. In Mobile systems, applications, and services. ACM, New York, NY, USA, 357-370.

[3] Pieter Simoens, Yu Xiao, Padmanabhan Pillai, Zhuo Chen, Kiryong Ha, and Mahadev Satyanarayanan. 2013. Scalable crowd-sourcing of video from mobile devices. In Mobile systems, applications, and services. ACM, New York, NY, USA, 139-152.

[4] P. Jain, J. Manweiler, A. Acharya, and K. Beaty, FOCUS: Clustering Crowdsourced Videos by Line-of-sight, in ACM Conference on Embedded Networked Sensor Systems, 2013.

[5] P. V. Seshadri, M. C. Chan, W. T. Ooi, and J. Chiam, On Demand Retrieval of CrowdSourced Mobile Video, National University of University, Singapore, Tech. Rep., April 2014. Available: http://www.comp.nus.edu.sg/%7epadmanab/movisodetechreport.pdf

[6] Engström, A., Zoric, G., Juhlin, O., & Toussi, R. (2012). The Mobile Vision Mixer: a mobile network based live video broadcasting system in your mobile phone. In Proceedings of the 11th International Conference on Mobile and Ubiquitous Multimedia (MUM ’12) (pp. 3–6).

[7] Shrestha, P., de With, P. H. N., Weda, H., Barbieri, M., & Aarts, E. H. L. (2010). Automatic mashup generation from multiple-camera concert recordings. In Proceedings of the international conference on Multimedia - MM ’10 (p. 541). Firenze, Italy.

[8] Cricri, F., Curcio, I. D., Mate, S., Dabov, K., & Gabbouj, M. (2012). Sensor-Based Analysis of User Generated Video for Multi-camera Video Remixing. In K. Schoeffmann, B. Merialdo, A. Hauptmann, C.-W. Ngo, Y. Andreopoulos, & C. Breiteneder (Eds.), Advances in Multimedia Modeling (Lecture No., pp. 255–265).

[9] Cricri, F., Roininen, M., Leppanen, J., & Mate, S. (2014). Sport Type Classification of Mobile Videos. IEEE Transactions on Multimedia, 16(4), 917–932.