Multimodal data acquisition on mobile devices

Aleš Láník, Jozef Mlích, Pavel Zemčík, Roman Juránek, Petr Chmelař
[ilanik, imlich, zemcik, ijuranek, chmelarp]@fit.vutbr.cz

The proposed paper deals with acquisition of data using mobile devices in order of its consequent processing, analysis and browsing. Mobile devices provide wide variety of stream data, for example audio-visual, accelerometer, location, wireless status and other data available on a particular device.

The motivation of the presented work is to acquire and provide information that is not available using other sensing methods. For instance, users may want to organize photos and videos not only according to the time, but also geographical position or according to the objects they contain. It can be e.g. castle, restaurant or a complex traffic crossing, similarly to Google Goggles, that different kinds of objects and places can be visually identified, using domain-dependant pictures only.

From the technological point of view, we need an appropriate data set for computer vision and image processing. An example of processing is creation of complex panorama photos, 3D objects reconstruction or classification from visual, position and motion information together. Moreover, camera motion estimation, moving object detection and camera lens intrinsic parameters may be estimated and refined from the data, either annotated or not. We suppose, the additional information to the visual can improve the performance and precision of these methods. The collected data should also provide information for general and spatio-temporal data mining challenges such as traffic density, closed road or Wi-Fi networks coverage.

For the above purposes, we attempt to record all information available on a particular device. Information, such as latitude and longitude, elevation, acceleration of gravity, directional information using compass, Wi-Fi network identifiers (SSIDs) and signal strength, information from GSM transceiver station (BTS) and, of course video. If a possibility exists to make use of GPS data and accelerometer, the position and orientation of the device can be estimated in 3D more precisely. This estimation is still not enough accurate to be used in e.g. medicine, but it is sufficient for other applications, where the directional or pointing information is interesting, e.g. for navigation guide, or a real-world played games (as Mafia). The refined spatio-temporal information may be also used to design a system with the necessarily synchronized collaborative environment, for instance for fire and rescue units.

Additionally, the obtained data may be extended by user-defined meta-data and other information using WWW – e.g. Google Maps, Picasa, Wikipedia, or Google Labs' Similar Images. Such approach could also improve the user experience in tasks requiring to browse (own history) related multi-modal data. Talking about any kind of forgetting, as the Alzheimer's disease, the application can help the user to navigate back home or to provide the necessary information to other attendants.

The information is acquainted using streaming protocols (RTSP), containing video information using MPEG-4 (3GP) and MPEG-7 (Multimedia Content Description Interface) for streaming the related information to a server. The visual information may be then extended by low level image information, i.e. dominant points descriptors, color and texture features supplemented

by high level information i.e. detection of certain objects (e.g. cars, signs, restaurants), events (doors opening, cars running) and concepts (indoor or outdoor scene) using an experimental software similar to that we developed for the TRECVid (NIST TREC Video Retrieval Evaluation) workshop in recent years. For such purpose, we use a database system capable of information retrieval functions and using image (features') similarity. Moreover, we plan to build a simple client (browser) of the multi-modal information recorded.

Current mobile phones using operating systems seems to fulfill the hardware requirements for use cases mentioned above. The hardware running operating systems, namely Android, Maemo, MeeGo, Windows Mobile, etc., allows complex accelerated data processing as is required using optimized processors as FPU, SIMD or DSP unit. The other algorithms, e.g. computer vision and object classification, which we cannot compute even using desktop computers are delegated over a network to remote servers (grid or cloud).

The paper will contain a description of the data acquisition implementation. Next, the synchronized complex data browser and experimental results achieved on the acquired data set.