https://doi.org/10.2352/ISSN.2470-1173.2017.19.AVM-012

Free-view multi-camera visualization and harmonization for automotive systems

Vladimir Zlokolica, Brian Deegan, Patrick Denny, Mark Griffin, Barry Dever; Valeo Vision System; Tuam, Galway, Ireland

Abstract

In this paper, we present framework for visualization of the vehicle-surround-views that include multiple cameras attached to the car exterior. The proposed framework harmonizes the input camera images in terms of brightness, colour and other related properties to enable advanced visualization, where the displayed image looks as it would have been captured by a single camera positioned at an arbitrarily chosen 3D point and oriented in 3D space around the vehicle. The rendering and harmonization framework is a hybrid based scheme that performs both adaptive camera tuning and post processing of the camera images. We discuss both algorithmic and implementation aspects of the image quality module within which the framework is designed. The algorithms involved in the image quality module perform camera image processing, which includes both image analyses and image post-tuning. In addition to algorithmic aspect of the framework, this paper also discusses real-time implementation aspects related to different embedded systems presently used in automotive systems.

Introduction

Multi-camera systems, in general, have become important topics of research in the field of computer vision and computer graphics. Automotive surround view systems pose a particular challenge, because they are increasingly used to perform both scene visualization and computer vision tasks. The main challenges are to optimally design the multi-camera configuration, while simultaneously optimizing individual camera performance independently to extract meaningful information, and/or combine it to display advanced views with enhanced overall visual quality and user experience.

In existing multi-camera automotive systems [1, 3, 6] different types of views are generated using multiple camera video inputs. In such visualization framework, the raw camera images (from different cameras) are first projected to the target surface (such as flat 2D plane bowl) and subsequently merged and rendered to the view-port by an arbitrarily chosen virtual camera. As a result, a panorama-mosaic image/video is obtained. One specific example is TopView, also referred to as bird view, where the images are projected to the 2D flat plane and the virtual camera is positioned exactly above the car and parallel to the car (see Fig. 1). On the other hand, a more advanced view, such as Bowl-View, for the same scene and the same four camera configuration system, is shown Fig. 2, where the 3D scene is acquired by a virtual camera positioned at an arbitrarily (referred to as free-view) chosen position in surrounding 3D space.

In the recent past, a considerable number of solutions have been proposed for multi-camera automotive systems [1, 2, 3, 4, 5], which aim at visualization of the 360 degrees vehicle surrounding area. The main purpose of this advanced visualization is to provide driver assistance within the emerging ADAS systems. In order to acquire as much as possible information about the vehicle 3D surrounding area, an increasing number of cameras are being included in the vehicle. Current solutions tend to include 4-6 outdoor cameras, but this number is likely to increase in future vehicles [7].

Currently, the most mature multi-camera view in the automotive marketplace is TopView - Surround View family of views, where the projection surface is flat and the virtual camera is placed above the car and parallel to it. It has been proven that using 4 camera images attached to the vehicle, one camera on the left side, one of the right side, one in front and one at the rear, one can generate a relatively accurate image mosaic that represents a view from above the car [1, 2, 3], as shown in Fig. 1. However, in most

Figure 1. *Example of Surround TopView using 4 surround cameras.*

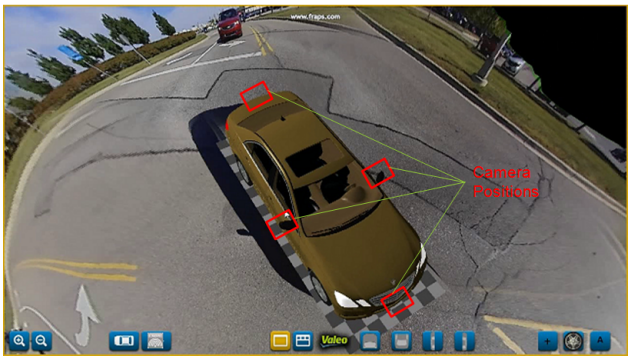

Figure 2. *Example of Free-View BowlView for camera configuration consisting of 4 cameras positioned on the vehicle.*

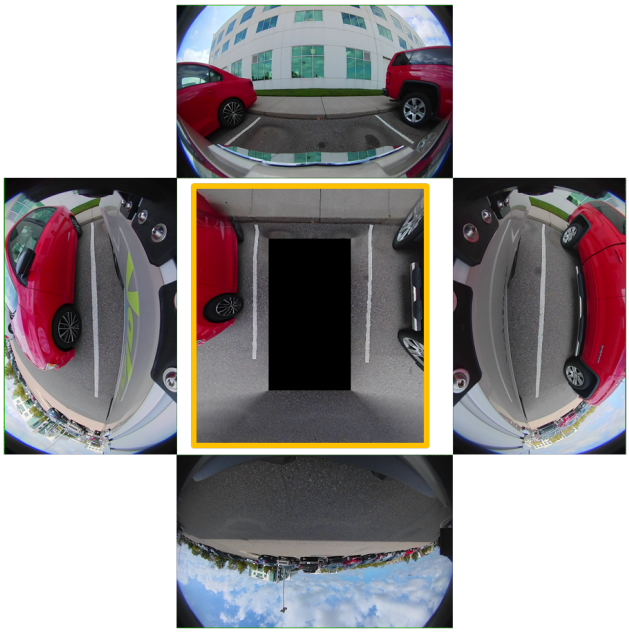

Figure 3. *Four input camera images and corresponding TopView image.*

existing (mostly commercial) real-time solutions there are still a number of issues that introduce artefacts in the final image that prevent it from being optimal in terms of visual quality, i.e., to be as it was acquired by ”real” single camera positioned above the car. These artefacts are: geometric misalignment of camera images, image intensity property variations (from different cameras) in brightness and colour hue, and other image quality properties, such as noise, sharpness and contrast.

In Fig. 3, an example of a TopView image is provided along with the 4 input camera images, where the vehicle surrounding also includes neighbouring 3D objects in addition to the flat ground area immediately surrounding the vehicle, which is usually only considered in existing literature [1, 2, 3]. Since the calibration and geometric image alignment is done for the ground plane, the ground plane image content is aligned to a certain extent but still the neighbouring objects are usually not in overlapping areas and in non overlapping areas they appear unnatural. This could potentially be solved by employing more than four cameras and performing additional sparse 3D cloud reconstruction but this adds additional costs into the system and introduces real-time issues in current embedded automotive systems.

In addition to geometric alignment, pixel intensity harmonization is another important factor in obtaining optimal visual quality of the TopView. The pixel intensity harmonization is mainly related to brightness (luma) and colour hue alignment between different camera images within the TopView. As can be seen in Fig. 3, the global brightness level of the neighbouring camera images differs considerably and therefore degrading the overall visual quality of the final TopView image. Specifically, in this example, the left and right camera images are of a much higher global brightness level than the front and rear camera images. The reason for this is the 3D content at which each camera is looking at as well as the direction of incoming outdoor light. Based on the surrounding 3D content, each camera adapts/tunes its parameters independently. As a result different global brightness level or colour hue can occur, which even in case of small differences is visible in ”overlapping” regions, where the two neighbouring cameras images merge. Such an artefact can be mitigated either by multi-camera tuning (from embedded system to each camera), post processing on the embedded system level or in best case combined (hybrid approach).

Finally, one of the most important and demanding objectives in automotive multi-camera systems is real-time processing (in current systems output video must be rendered within frames of 33ms) on limited embedded system resources available on the market that offer the best performance with minimal costs and aggressive design cycles. Besides the cost, currently available automotive embedded systems are constrained by specific automotive requirements in terms of temperature, safety issues and other specific automotive standards. While the emerging embedded system tend to increase their computational power each year, the resolution of the input cameras and performance demands are also increasing. Consequently, the overall system architecture, as well as each computer vision algorithm/block needs to be carefully designed, implemented and optimized.

In this paper, we discuss concepts in automotive multi-camera systems and propose multi-camera free-view visualization framework. Within this framework, we mainly focus on brightness and colour harmonization within the multi-camera view system. Additionally, we also propose and explain camera tuning and adaptation scheme that is applied in particular system modes; this mainly concerns different light modes, as an example. After that we detail on implementation aspects of the multi-camera system and provide experimental results of the proposed system. Finally, we provide conclusion along with additional discussion.

Multi-camera free-view visualization framework

In this Section, we discuss general multi-camera visual system applied for different views and propose a view-independent framework that is semi-automatically adaptive to different views and case scenarios.

Generally, each multi-camera framework consists of camera interface, camera image processing block and rendering block used for displaying particular view on the output, i.e., on the head unit in the vehicle. Each particular view has specific camera configuration and view-port parameters. While camera configuration parameters describe the camera set and positions of the cameras, the view-port parameters specify the view type and how it is set to the display screen.

The block diagram of the proposed multi-camera framework is shown in Fig. 4. As can be seen in Fig. 4, the input camera images are fed into the video processing and texture generation block, for the given view-port parameters. The pre-processing of the input camera video is applied to improve visual quality of the projected textures in terms of noise, sharpness and contrast.

The most common example of the pre-processing applied to input camera images is temporal filtering for de-noising in case of low light video sequences, where the noise level increases significantly. Therefore, in order to improve final surround view visual quality, the noise has to be reduced in input images prior to the texture projection (to the selected plane - in case of TopView this is 2D flat ground plane); otherwise the noise gets spatially

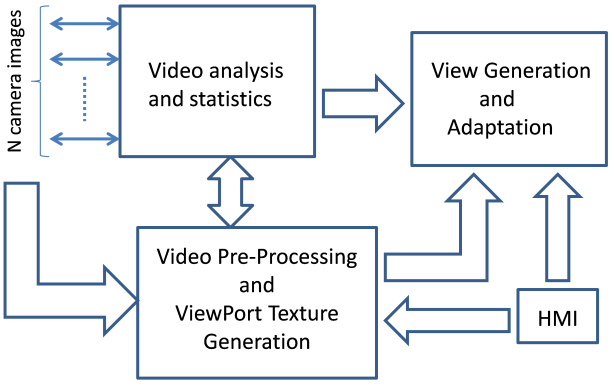

Figure 4. *Block scheme of the proposed multi-camera framework (HMI - Human Machine Interface).*

stretched due to perspective correction, and forms spatially correlated ”blob-like” noise patterns, which are very difficult to remove.

Further on, the pre-processing step, consisting of different image pre-processing tasks, can also be divided between the camera post-processing unit and the ECU pre-processing units, because in most of the cases these are camera independent tasks, such is the case in de-noising. Specifically, part of the image pre-processing can be done within the camera chip-set that is practically done through camera parameter tuning, through which certain spatial and/or temporal filtering can be applied in particular modes. However, the available filters within the camera are often of limited performance (due to runtime and memory bandwidth restrictions), and such they can be incorporated only to limited extent.[1] The rest of the necessary processing has to be done on the ECU where a more customized solution , with better performance is possible.

Camera Tuning and Camera Based Visual Adaptation

Camera tuning for surround view applications presents additional challenges. In terms of harmonizing brightness and colour for TopView images, there are, broadly speaking, two approaches that may be pursued. The first is to have a centralized camera control architecture. In this design, image parameters are extracted from each of the raw input streams (e.g. current exposure time, average brightness, colour etc.) are fed into a central algorithm, which then sets the exposure time, gains, and colour corrections to common values on all input cameras. While in theory, this will provide optimal harmonization, there are a number of problems with this architecture. In the case of TopView images, only a small region of each camera input is cropped and used for the final output image. If the camera parameters are chosen to harmonize for this small section of the image, there is a significant risk that the image quality of the rest of the image is affected. This is particularly relevant in cases where images from a single camera are used for multiple views, or as input for both viewing and machine vision. Even more fundamentally, such a system must have aggressively small latencies in the camera sensor - central ECU control loop to respond to a quickly changing environment and this places significant demands on the bus systems and on the prioritisation of control loop handling software on both the sensor and the central ECU. Also, the trend in the automotive industry is towards high dynamic range (HDR) imaging. HDR is useful in automotive, because of the challenging nature of many automotive use cases (e.g. headlights at night time, or entering/exiting a garage). Currently, a number of different HDR schemes are available for automotive, and each of these require different signal processing and control algorithms. A central camera control algorithm would have to set numerous parameters on each camera, and would have to be re-designed for each new sensor and related Image Signal Processing (ISP) that becomes available. This is an impractical approach.

A preferable solution is to allow each camera to adapt to its individual scene independently, and to harmonize the image in terms of brightness and colour in post processing. There are some additional steps that can be considered at the camera level, which can aid harmonization. A good example is image contrast. Most Image Signal Processing (ISPs) on the market employs contrast enhancement algorithms, including histogram stretch and equalization. The degree of contrast enhancement applied is typically dynamic, and depends on the statistics of the image. In a TopView configuration, each camera ”sees” a very different scene, and will therefore have different levels of contrast enhancement applied. This will hinder efforts to harmonize a TopView image later in the video chain. The simplest solution to this problem is to disable contrast enhancement at the camera ISP level, and to apply contrast enhancement on the combined topview image later in the video chain.

Similarly, tone mapping also presents a problem. If you consider the example of a multi-capture HDR scheme, 2 or more images are captured, combined, and then typically tone mapped down to an 8-bit or 10-bit bit image for display. Again, in a TopView configuration, one camera may view a HDR scene (e.g. dark shadows and a low sun), and another camera may face a low dynamic range scene (e.g. a dark, unlit garage). In this use case, both cameras will apply different tone mapping to the image. This can be very difficult to correct when using post-processing to harmonize the image. One solution is to linearize the input images by reversing the tone mapping applied, performing harmonization on the linearized image, and then tone-mapping the combined image. Such a process is complex and requires significant processing and memory.

Another use case to consider is adaptation to low light. It is not uncommon for ISPs to dynamically adjust brightness, colour saturation, denoise, edge enhancement and other parameters for low light scenes. There are several use cases for TopView systems whereby the individual cameras see scenes with vastly different brightness levels. In such use cases, two cameras may, for example, have different colour saturation levels - the same object may appear colourful in one view, and desaturated in another. Similarly, if different denoise levels are applied, an object may appear blurred in one view, and sharp and noisy in another view. It is therefore quite important to take into account how ISPs adapt to low light scenes when optimizing a surround view system.

[1] In Subsection ”Camera Tuning and Camera Based Visual Adaptation” we go more into details about camera tuning and application of particular algorithms within the camera chip-set.

Rendering different multi-camera views

After the camera tuning and camera image pre-processing tasks are performed, the next step is to generate textures corresponding from each camera that will be used to build the selected multi-camera view. Each view, as previously mentioned in introduction, is defined by view-port parameters and camera configuration (intrinsics, extrinsic and lens distortion parameters).

In current automotive surround view systems, the most common type of cameras used are wide-lens cameras, which can maximise the field of view for a constrained number of cameras attached to the vehicle. However, such cameras introduce undesired radial distortion into the input camera images, which have to be subsequently rectified to perform suitable distortion correction [8, 9]. However, first the calibration of cameras must be done to obtain camera intrinsic and extrinsic parameters, as well as lens distortion model.

Calibration can be executed in many ways. The calibration can be done either off-line using specifically designed calibration pattern placed on the ground beneath vehicle or automated that is performed based on the surrounding 3D content. While off-line calibration using predefined pattern is considered more reliable it can be impractical in certain situation and while driving cameras attached to vehicle can move slightly. Consequently, automatic calibration based on surrounding feature extraction and matching is considered more suitable in real-case scenarios. In our approach, we use automated calibration which we do not describe in this paper.

Once the calibration is done, reverse texture mapping is performed starting from the predefined view-port parameters for the selected view (via Human Machine Interface - HMI). Using the view-port parameters, screen coordinates are converted to 3D world coordinates after which the 3D world points are converted to the fish-eye corrected coordinate, which corresponds to the input camera image. As a result, we generate projected textures for each camera image corresponding to selected view-port. Finally, the generated camera textures are tone mapped (as explained in the next section) and merged via blending to obtain a particular view, as depicted in Fig. 4 in block "View Generation and Adaptation".

Visual Quality Optimization via Post Processing

A multi-camera system consists of more than one camera, each exposed to different 3D area and content, as well as exposed to different outdoor illuminations and colour temperatures. As such, each camera automatically aims at adapting to the scene by adjusting its camera parameters, such as exposure, gain and Auto White balancing (AWB). This results in variable brightness and chrominance hue introduced relatively between cameras. Looking at the acquired camera images the difference in chroma and luminance might not be so apparent; however when stitched together the difference becomes more visible and unpleasant for human visual system, especially in merging areas and nearby areas.

Brightness and chrominance alignment, i.e., "harmonization", can mainly be done in two ways: (i) through multi-camera re-tuning and adjustment, (ii) as a post-processing step via multi-camera tone mapping, or through combined via hybrid approach. In case of multi-camera re-tuning, parameters are monitored for all cameras and certain statistics is computed on such outputs. Based on the calculated statistics, the camera parameters are tuned from ECU to camera, via specifically designed interface. In such case, a specifically designed two-direction ECU-camera communication interface has to be implemented, which might be problematic due to difficult constraints such as exact camera synchronization and time delay - camera communication does not fall into the first-class priority task in real-time embedded systems.

Another way to perform photometric alignment is within the projected texture post-processing by tone mapping, as proposed in Fig. 4. This step is usually a part of "View Generation and Adaptation", shown in Fig. 4. As can be seen in Fig. 4, the input to this block is "Video analysis and statistics", besides input camera projected textures. Specifically, the input camera raw video frames are processed and analysed within the ROI corresponding to the chosen surround view. The exact ROI positions and size, for input camera images, are computed using specifically designed algorithms, which determine input image areas most applicable to statistics computation for brightness and chrominance harmonization.

The extracted brightness and chrominance values for the ROI camera textures are then transformed in a specific domain in which statistics can be computed more precisely, reliably and computationally less expensive. The output of this block are correction values for luminance and chrominance components. These correction values have their global (fixed for whole camera image) and local (varies for different spatial positions within the image) parts that are estimated independently. The final correction values are obtained as a combination of global and local parts, and applied via tone mapping within the View Generation block.

In the proposed approach we combine both mentioned approaches, where the second one (post-processing) is the main (fine tuning) and camera-tuning based one is used only for rough photometric adjustments (fine camera tuning in real-time is difficult via this approach).

Real-time Implementation aspects

Multi-camera system concepts and related computer vision algorithms have been available for a considerably long time. However its real-time implementation on target platforms, with limited resources, have always been a bottle neck. Since automotive multi-camera systems are greatly limited by availability and appropriately strong enough target platforms, implementation aspects of the automotive multi-camera systems is very important. The complete system design has to be carefully designed to not only support visualization but all supporting computer vision tasks, such as pedestrian detection, lane sensing, etc. In extremely high demanding and such complex environments all processing blocks have to be optimally designed and computationally optimized to meet computation and memory limitations.

Since recently, most of the automotive multi-camera systems have been based on DSP-only-based embedded platforms, where several DSP units (of different types) were available. In such systems each DSP is allocated to a particular set of tasks which are performed in a particular order, set by the priority list. Currently, new automotive target platforms for multi-camera systems include advanced GPU units which are shown to significantly boost processing power, especially for rendering multiple different views in real-time. Namely, one of the main drawbacks of only DSP based systems was difficulty to render in real-time

lots of different views in a short time instance, which is relatively easily done on GPU.

In the proposed visualization framework, the system (shown in Fig. 4) is implemented on a hybrid platform consisting of DSP and GPU units, along with other supporting parts. All processing related to harmonization and visualization is implemented on GPU, while all other computer vision tasks and certain additional pre-processing is done on DSP.

Experimental results

In the experimental results section we first demonstrate harmonization performance in topView, in "high light" case corresponding to the day light outdoor scene. After that we present TopView results of the de-noising and harmonization in "low-light" case that corresponds to the darker condition in night time. Finally, we also present harmonization results for additional view (in addition to TopView and Bowl View), referred to as front View Overhead.

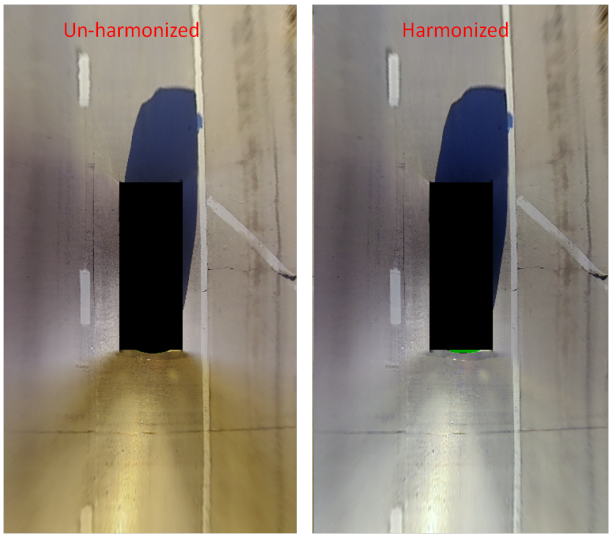

Figure 5. *TopView example 1, where on the left hand side is example of un-harmonized view and on the right hand side is the harmonized view by the proposed scheme.*

In Fig. 5 we show colour correction and harmonization for the TopView, where it can especially be seen that the colour hue of the Rear Camera has been removed efficiently and the rest of the TopView image is harmonized in terms of the colour hue. Next, we show in Fig. 6 the results for the input camera images shown in Fig. 3. In this example the brightness correction harmonization is shown to be efficient, even in this more complex vehicle surrounding where 3D objects are also present. Namely, the neighbouring vehicles with strong colours and brightness introduce considerable difficulty in brightness and colour harmonization, where the proposed algorithm is shown to perform well.

In case of low light case, we present example in Fig. 7 of one selected scene and show experimental results for the TopView projection with different algorithms applied. The upper left part of the image represents topView in low light without harmonization and without camera tuning for de-noising. After that, right-up, we show the same view with camera-tuned based de-noising, while in bottom left we show case when both harmonization and

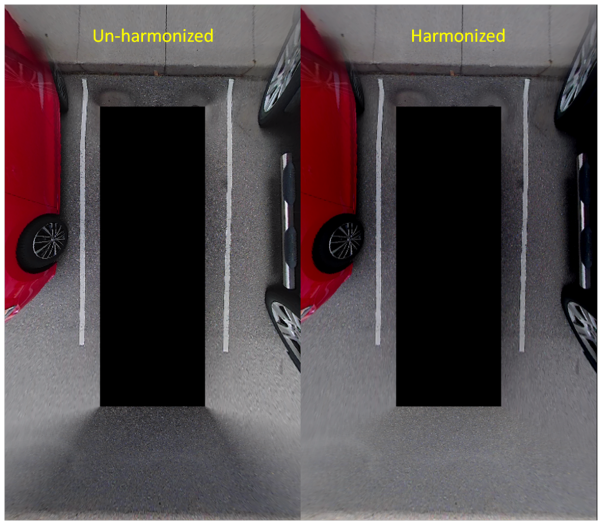

Figure 6. *TopView example 2, from Fig. 3, where on the left hand side is example of un-harmonized view and on the right hand side is the harmonized view by the proposed scheme.*

camera-tuned based de-noising is applied. Finally, we show in bottom right corner the case where in addition the temporal filtering is applied for de-noising also on ECU in the pre-processing and texture generation block, as shown in Fig. 4.

Additionally, we show one example (see Fig. 8) of different view - Front View Overhead, on input camera images ("high light case") shown in Fig. 3, where the same corrections have been applied as in the TopView case. This represents a proof of concept that statistics and corrections can be computed view-independently and applied subsequently on different views.

Conclusions and Discussions

It is clear that the multicamera projections have demonstrated a progression, from relatively straightforward projections with non-HDR cameras, to the more complex automatic vision systems that react to safety related, rapidly changing, high dynamic range scenes with low light content. All such systems should be considered as intuitive presentations of the environment to a human. Thus the evolution of such systems has been driven by a need to maximize the fidelity, utility and naturalness of the visual presentation of a vehicle's environment to a user. As the technologies have become available, the proportion of use cases for the user has been able to expand significantly. Progressions of this in the future are the basis of considerable research (and indeed intellectual property protection) and focus on enriching the user's proprioception in the vehicular environment. This includes all aspects of the sensor and system enhancements, with a limiting factor being an overprovision of information and detail which would reduce information to data overload which the human perception system is vulnerable to. So considerable end user testing is required in order to maximize the benefits of the technologies concerned.

References

[1] Chien-Chuan Lin, Ming-Shi Wang, "A Vision Based Top-View Transformation Model for a Vehicle Parking Assistant", Sensors,

Figure 7. *Results for TopView image in low light scenario: (i) un-harmonized (left-up); (ii) un-harmonized but with camera based temporal filter (right-up); (iii) harmonized and camera based temporal filter (left-bottom); (iv) harmonized, camera based temporal filter and ECU temporal filter.*

No.12, p. 4431-4446 (2012).

[2] B. Zhang, V. Appia, I. Pekkucuksen, A. U. Batur, P. Shastry, S. Liu, S. Sivasankaran, K. Chitnis, Y. Liu, "A Surround View Camera Solution for Embedded Systems", IEEE Conference on Computer Vision and Pattern Recognition Workshop, CVPRW, (2014).

[3] V. Appia, H. Hariyani, S. Sivasankaran, S. Liu, K. Chitnis, M. Mueller, U. Batur, G. Agarwal, "Surround view camera system for ADAS on TIs TDAx SoCs", Texas Instruments, 2005.

[4] M. Yu, G. Ma, "360 Surround View System with Parking Guidance", SAE Int. J. Commer. Veh. 7(1):19-24, 2014.

[5] Y.-C. Liu, K.-Y. Lin, and Y.-S. Chen, "Birds-Eye View Vision System for Vehicle Surrounding Monitoring", Springer-Verlag, RobVis 2008, LNCS 4931, pp. 207218, 2008.

[6] M. Sainz, T. Stich, "Next Generation Surround-View for Cars", GPU Technology Conference (GPC), 2015.

[7] U. Ayyer, S. Husain, T. Jiang, S. Petrovic, A. Tolani, "Self-Driving Cars: Disruptive vs Incremental", Applied Innovation Review, Issue 1 June 2015.

[8] C. Hughes, M. Glavin, E. Jones, P. Denny, "Wide-angle camera technology for automotive applications: a review", IET Intelligent Transport Systems, 2008.

[9] M. Friel, C. Hughes, P. Denny, E. Jones, M. Glavin, "Automatic calibration of fish-eye cameras from automotive video sequences", IET Intelligent Transport Systems, 2009.

Author Biography

Vladimir Zlokolica PhD degree in Applied Sciences at Ghent University, Belgium in 2006, where he also worked as post-doc researcher until 2007. Since 2007 he has worked for Micronas/RT-RK and Valeo Vision Systems, as an video architect/algorithm developer and senior vision research engineer, respectively. Since 2008 he also holds Assistant Professor position at University of Novi Sad, Serbia. His main expertise and interests are in domain of video processing, camera imaging and computer vision.

Brian Deegan received a PhD in Biomedical Engineering from the National University of Ireland, Galway in 2011. Since 2011 he has worked in Valeo Vision Systems as a Vision Research Engineer focusing on Image Quality. His main research focus is on high dynamic range imaging, topview harmonization algorithms, LED flicker, and the relationship between image quality and machine vision. Since 2014 I have been an official Valeo Expert on Image Quality.

Patrick Denny received his PhD in Physics in 2000 from the National University of Ireland, Galway, where he is also an Adjunct Professor of Automotive Electronics. He is a Senior Research Engineer and a Valeo Senior Expert and has worked for the last 15 years at Valeo Vision Systems and its previous incarnation, Connaught Electronics Limited, initially as the Team Leader of RF Design, before moving into the development and innovation associated with automotive vision systems. His research interests include several aspects of automotive vision system image quality, components, algorithmic design, systems and more recently data analytics.

Mark Griffin received his Bachelor Degree in Applied Physics from University of Limerick in 2010. From 2010 to 2011 he worked as an engineer in ASML, Holland. Since 2011 he is working for Valeo Vision Systems, Ireland, as a Vision Research Engineer in Image Quality Group. His work mostly concerns video algorithm design and development, camera imaging and visualization system development for automotive systems.

Barry Dever received his Bachelor Degree in Electronic Engineering from Limerick Institute of Technology in 2003. From 2004 to 2011 he worked for General Electric in the development of their Security Applications. Joining Valeo in 2011 he has lead teams in the research and development of high precision Active Alignment platforms and real time Image Quality tools and is officially recognised as a Valeo Expert in this area since 2014. He currently leads the Valeo Vision Image Quality team from the Research and Development headquarters in Galway, Ireland.

Figure 8. *Front Over View example, from Fig. 3, where on the upper image corresponds to un-harmonized view and the bottom one corresponds to the harmonized view by the proposed scheme.*