

A software video stabilization system for automotive oriented applications

A. Broggi, P. Grisleri
Dipartimento di Ingegneria dell'Informazione
Universita degli studi di Parma
43100 Parma, Italy
Email: {broggi, grisleri}@ce.unipr.it

T. Graf, M.Meinecke
Electronic Research
Volkswagen AG
Wolfsburg, D-38436, Germany
Email: {thorsten.graf, marc-michael.meinecke}@volkswagen.de

Abstract—Vision applications in the vehicular technology field can take big advantages by electronic stabilization of video sequences since it reduces calibration based feature measurement and tracking errors. In this paper a new video stabilization system expressly designed for automotive applications is presented. Image correction is fast and can be included in real time applications. The system has been implemented on one of the vehicles in use at the Department of Information Technology of the University of Parma and tested in a wide range of cases. A test using a vision based pedestrian detector is presented as case study showing promising improvements in detection rate.

Keywords: Video stabilization, Vision, Automotive applications.

I. INTRODUCTION

Vehicular and automotive vision applications [1] are usually affected by serious problem related to vibrations and oscillations transmitted to the camera mainly from the road and vehicle bumpers. These problems are usually connected to the tight dependence between camera movements and the calibration parameters used for measurements. In most cases - such as obstacle detection, pedestrian detection [2] or lane detection - some features are extracted from the image and, using calibration parameters, their position and/or size is computed in the 3D world. Any moving vehicle is affected by oscillations. This causes a continuous variation of camera orientation angles (with respect to a fixed reference system) that are used for distance computations. The problem becomes even harder on off-road tracks, where road coarseness causes larger oscillations. However, during normal driving on urban roads or on highways, pitch variations are dominant on roll and yaw. Systems for video stabilization can partially cope with this problem by compensating camera movements. In order to achieve this goal, it is necessary to estimate or measure camera orientation and possibly its position.

II. STATE OF THE ART IN VIDEO STABILIZATION

In the literature several approaches to the video stabilization problem are available.

Some systems [3] use sensors different from vision to detect camera movements, such as inertial systems. This type of stabilizers generates a robust result with a low delay (the acquired frame is directly stabilized), since the correction to be applied is measured. The drawback of this approach is the

cost: inertial measurement systems are not yet suitable for integration on commercial vehicles, since their cost is about 500 USD and their usage is mainly limited to avionics or military applications. Other systems [4] use the movement of a group of lenses in order to obtain an optical stabilization. This kind of stabilization is particularly good for handheld cameras but presents some disadvantages in correspondence to fast movements. The reaction time is limited by physical factors and cannot be too quick. This approach is suitable for low cost applications, but integration in industrial optics is needed.

Vision based systems have also been proposed [5]. Systems implementing stabilization based on feature correlation across subsequent frames are also available on the market. Preferred application fields for these devices are in handheld cameras and security field. A typical 2-3 frames delay is introduced and, for external units, resampling noise may be critical for image analysis.

Automotive applications are critical for feature extraction methods because the background is continuously changing not allowing to use entire features found in a certain frame for a matching in the subsequent ones. The proposed system is specifically designed to handle vertical camera movements, which represent the most common oscillation in the automotive field. In this work a software system for pitch estimation is proposed. The system can estimate camera pitch variations extracting a signature from horizontal edges. The signature is compared to the current mean signature and the pitch information is extracted. The algorithm is also able to evaluate if there is not enough information for pitch estimation. This usually happens when the vehicle is approaching a steep hill or is steering.

III. DATA CHARACTERIZATION

The input image is acquired using a camera mounted on a vehicle as shown in figure 1.

The stream of frames produced by the camera has a resolution of 320x240 pixels. Each pixel is described by an 8 bit intensity value. The camera is analog and based on the NTSC standard (30 fps). Due to vehicle motion, an image sequence presents three different kinds of transformations.



Fig. 1. The test vehicle with different camera mounting examples: front camera above the license plate and two cameras on the top of the vehicle.

- *Perspective zoom.* The image appears to be continuously zoomed in the motion direction. Far objects get closer and change their position and size.
- *Horizontal movements.* If the vehicle is steering, all the points in the image move in the opposite direction. This movement is typically horizontal, with some occasional roll due to bumpers deformation.
- *Vertical movements.* When the vehicle approaches a hill or hits a bump all the points in the image move in the vertical direction.

The main problem that may be observed in image sequences taken from real roads is due to the presence of road coarseness (potholes, speed bumpers). These road defects cause oscillations in the camera pitch value. The oscillation amplitude is also dependent on the vehicle speed bumpers behavior. Camera calibration is necessary to obtain obstacle distance measurement using vision. During normal driving, grabbed images are mainly affected by pitch and roll camera orientation variations with respect to the original calibration setpoint. These variations prevent vision algorithm based on calibration to work correctly. During urban and highway driving it can be assumed that roll variations are neglectable. Thus the system developed deals with pitch variations only.

Images must be acquired using fast a shutter time (below 1/200 s) in order to avoid motion blur problems which prevent the feature extraction process to work.

IV. THE ALGORITHM

The proposed stabilization algorithm scheme is depicted in figure 2. Specific features are extracted for each frame and compared to those found in previous frames. If a vertical shift is detected, its estimation is produced as output; otherwise, if no correlation is found, such as during a fast steering, the output is a conventional out-of-range value indicating the inhibition status. In the following, each algorithm step is described in detail.

A. Low level processing and horizontal edges extraction

The first step is the application of a simple Sobel operator in order to extract horizontal edges. A contrast stretching filter

is used to prepare the image to be filtered with the aim of using a good luminance-equalized source. The average frame luminance is also computed and used as threshold for the edge image. Since this is the most time consuming operation, it can eventually be performed on dedicated hardware such as SVM, speeding up the processing time for each frame. A reduction in the amount of computation to be executed in the next steps is reached by reducing the number of details extracted from the Sobel filter. An example of the input image is depicted in figure 3.a, the correspondant edge image is reported in figure 3.b.

B. Horizontal edges signature

The second step consists of the analysis of a fixed interest area of the frame in order to extract horizontal edges histogram. The area of interest is as wide as the whole image; it starts from a 10% from the top of the frame and stops 10% from the end of the frame. This choice allows to reduce computational time and improves the final result: avoiding the processing of the upper and lower areas means to avoid problems related to fast, vertical-moving objects that may confuse the algorithm. In the interest area the horizontal histogram is generated by computing the number of white pixels in each row. During normal driving, framed object contours change slowly and smoothly. In correspondence to an oscillation, this continuity is broken and all features points move mainly vertically, up or down according to oscillation direction. A maximum value has been introduced for each row of the histogram in order to avoid that fast approaching horizontal structures (such as bridges and underpasses) be misinterpreted as a feature oscillation requiring stabilization. In this phase the number of points characterizing the histogram is also computed analyzing the histogram variations. The height of the interest area is divided into eight zones and for each zone histogram variation is computed. When a variation is bigger than a fixed threshold, the zone is marked as valid for vertical shift computation. At the end of this step, as necessary condition for continuing the processing, at least 3 zones must be marked as valid. This is basically a way for measuring the quantity of information present in the image. A low information content disables the stabilizer in situations such as during fast steering where consecutive frames content is too different for searching for vertical features correlations. The eight zones are marked with horizontal green lines on the left hand side of figure 3.b.

C. Shift estimation

Once the histogram data have been collected it is possible to obtain the vertical offset between the current frame and the previous one. This operation is performed by finding the maximum value of the convolution between a reference histogram (dynamically updated in time) and the current histogram. The position of this maximum is maintained as raw shift value and then filtered using shift values obtained for previous frames.

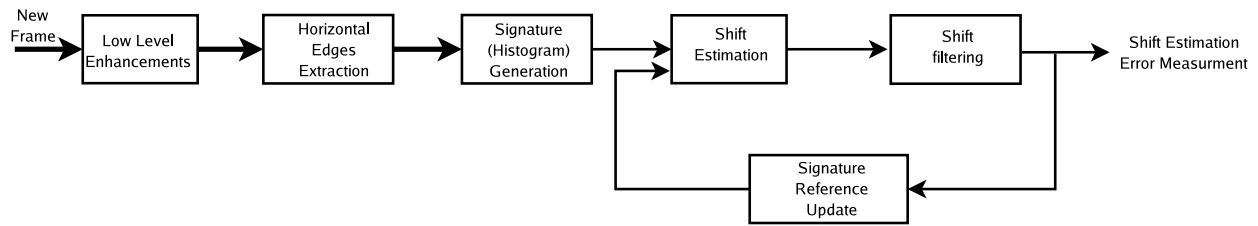


Fig. 2. Algorithm processing flowchart.

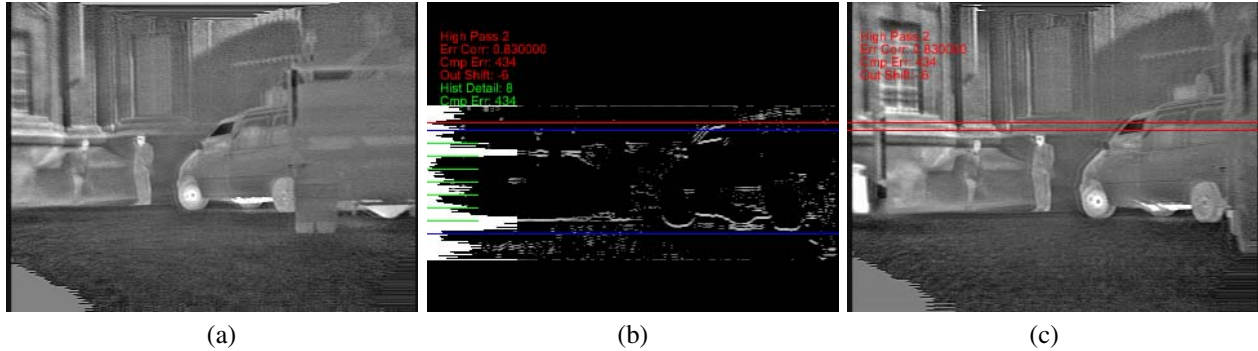


Fig. 3. Stabilization algorithm processing step. (a) Far infrared input image. (b) Output: blu lines indicate the region of interest, green lines indicate features detection areas, the intensity histogram related to edge image is depicted on the left-hand side, computed values are reported in the top-left hand side. (c) the original image translated 6 pixels down (accordingly with the detected shift).

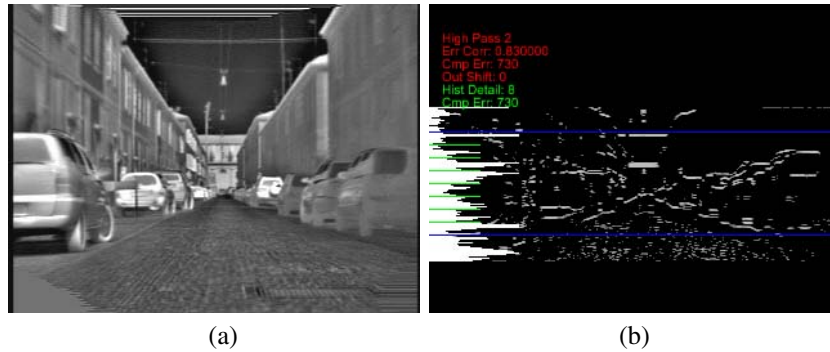


Fig. 4. Example of significant horizontal component generated by the top of the roof at the end of the road. This kind of features may compromise the stabilization algorithm behavior since they move vertically during normal driving.

- *Spurious variations filter*: isolated low amplitude variations are eliminated since they are usually caused by image noise or by errors in the reference histogram update.
- *Perspective filter*: during normal driving, far details get closer and behave in different ways according to their position relative to the vanishing point. Objects with a significant horizontal structure (such as roofs or underpasses or bridges), when approached (see figure 4), generate large and evident peaks in the histogram. These peaks mainly move up or down in the image according to their position in the image (over or below the vanishing point). In order to avoid considering these normal variations as oscillations to be corrected, a filter that removes this effect has been introduced.

The filtered shift value can be used, together with calibration parameters to estimate the new camera pitch value.

D. Reference Horizontal Signature Update

Each histogram needs to be compared with another histogram that summarizes the past history. Thus the last algorithm step is to use the new histogram as reference histogram for the next frame.

When a vehicle is driving through a road slope change, it can be observed that there is a set of features moving in a preferred direction without returning back. For example when driving through the base of an hill, features moves down, but this oscillation has a non zero mean. This is an intrinsic weakness of vision based stabilizer, since without an inertial measurement it is impossible to establish that such oscillation should be ignored.



Fig. 5. Image from the sequence used for static testing.

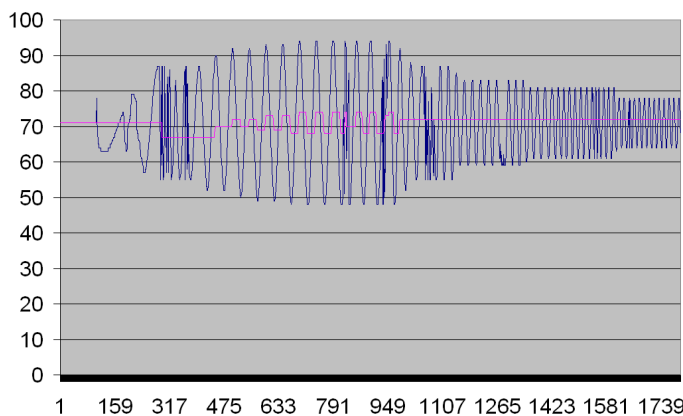


Fig. 6. Result of static testing: this graph shows the vertical position of a specific feature in the scene versus the frame number for both unstabilized (blue) and stabilized (red) images. Saturation is clearly visible when oscillation amplitude exceeds algorithm limits.

In order to allow the system return in the original condition after this kind of events (even normal oscillation have a non zero mean due to perspective effect) an appropriate smoothing factor allows to restore the stabilizer original condition.

This method prevents the reference histogram from progressively drifting from the initial setpoint and restores its original pixel shift. The speed used to converge to the starting point is an algorithm parameter.

V. RESULTS AND CONCLUSIONS

The system has been tested on 7 different video sequences (for over 15.000 frames) acquired from the vehicle in use at the Department of Information Technology of the University of Parma. Qualitative results appear to improve the image quality in several real cases.

The performance result has been quantitatively evaluated using the following procedure. The vehicle was parked in a fixed position in front of a reference grid framing images depicted in figure 5. A video sequence was acquired (named "jumping on the hood"). The image sequence (200 frames) has

been analyzed offline measuring the position (the vertical pixel coordinate in the image) of a specific feature of the framed scene. Test results are depicted in figure 6.

The stabilizer corrects the current acquired frame before the processing, thus there is no frame delay introduced in the processing pipeline. The maximum corrected amplitude has been fixed to 24 pixels (10% of the frame height) in order to avoid excessive corrections that determine bad behavior for algorithms.

The average execution time measured on a 2.8 GHz Pentium 4 class machine, running a 4100 frame sequence, is 4ms. This confirms the system is ready for real time video applications. The most time consuming operation is the Sobel filter computation, executed on the whole image. This step can be accelerated using appropriate hardware (or performed on a smaller region of interest).

Since the stabilizer has been designed to improve the robustness of vision systems, a far infrared pedestrian detector has been used as case study. An infrared video sequence framing pedestrians acquired from a moving car has been manually annotated [7] twice (with and without stabilization). The sequence contained a collection of scenes taken in proximity of holes and bumpers and thus critical due to camera pitch movement. Statistics on pedestrian detection algorithm performance has been computed in both cases, obtaining a 5.4% increment in correct detection. The developed systems seems to be promising for stabilizing image streams coming from cameras mounted on vehicles. Hardware acceleration of low level processing steps can heavily improve the execution speed.

ACKNOWLEDGMENT

This work has been funded by Volkswagen AG.

REFERENCES

- [1] M. Bertozzi and A. Broggi, *Vision-Based vehicle guidance*, Computer Vision, Vol. 30, pp. 49-55, July 1997.
- [2] M. Bertozzi, A. Broggi, P. Grisleri, T. Graf, M. Meinecke, *Pedestrian detection in infrared images*, Intelligent Vehicles Symposium, 2003. Proceedings. IEEE, 9-11 June 2003, Pages:662 - 667
- [3] R. Kurazume; S. Hirose, *Development of image stabilization system for remote operation of walking robots*, Robotics and Automation, 2000. Proceedings. ICRA '00. IEEE International Conference on, Volume: 2, 24-28 April 2000, Pages:1856 - 1861 vol.2
- [4] K. Sato; S. Ishizuka, S., A. Nikami, M. Sato, *Control techniques for optical image stabilizing system*, Consumer Electronics, IEEE Transactions on, Volume: 39, Issue: 3, Aug. 1993, Pages:461 - 466.
- [5] Yu-Ming Liang; Hsiao-Rong Tyan; Hong-Yuan Mark Liao; Sei-Wang Chen, *Stabilizing image sequences taken by the camcorder mounted on a moving vehicle*, Intelligent Transportation Systems, 2003. Proceedings. 2003 IEEE, Volume: 1, 12-15 Oct. 2003 Pages:90 - 95 vol.1.
- [6] J. S. Jin, Z. Zhu, and G. Xu, *A stable vision system for moving vehicles*, IEEE Transaction on Intelligent Transportation Systems, Vol. 1, No. 1, pp 32-39, 2000.
- [7] M. Bertozzi, A. Broggi, P. Grisleri, A. Tibaldi and M. Del Rose, *A Tool for Vision based Pedestrian Detection Performance Evaluation*, Intelligent Vehicles Symposium, 2004. Proceedings. 15-18 June 2004 Parma, Pages: 784-789.

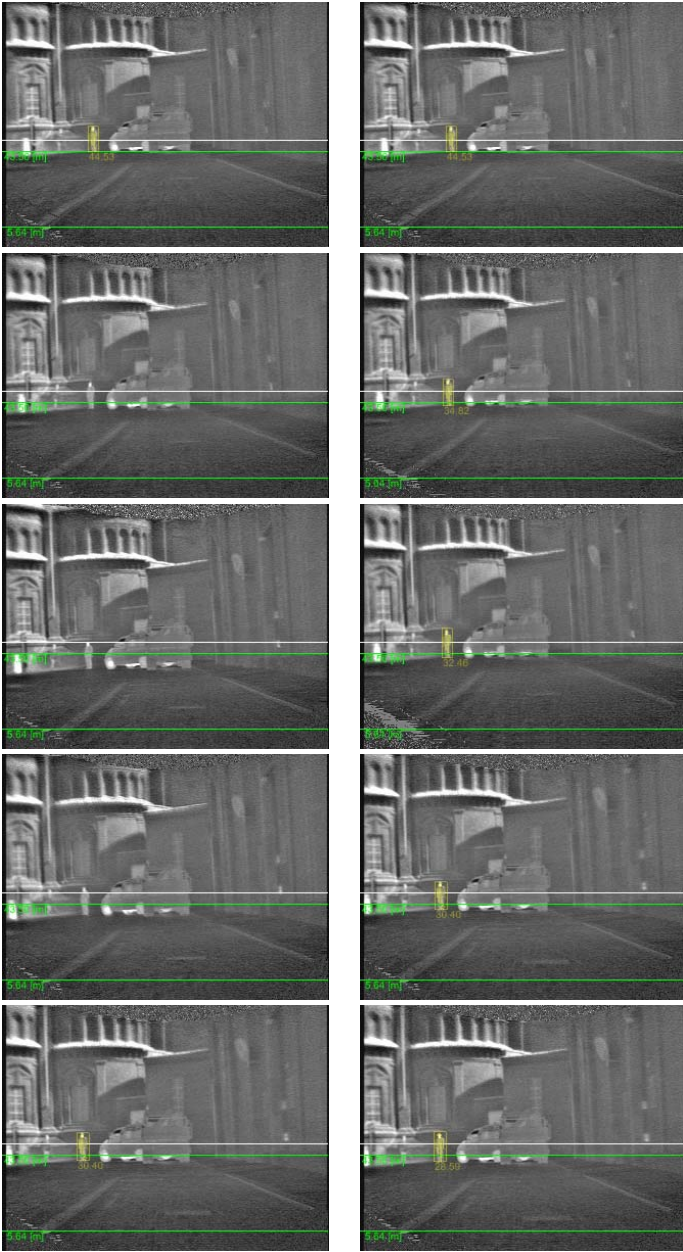


Fig. 7. Example of improved performance in pedestrian detection matching due to stabilization...