Matching in video images: camera registration and CNN based feature extraction

Theses of the Ph. D. dissertation

Zoltán Szlávik

Scientific adviser:
Tamás Roska, D. Sc.

Supervisor:
Tamás Szirányi, D. Sc.

Analogic and Neural Computing Laboratory
Computer and Automation Research Institute
Hungarian Academy of Sciences

Budapest, 2006

Table of Contents

Introduction

Computer-assisted observation of human or vehicular traffic movements, natural reserves or any kind of activities using multiple cameras constitutes a great interest in many applications. Examples are semi-mobile traffic control using automatic calibration, or tracking of objects in a surveillance system. The use of single camera limits the number of possible applications – even simplest applications need multiple cameras. Typical scenarios of multiple camera surveillance could be found in banks, airports, parking lots, stations. Here, the observation with a single camera is not possible because of occlusions and cameras' limited field of view. Such systems/algorithms should work robustly and real-time. In computer vision the development of fast and/or robust algorithms is still a great challenge. Typical multi camera system performs the following steps:

1. change or motion detection;
2. object detection;
3. classification of objects – position, class, features etc.
4. tracking of objects;
5. event detection.

During my work I tried to answer some of the above problems. On the one hand, I have developed real-time algorithms for the analysis of face images and on the other hand, I have developed robust algorithms for the matching of images.

Object detection and tracking is a well defined problem for single images and image sequences. The detected object must be found in the later frames, which is a correspondence problem between the current frame and next frames. Multiple camera tracking means finding the same object in different views. To exploit information exchange between different cameras a correspondence must be established between them. A possible solution is the matching of different views of the same scene.

Matching different images of a single scene may be difficult, because of occlusion, aspect changes and lighting changes that occur in different views. Still-image matching algorithms [21][22][23][24] search for still features in images such as: edges, corners, contours, color, shape etc. They are useful for image pairs with small differences; however they may fail at occlusion boundaries, within featureless regions and may fail if the chosen primitives or features cannot be reliably detected. The views of the scene from the various cameras may be very different, so we cannot base the decision solely on the color or shape of objects in the scene. In a multi-camera observation system the video sequences recorded by cameras can be used for estimating matching correspondences between different views. Video sequences in fact also contain information about the scene dynamics, besides the static frame data. Scene dynamics is an inherent property of the scene independently of the camera positions, the different zoom-lens settings and lighting conditions. References [25] and [26] present approaches in which motion-tracks of the observed objects are aligned. In practice, the existing algorithms can be used only in restricted situations. In case of scenes including several objects in random motion, successful registration of images from separate cameras conventionally requires some *a priori* object definition or human interaction. But in most cases the extra information of *a priori* object models or human interaction is not available. During my work I focused on approaches that can establish a correspondence between different views without a priori defined object models and scene structures in fully automatic way.

Fast algorithms must be developed to be able to integrate several algorithms, e.g. object detection, tracking, classification, into a single system. The most natural way of identification of human in images or videos is the analysis of their faces' image. Humans' face is a non-intrusive biometric feature, which means that the identification can be done without disturbing the observed human.

In the literature of the computer vision many examples about how the face images can be analysed are described. Usually a face detection of face recognition algorithm [29][30] is built up from the follwing steps: i) detection of face like images (face candidates); ii) verification of face candidates; iii) identifiaction. In the verification step about each face candidate the decision of being a face or not must be made. A possible solution when the face candidate is compared to an average face or to a face model [28]. Another possible solution is when facial features are extracted and their geometrical relationship verifies the face candidate [27]. Such algorithms need a fast facial feature extractor method. During my work I have developed a real-time facial feature extraction algorithm, which was implemented on ACE4K Cellular Visual Microprocessor, a real-hardware for real time image processing.

Methods used in experiments

In the course of my work, theorems and assertions from the field of ordinary and partial differential equations, mathematical statistics, numerical geometry, optimization, reported results of image and video processing were explored.

The experiments for camera registration were performed by using the MDICAM multi-camera software system that was designed in the Analogical and Neural Computing Laboratory. For unique experiments I have also designed simulation systems in Matlab. Testing of the proposed algorithms was performed on various video sequences from personal experiments and from publicly available video databases.

For the design and testing of analogic algorithms I have used a software-hardware system developed in the Analogical and Neural Computing Laboratory. Designed CNN templates and algorithms were tested on software simulators, such as Aladdin System and on Cellular Visual Microprocessors: ACE4k and ACE16k. The simulating system was developed in the Analogical and Neural Computing Laboratory. The implementation of different methods was completed in different CNN languages (Alpha, AMC, UMF) ensuring their applicability on different platforms. The proposed methods were tested on images from a publicly available image database [31].

Theses of the dissertation

First Thesis

Accumulated motion statistics for automatic camera registration.

a. I gave a new camera registration method based on the use of co-motion statistics.

I defined the notion of co-motion statistics that is based on the accumulation of motion statistics for different camera views. I have shown experimentally that co-motion statistics can be used for camera registration without any *a priori* knowledge about the objects' appearance or motion and without human interaction. The method's accuracy is in the subpixel range.

Published in [2][6][16]

b. I showed that temporal alignment of image sequences in the case of co-motion statistics based camera registration is possible through the minimization of the estimated alignment error.

In camera registration, it is usually assumed that the cameras are synchronized. I have solved the problem of temporal alignment of image sequences during camera registration for unsynchronized cameras. The proposed method aligns the camera views for different time offsets and then searches for offset with minimum alignment error.

Published in [2][6]

Second Thesis

Automatic Bayesian method for camera registration without human interaction and assuming any *a priori* information about scene structure, appearance of objects or movement.

a. **I showed that the extraction of cameras' overlapping field of views can be done by a new automatic Bayesian iterative algorithm.**

   The knowledge of cameras' overlapping field of view is a prerequisite of the fast and robust matching of wide baseline stereo images. I worked out a Bayesian model for modelling concurrent changes in different camera views. I have shown that the solution of this model is equivalent to a solution of a periodic Markov chain. The developed method provides an automatic solution to the problem without any human interaction and *a priori* object model.
   Published in [1][8]

b. **I showed that an entropy based analysis can be applied for the extraction of pixels of significant non-continuous changes.**

   I worked out an entropy based method for the classification of motion histories into two classes: (i) significant non-continuous changes; (ii) significant continuous changes or camera noise.
   Published in [1][8]

Figure 1. Result of alignment of two images of the same scene captured by different cameras.

In practice, the existing algorithms can be used only in restricted situations. The reported methods focus on the solution of the view-registration problem in respect of outdoor scenes, and neglect the additional difficulties, which tend to arise for indoor scenes. In the case of indoor cameras, the still-image based methods may fail due to the variability of conditions: occlusions, changing illumination etc. Due to the larger size of the moving objects, the existing motion-based methods also fail; the observed motions are not necessarily on the ground-plane – while for outdoor scenes, such an assumption can safely be made. But if we could detect feature points that are on the groundplane then the motion based methods still can be applied for the matching of two views.

c. **I introduced shadow as salient features on the ground and I have shown experimentally that shadows efficiently can be used for camera registration.**

   Shadows are excellent features for indoor scenes, because they are mainly on the ground plane. I have experimentally shown that if we could detect shadows of moving objects then we have a lot of concurrently moving points in the cameras. By extracting them, the matching of the views can be done.
   Published in [7]

Third Thesis

Working out robust feature extraction algorithms by using spatio-temporal dynamics.

a. **I did show that basic face features can be extracted by a set of procedures which can run on the reduced set of parallel operations in the CNN UM.**

I have developed new analogic algorithms for the extraction of main, eyes, nose and mouth, facial features. The extraction of primary facial features such as eyes, nose and mouth relies on the fact that the distance between the eyes is proportional to the distances between the facial features. All the algorithms run on the ACE4K CNN chip, and the achieved speed of the algorithms is equivalent to 50 frames per second.

Published in [14]

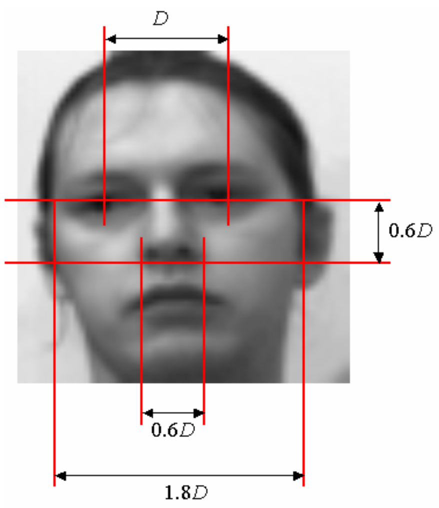


Figure 2. The defined geometry of the face model

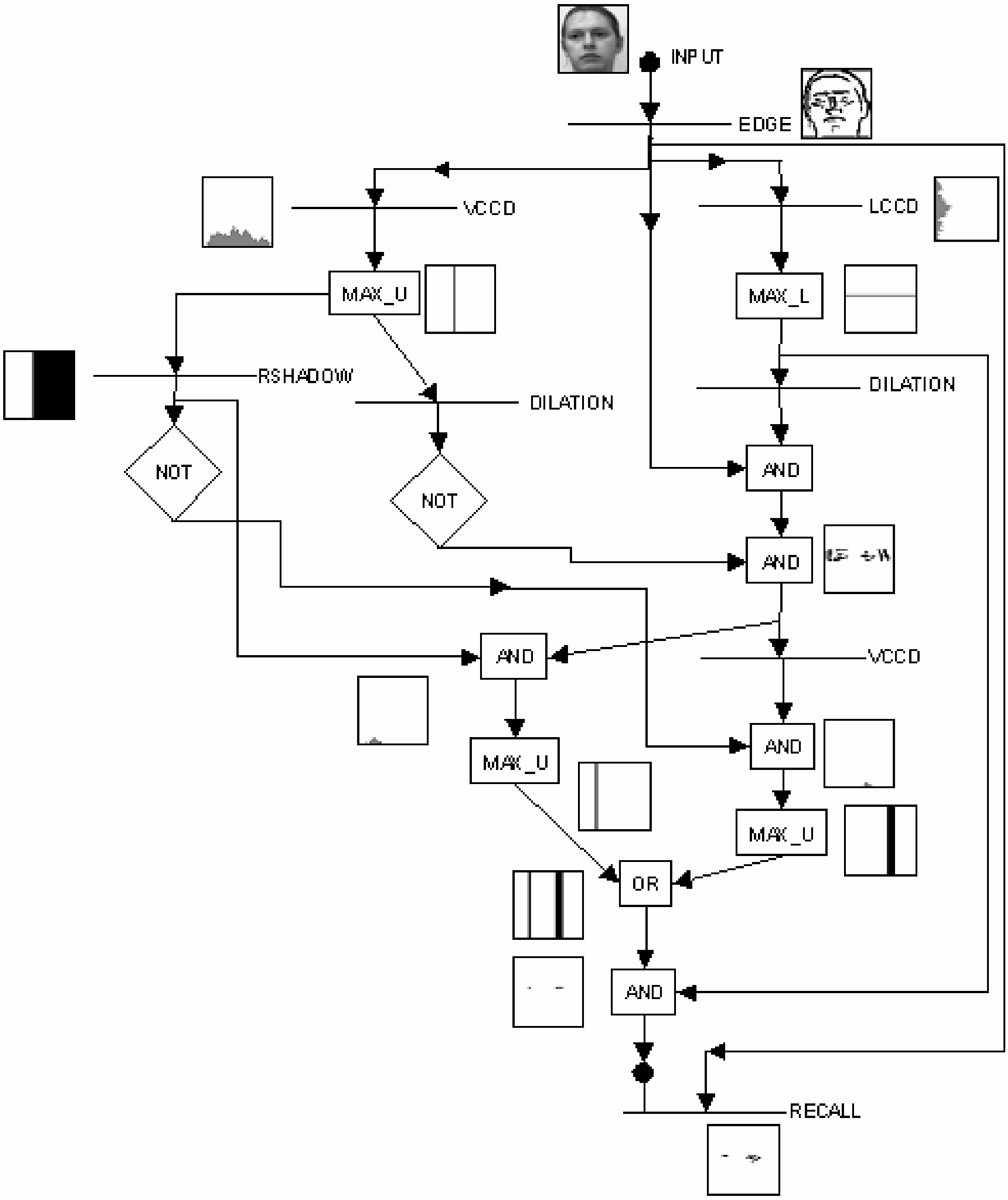


Figure 3. The UMF diagram of the proposed eye detection algorithm

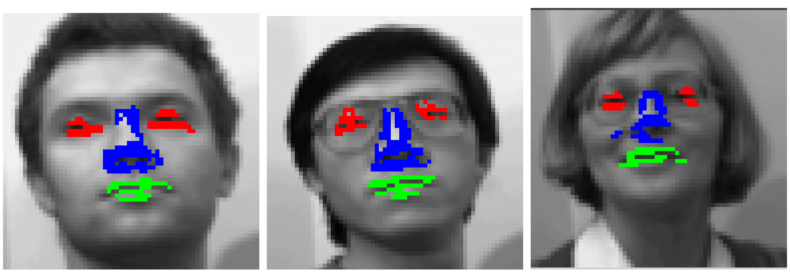
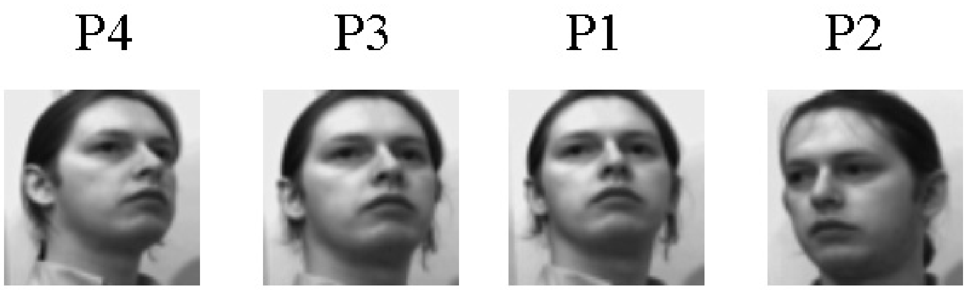


Figure 4. Sample results of facial feature extraction on CNN-UM

b. I worked out an analogic algorithm for the identification of rotated faces by using CNN.

I have designed and implemented the algorithm in the CNN-UM framework. The algorithm is based on the estimation of face position by using analogic facial feature extraction algorithms. The steps of the algorithm are as follows: extraction of facial features, estimation of face position, transforming back the face to the frontal position and comparing to images in the database.

Published in [19]



P4	89%	80%	70%
92%	**P3**	83%	67%
84%	86%	**P1**	88%
72%	81%	90%	**P2**

Figure 5. Results of identification of rotated faces. The rows are the training classes and the columns are the test classes in the table. Example images from each class are also shown.

Possible applications

All the developed methods and implementations offer solutions for real applications problems.

Methods of First and Second Theses provide general solutions for camera registration in real circumstances. They can be used in any multi-camera surveillance system where automatic functioning is needed.

Algorithms of the Third Thesis provide solutions for computer systems where real-time face analysis is needed, e.g. face detection, identification.

The author's publications

Journal publications

[1] **Zoltán Szlávik**, Tamás Szirányi, "Stochastic view registration of overlapping cameras based on arbitrary motion", *IEEE Transactions on Image Processing*, **accepted**, 2006

[2] **Zoltán Szlávik**, Tamás Szirányi, László Havasi, "Video camera registration using accumulated co-motion maps", *ISPRS Journal of Photogrammetry and Remote Sensing*, **accepted**, 2006

[3]László Havasi, **Zoltán Szlávik**, Tamás Szirányi, "Detection of gait characteristics for scene registration in video surveillance system", *IEEE Transactions on Image Processing*, **accepted**, 2006

[4] László Havasi, **Zoltán Szlávik**, Tamás Szirányi, "Higher order symmetry for non-linear classification of human walk detection", *Pattern Recognition Letters*, vol. 27(7), pp. 822-829, 2006

International conference publications

[5] László Havasi, **Zoltán Szlávik**, Tamás Szirányi, "Use of human motion biometrics for multiple-view registration", Advanced Concepts for Intelligent Vision Systems, ACIVS 2005, LNCS, vol. 3708, pp. 35-44, 2005

[6] **Zoltán Szlávik**, László Havasi, Tamás Szirányi, "Estimation of common groundplane based on co-motion statistics", International Conference on Image Analysis and Recognition, ICIAR'04, LNCS, vol. 3211, pp. 347-353, 2004

[7] **Zoltán Szlávik**, Tamás Szirányi, László Havasi, "Analysis of dynamic scenes by using co-motion statistics", IEEE International Workshop on Visual Surveillance, 2006

[8] **Zoltán Szlávik**, Tamás Szirányi, "Bayesian estimation of common areas in multi-camera systems", IEEE International Conference Image Processing, ICIP, Atlanta, **accepted,** 2006

[9] Cs. Benedek, L. Havasi, T. Szirányi, **Z. Szlávik**, "Motion-based Flexible Camera Registration", in Proc. of IEEE Advanced Video and Signal-Based Surveillance, AVSS'05, pp. 439-444, 2005

[10] **Zoltán Szlávik**, László Havasi, Tamás Szirányi, Csaba Benedek, "Random motion for camera calibration", European Signal Processing Conference, EUSIPCO, Antalya, 2005

[11] **Zoltán Szlávik**, Tamás Szirányi, László Havasi, Csaba Benedek, "Optimizing of searching co-motion point-pairs for statistical camera calibration", IEEE International Conference on Image Processing, ICIP, pp. 1178-1181, Genova, 2005

[12] László Havasi, **Zoltán Szlávik**, Tamás Szirányi, "Eigenwalks: walk detection and biometrics from symmetry patterns", IEEE International Conference on Image Processing, ICIP, pp. 289-292, Genova, 2005

[13] László Havasi, **Zoltán Szlávik**, Csaba Benedek, Tamás Szirányi, "Learning human motion patterns from symmetries", International Conference on Machine Learning Workshop on Machine Learning for Multimedia, Bonn, 2005, (on CD-ROM)

[14] **Zoltán Szlávik**, Tamás Szirányi, "Face analysis using CNN-UM", Proc. of Cellular Neural Networks and Aapplications, CNNA'04, pp. 190-196, Budapest, 2004

[15] László Havasi, **Zoltán Szlávik**, "Symmetry feature extraction and understanding", Proc. of Cellular Neural Networks and Aapplications, CNNA'04, pp. 255-261, Budapest, 2004

[16] **Zoltán Szlávik**, László Havasi, Tamás Szirányi, "Image matching based on co-motion statistics", 2nd IEEE Int. Symposium on 3D Data Processing, Visualization, and Transmission, 3DPVT, 2004, (on CD-ROM)

[17] László Havasi, Csaba Benedek, **Zoltán Szlávik**, Tamás Szirányi, "Extracting structural fragments of overlapping pedestrians", Proc. Of the 4th IASTED Int. Conference VIIP'04, pp. 943-948, Marbella, 2004

[18] László Havasi, **Zoltán Szlávik**, Tamás Szirányi, "Pedestrian Detection using derived Third-Order Symmetry of Legs", ICCVG'04

[19] **Z. Szlávik**, T. Szirányi, "Face identification using CNN-UM", Proceedings of European Conference on Circuit Theory and Design, ECCTD'03, vol. 2, pp. 81-85, Krakow, 2003

[20] **Z. Szlávik**, D. Bálya, T. Roska, "Properties of the adaptive integration formula to compute the CNN dynamic equation", Proceedings of European Conference on Circuit Theory and Design, ECCTD'03, vol. 1, pp. 105-109, Krakow, 2003

References

[21] S. T. Barnard, W. B. Thompson, “Disparity analysis of images,” IEEE Trans. PAMI, vol. 2, pp. 333-340, 1980.

[22] J. K. Cheng, T. S. Huang, “Image registration by matching relational structures,” Pattern Recog., vol. 17, pp. 149-159, 1984.

[23] J. Weng, N. Ahuja, T. S. Huang, “Matching two perspective views,” IEEE Trans. PAMI, vol. 14, pp. 806-825, 1992.

[24] Z. Zhang, R. Deriche, O. Faugeras, Q.-T. Luong, “A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry,” Artificial Intelligence Journal, vol. 78, pp. 87-119, 1995.

[25] L. Lee, R. Romano, G. Stein, “Monitoring activities from multiple video streams: establishing a common coordinate frame,” IEEE Trans. PAMI, vol. 22, 2000.

[26] Y. Caspi, D. Simakov, and M. Irani, “Feature-based sequence-to-sequence matching,” in Proc. VAMODS (Vision and Modelling of Dynamic Scenes) workshop, with ECCV’02, Copenhagen, 2002.

[27] Brunelli R, Poggio T. Face recognition through geometrical features. Proceedings of ECCV, S. Margherita Ligure, 1992; 792-800.

[28] Brunelli R, Poggio T. Face recognition: features versus templates. IEEE Transactions on Pattern Analysis and Machine Intelligence 1993;15(10):1042-1052.

[29] Chellappa R, Wilson C L, Sirohey S. Human and Machine Recognition of Faces: A Survey. Proceedings of IEEE 1995; 83:705-740.

[30] Hjelmas E, Low B K. Face detection: A survey. Computer Vision and Image Understanding 2001:83(3):236-274.

[31] Graham D B, Allinson N M. Characterizing Virtual Eigensignatures for General Purpose Face Recognition. Face Recognition: From Theory to Applications, NATO ASI Series F, Computer and Systems Sciences 1998;163:446-456.