

USING COMPUTER VISION IN SECURITY APPLICATIONS

Peter Peer, Borut Batagelj, Franc Solina

University of Ljubljana, Faculty of Computer and Information Science
Computer Vision Laboratory
Tržaška 25, 1001 Ljubljana, Slovenia
{peter.peer, borut.batagelj, franc.solina}@fri.uni-lj.si

Abstract: *In this paper we present projects developed in the Computer Vision Laboratory, which address the issue of safety. First, we present the Internet Video Server (IVS) monitoring system [5] that sends live video stream over the Internet and enables remote camera control. Its extension GlobalView [1,6], which incorporates intuitive user interface for remote camera control, is based on panoramic image. Then we describe our method for automatic face detection [3] based on color segmentation and feature extraction. Finally, we introduce our SecurityAgent system [4] for automatic surveillance of observed location.*

Key words: *computer vision, security, panoramic image, Internet, face detection, automatic surveillance.*

1. INTRODUCTION

Computer vision is relatively young branch of computer science that tries to get as much information from the image or sequence of images as possible. This information is than used for scene reconstruction, object modeling, visualization, navigation, recognition, surveillance, virtual reality or similar tasks.

Video information is an important component of each security application. Intelligent or automatic examining of this information is becoming exceptionally important.

In this paper we present projects developed in Computer Vision Laboratory, which address the issue of safety. The Internet Video Server (IVS) system [5] and its extension GlobalView [1,6] are discussed in Section 2. Automatic face detection system [3] is presented in Section 3. Section 4 reveals the idea of the SecurityAgent system [4]. Conclusion of the paper is given in Section 5.

2. IVS AND ITS GLOBAL VIEW EXTENSION

The IVS system [5] was developed in 1996 (it was one of the first such systems in the world) to study user-interface issues of remotely operable cameras and to provide a testbed for the application of computer vision techniques such as motion detection, tracking and security.

The first version of the IVS system enabled live video transmission and remote camera control over the World Wide Web (Figure 1). Navigation is possible by clicking the buttons, which issue the command to the pan-tilt unit to move the camera in the predefined direction.



Figure 1. The plain IVS user interface showing buttons for controlling the camera direction.

Extended testing of the IVS prompted us to improve its user interface. We noticed that because of the slow and irregular responses to the user requests, the system does not seem to be very predictable. The second problem with such a user interface is even more perceivable. If the focal length of the lens is large then one can easily lose the notion where the camera is pointing. When moving the camera step by step, we can not predict the exact final position of it, nor do we know what are the relations between the object in the image and other objects on the scene. We can solve all these problems with an improved interface, which includes panoramic view of the whole observed environment.

The GlobalView extension [6] of IVS was developed, which enables the generation of panoramic images of the environment and more intuitive control of the camera (Figure 2). At the system startup, the panoramic image is first generated by scanning the complete surrounding of the camera. When a user starts interacting with the system he or she sees the whole

panoramic image. A rectangular frame in the panoramic image indicates the current direction of the camera. The user can move this attention window with a mouse and in this way control the direction of the camera. From the position of the attention window within the panoramic image the physical coordinates of the next camera direction are computed and appropriate command is issued to the pan-tilt unit of the camera. Before the camera moves, the last live image is superimposed on the corresponding position in the panoramic image. In this way the panoramic image is constantly updated.

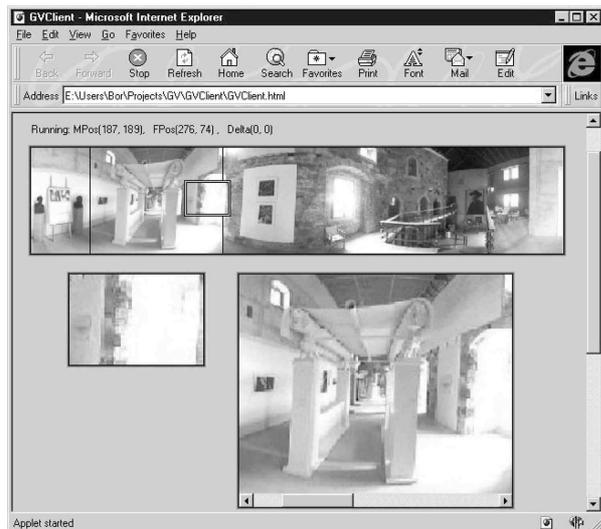


Figure 2. The intuitive GlobalView user interface. The rectangle in the panoramic image indicates the attention window, which presents the current direction of the camera. By moving the rectangle the user controls the direction of the camera. On the bottom of the web page the live video image in actual resolution (left) and a zoomed panoramic view (right) are shown.

The whole interaction with the system is carried out through the attention window. Moving this attention window results in the camera movement. At any time, only one user can control the camera, since allowing more than one user to interact with it could be confusing. To allow a fair time-sharing every user can control the camera for a few minutes and after that the control is passed to the next user waiting in the queue.

The system described so far is based on the analog camera that is attached to the special pan-tilt unit (Figure 8). The latest version of the system works with the JVC network camera [1] (Figure 3), which is digital camera with the pan-tilt unit. It contains a computer and a compression chip. The computer is small and specialized for network applications. The JVC network camera has its own IP address. It is connected to the network as a network device and it has build-in web server. It has 10x optical zoom and input/output alarm capability.



Figure 3. The JVC network camera VN-C3WU with a build-in web server is used in the latest version of the IVS system.

3. AUTOMATIC FACE DETECTION

Automatic face detection is like most other automatic object-detection methods difficult, especially if sample variations are significant. Sample variations in face detection arise due to individual face differences and due to differences in illumination.

We developed a face detection method, which consists of two distinct parts: making face hypotheses and verification of these face hypotheses [3]. This face detection method tries to join two theories: it is based on detection of shape features, i.e. eye pairs (bottom-up feature-based approach), and skin color (color-based approach). The method assumes certain circumstances and constraints, respectively. Therefore, it is not applicable universally.

The two basic limitations of the method thus originate from constraints of these approaches:

- input image must have high enough resolution (the face must be big enough) and
- it is sensitive to the complexion (i.e. fair skin).

The basic idea of the algorithm is as follows: find in image all regions, which contain possible candidates for an eye, then on the basis of geometric face characteristics try to join two candidates into an eye pair and finally, confirm or refuse the face candidate using complexion information.

The method was projected over a set of quite different pictures, i.e. the training set. The goal of the method was to reach maximum classification accuracy over the images, which meet the following demands and constraints, respectively (beside already mentioned two):

- real-time operation on standard personal computer,
- plain background,
- uniform ambient illumination,
- fair-complexion faces, which must be present in the image in their entirety (frontal position) and
- faces turned round for at most 30 degrees.

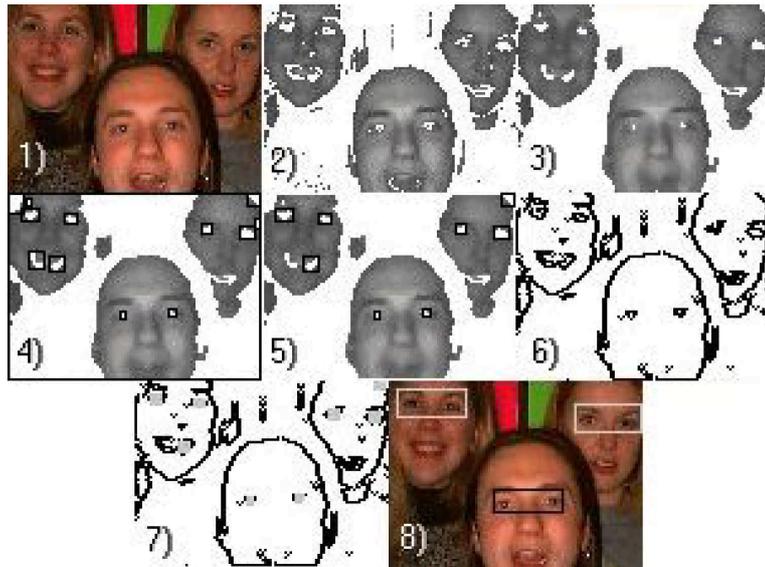


Figure 4. Basic principles of our face detection method: 1) input image, 2) eliminated insignificant colors, 3) image filtered with median filter, 4) segmented white regions, 5) eliminated insignificant regions, 6) traced edges, 7) best possible circles within regions of interest, 8) output image. Also next feature of the method is evident from the figure: because of greater distance of the faces in the background, the algorithm ascribe to the latter lower probability based on the fact, that they have lower resolution and worse illumination then the face in the foreground.

The effectiveness of the method was tested over an independent set of images.

The method uses the captured picture in a small resolution as input. So the input can contain more than just one face candidate. The basic principle of the method described in [3] is illustrated in Figure 4.

The method requires some thresholds, which play a crucial role for proper processing. They are set quite loosely (tolerantly), but they become effective as a sequence. All thresholds were defined experimentally using the training set.

In order to get reliable results regarding algorithm's effectiveness, this method was tested over an independent testing set of two public image-databases with good results [3].

Given the constraints, our method is effective enough for applications where fast execution is required. One example of such an application is our interactive, computer-vision based art project called *15 seconds of fame* [2].

3.1. 15 SECONDS OF FAME

15 seconds of fame [2] is an interactive art installation, which elevates the face of a randomly selected gallery visitor into a "work of art" for 15 seconds. The installation was inspired by Andy Warhol's statement that "In the future everybody will be world famous for fifteen minutes" as well as by the pop-art style of his works.

The installation intends to make instant celebrities out of common people by putting their portraits on the museum wall for 15 seconds. Instead of 15 minutes as presaged by Warhol we opted for 15 seconds to make the installation more dynamic. This also puts a time limit for the computer processing of each picture. Since the individual portrait which is presented by the installation is selected by chance out of many faces of people who are standing in front of the installation by a random number generator, the installation tries to convey that besides being short-lived, fame tends to be also random.



Figure 5. Flat-panel computer monitor dressed up like a precious painting for the *15 seconds of fame* installation. The round opening above the picture is for the digital camera lens.

The installation consists of a computer with a flat-panel monitor which is framed and displayed as a painting on the wall (Figure 5), a digital camera pointing at the

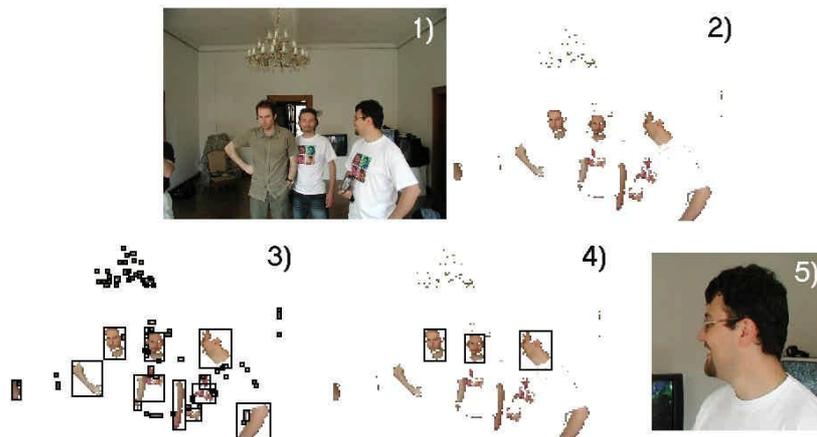


Figure 6. Finding face candidates in the *15 seconds of fame* installation: 1) downsize the resolution of the original image (to speed up the process), 2) eliminate all pixels that can not represent a face, 3) segment all the regions containing face-like pixels, 4) eliminate regions which can not represent a face, 5) randomly select one face candidate and crop it from the original image. After the face is cropped from the image, it is transformed using again a randomly selected pop-art color filter (the number of filters goes in millions) and displayed on the monitor for 15 seconds.

viewers of the monitor and proprietary software that can detect human faces in images and graphically transform them.

The described face detection algorithm is quite complex, so we decided that we integrate a simpler version of it in this installation. Figure 6 illustrates the modified process.

The first public showing of the *15 seconds of fame* installation was in Maribor, at the 8th International festival of computer arts, 28 May – 1 June 2002. The installation was active for five consecutive days and the face detection method turned out to be very accurate.

4. SECURITY AGENT

Most of the security systems are based on internal TV. This kind of monitoring demands that the security officer is present. Video content is written on tape, which is then put into the archive. It is obvious that the examination of the material is time consuming and the maintenance is expensive.

Modern systems for video monitoring and security providing should combine new technologies like Internet technologies, database technologies, image processing technologies and telecommunication technologies. Each technology can add a specific functionality to such system.

Basic video monitoring computer application, which can be thought of as an equivalent to the standard video monitoring non-computer based systems, can be presented simply by the connection between the computer and the camera. Computer captures the video and shows it on screen. In this case the security officer must be present to monitor the premises while the

computer does not do anything else but captures and shows the images. This means that the internal TV can be replaced by the computer application. However, this basic application also has similar problems as an internal TV based system. In general, we are dealing with the flexibility problem.

We would like to have such an application that can at least partially, if not completely, eliminate the need of the security officer presence.

SecurityAgent presents solutions of mentioned problems of standard security systems. It is an application for security providing and monitoring without the presence of the security officer.

When we monitor one location we are limited with the camera's field of view. Nevertheless, we would like to monitor the whole location. To achieve this, we can use a robotic arm, which can be turned in any direction (Figure 8). In this way the security officer can see every small corner of the location. However, we would prefer if we could unburden the security officer at least a bit. That is why we use image processing algorithms in order to detect changes on the scene, i.e. we detect motion. When the change on the scene is detected, the system can immediately inform the security agent or system's operator about an event. He or she can then focus on the event. To go even further, we can demand that the application makes visual and numerical summaries of events. The operator can then be almost totally discharged and his or her presence is usually not needed. Let us illustrate an example of the summary: when the motion is detected, the system starts following the motion, storing the images of the event and some important numerical data like date and time stamp of each image.

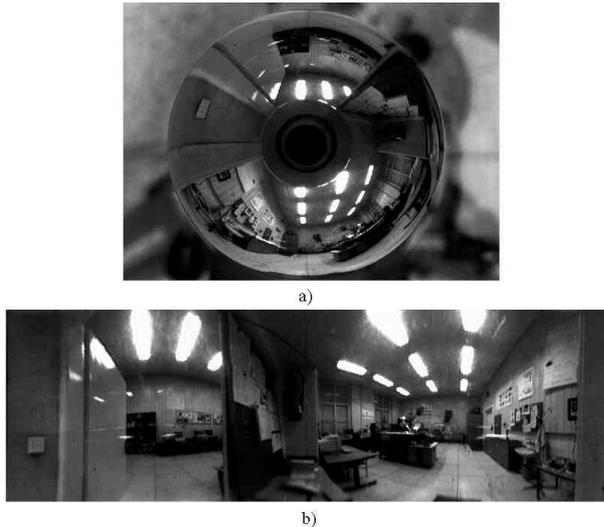


Figure 7. We are using the catadioptric panoramic camera to monitor the whole location in every moment: a) the input image taken with the catadioptric panoramic camera using spherical mirror, b) the panoramic image after applying the cylindrical transformation on the input image. An example shown in b) is much more intuitive for human perception, but the computer does not care how this panoramic data is presented to it. This is why the SecurityAgent is using panoramic data presented in a).

However, the solution with using the robotic arm is not optimal, since we would like to be able to monitor the whole location in every moment.

Standard cameras have a limited field of view, which is usually smaller than the human field of view. Because of that, people have always tried to generate images with a wider field of view, up to a full 360 degrees panorama. And the panoramic image presents the solution of our problem.

But how can we capture the panoramic image in every moment? We can use the catadioptric panoramic camera. Catadioptric panoramic camera consists of a mirror and the camera, which captures the image reflected from the mirror. The simplest catadioptric camera combines the standard camera and the planar mirror, but this combination does not enable us to capture panoramic images. Already by using partially planar mirror, we can capture images with wider field of view than with only the standard camera. An example of panoramic image generated with the help of the catadioptric panoramic camera is shown in Figure 7.

Now we have the solution of the problem how to monitor the whole location in every moment at hand. Besides using the standard camera fixed on the robotic arm, we are using the catadioptric panoramic camera (Figure 8). The catadioptric panoramic camera has a lower resolution than the standard camera. The lowest resolution is on the border of the image, i.e. on the border of the circle, which presents the contour of the mirror (Figure 7a), while we are interested only in the

information within this circle. Unfortunately, most events are detected in regions close to this border, but we can dispatch this deficiency with the use of standard camera. Images captured by the catadioptric panoramic camera are used for global motion detection, which can be a very effective procedure also in the case of low resolution images. When the global motion is detected, the SecurityAgent turns the robotic arm in the direction where the motion was detected and the standard camera that is attached on the top of robotic arm enables us to see what is going on at the location in bigger resolution.

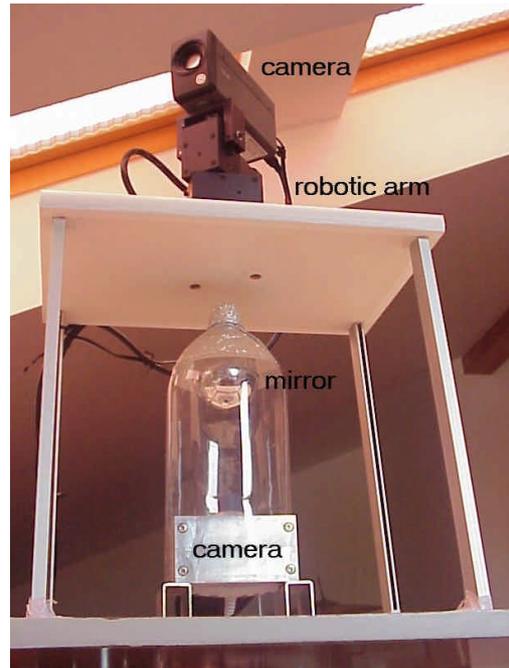


Figure 8. The SecurityAgent basic hardware part combines one standard camera attached on the robotic arm and one catadioptric panoramic camera (it consists of standard camera and spherical mirror), which enables to monitor whole location in every moment.



Figure 9. The basic user interface of the SecurityAgent.

The system can work in two modes (Figure 9): operator mode or automatic mode, i.e. without the presence of the operator. The features of the automatic mode are motion detection, motion following, switching between standard and panoramic camera, making visual summaries of activities on an observed location,

querying the visual summaries database over the Internet, notifying operator about activities on an observed location (on screen and e-mail notification, but in the future the message will be send over one telecommunication protocol, like SMS (Short Message System) protocol) and sending live video over the Internet. The basic flow chart for fully automatic surveillance is shown in Figure 10.

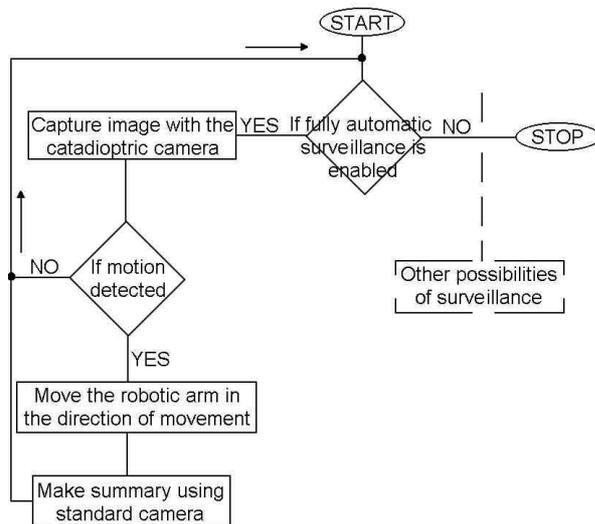


Figure 10. The basic flow chart for fully automatic surveillance in SecurityAgent system.

5. CONCLUSION

In the eyes of recent events marked with terrorist attacks, the safety of transportation systems became one of the most important issues in the world. The surveillance material is basically pictorial and normally we are dealing with enormous amount of such material. The examination of the material is time consuming and the maintenance is expensive. Computer vision can help us overcome these problems, if not completely automate them.

Our future work in development of security applications is mainly focused on integration of presented projects. Face detection could represent one more feature of automatic surveillance system, while remote manual observation and control of the camera in this system could be upgraded with our friendly panoramic user interface.

REFERENCES

- [1] B. Batagelj, P. Peer, F. Solina, System for Active Video Observation over the Internet, *International Symposium on Video / Image Processing and Multimedia Communications VIPromCom'02*, IEEE Region 8, pp. 221-226, Zadar, Croatia, 2002.
- [2] F. Solina, P. Peer, B. Batagelj, S. Juvan, 15 seconds of fame – an interactive, computer vision based art installation, accepted to *International Conference on Control, Automation, Robotics and Vision ICARCV'02*, Singapore, 2002.
- [3] P. Peer, F. Solina, Automatic Human Face Detection Method, *Computer Vision Winter Workshop CVWW'99*, pp. 122-130, Rastendorf, Austria, 1999.
- [4] P. Peer, F. Solina, SecurityAgent: Security and Monitoring Computer Enhanced Video System, *Electrotechnical and Computer Science Conference ERK'01*, IEEE Region 8, Volume B, pp. 119-122, Portorož, Slovenia, 2001. (Slovenian version)
- [5] B. Prihavec, F. Solina, Sending live video over Internet, *Workshop of the Austrian Association for Pattern Recognition (ÖAGM/AAPR'97)*, pp. 299-303, Hallstatt, Austria, 1997.
- [6] B. Prihavec, F. Solina, User interface for video observation over the internet, *Journal of Network and Computer Applications*, 21:219-237, 1998.

Peter Peer is a research assistant at the Faculty of Computer and Information Science, University of Ljubljana, Slovenia. He received B.Sc. and M.Sc. in computer science from the University of Ljubljana in 1998 and 2001, respectively. Currently he is moving towards his Ph.D. in computer science. His research interests are focused on computer vision methods.

Borut Batagelj is a research assistant at the Faculty of Computer and Information Science, University of Ljubljana, Slovenia. He received B.Sc. in computer science from the University of Ljubljana in 2001, respectively. Currently he is moving towards his M.Sc. in computer science. His research interests are focused on computer vision methods.

Franc Solina is a full professor of computer science at the University of Ljubljana, Slovenia and head of Computer Vision Laboratory at the Faculty of Computer and Information Science. He received B.Sc. and M.Sc. in electrical engineering from the University of Ljubljana in 1979 and 1982 and Ph.D. in computer science from the University of Pennsylvania, USA in 1987, respectively. His research interests are focused on computer vision methods.