Aleš Leonardis
Franc Solina (Eds.)

Computer Vision – CVWW'98

Proceedings of the Computer Vision Winter Workshop
Gozd Martuljek, Slovenia
9–12 February 1998

IEEE Slovenia Section

Contents

A. Leonardis, F. Solina (Eds.): Computer Vision – CVWW'98,
Proceedings of the Computer Vision Winter Workshop,
Gozd Martuljek, Slovenia, 9–12 February 1998, pp. 134–146

User Interface for Video Observation Over the Internet *

Bor Prihavec, Franc Solina
Computer Vision Laboratory
Faculty of Computer and Information Science,
University of Ljubljana
Tržaška 25, 1000 Ljubljana, Slovenia
Tel. +386 61 1768 381
Fax. +386 61 1264 647
Bor.Prihavec@fri.uni-lj.si Franc.Solina@fri.uni-lj.si

February 26, 1998

Abstract

We present the design and application of a system for live video transmission and remote camera control over the World Wide Web. Extensive testing of the Internet Video Server (IVS) prompted us to improve its user interface. We developed the GlobalView extension of IVS which enables the generation of panoramic images of the environment and a more intuitive control of the camera. The live video frame is superimposed on a 360^o panoramic picture. By interactively moving a rectangular frame in the panoramic picture the user locally selects the new direction of the camera. Once the view is selected the users prompts the selection and the command is issued over the Internet to the remotely controlled camera. Visual summaries of activities on an observed location can be generated and custom queries are possible with a similar intuitive user interface.

---

*This work was supported by the Ministry of Science and Technology of Republic of Slovenia (Project L2-8721) and by the European Cultural Month Ljubljana 1997 (Project Netropolis – Cyborg's Eye).

1 Introduction

Live video transmission over the Internet and interactivity are becoming more and more popular. This very moment we can find on the World Wide Web hunderts of cameras all across the world that we can use as our remote eyes [1, 2]. Video can give us information that static images can not (telepresence) and with further development of technology and Internet infrastructure the speed of transmission and the amount of video imagery will only increase. Therefore, intelligent control of video capture by means of changing the view direction of the camera, spatial structuring of visual information, as well as generation of visual summaries are essential for successful application of this technology. To study user-interface issues of remotely operable cameras and provide a testbed for the application of computer vision techniques (motion detection, tracking, security) we developed the Internet Video Server (IVS) system [3] which we recently expanded with the GlobalView interface.

The next chapter is on the GlobalView extension of IVS which enables a much more intuitive remote control of the camera, especially if the connection is slow. Also the different approaches to panoramic image generation are considered. The third chapter describes how IVS and GlobalView can be used to make visual summaries. Conclusions in chapter four give a short summary and some ideas for future research.

2 GlobalView extension

Internet Video Server (IVS) [3] enables live video transmission and remote camera control (pan & tilt) over the Word Wide Web. We expanded the IVS system with the GlobalView interface to give IVS users a better spatial perception of the observed location and a more intuitive control of the camera platform. We first describe how the panoramic views are acquired and then how the panoramic view is used in the IVS interface.

2.1 Panoramic views

Panoramic views have been traditionally generated by special photographic cameras and photographic techniques by means of rotation of the whole camera or just the aperture in the camera lens. To be used in a computer system the photographic film must be first scanned which, needless to say, prevents any real-time application.

To generate panoramic images using video camera two general approaches are known:

1. using special omidirectional sensors and
2. using conventional image-based systems

2.1.1 Omnidirectional video

The first approach, which is becoming more and more popular, is using specialised cameras or camera systems that are able to acquire omnidirectional visual information [5]. Optics of such sensors use a fish-eye lens or combine standard lens on a video camera with a conic mirror [7], a spherical mirror [8], or a paraboloidal mirror [5].

These images, which cover a complete half sphere, must be mathematically processed to free them of severe distortions to get a proper perspective view. The advantage of this approach is that the whole panorama is imaged at once and that several users can each move their own "virtual camera" over this image to observe the part they are interested in. However, the benefit of of such single step image capture is reduced by very uneven resolution of these panoramic images. The majority of the image covers the sky or the ceiling of indoor spaces while the usually more interesting parts of the image are on the boundaries where the resolution is the poorest. To get information below the horizon, two such parabolic mirrors and cameras must be applied at once.

2.1.2 Camera rotation

The second approach involves camera rotation and/or integration of overlapping images taken with a regular camera. By panning the camera over a scene and composing the video frames, large panoramic images of arbitrary shape and detail can be created[4]. To automatically construct those panoramic images, however, we must derive the alignment (warping) transformations based directly on images or some other parameters that are gained automatically, rather than relying on manual intervention. If the camera direction information is automatically available it can be used as a warping parameter. This makes possible fast automatic panoramic image composition which can be applied on static scenes. This method in unappropriate for dynamic scenes since the panoramic image is generated gradually.

2.2 Generating panoramic views with a pan-tilt manipulator

We generate 360^o panoramic views by turning the camera (in horizontal and, if necessary, in vertical direction) assembling the pictures into a single slit. To get a rectangular panoramic image the individual images must be transformed from sphere to cylinder coordinates. If we are scanning only the area in the level of the camera horizon with a small vertical angle this transformation is not necessary since the image distortion is small. In our case, the vertical angle of panoramic images is about 90^o and therefore the transformation is necessary to assure smoothness of panoramic images. The panoramic images obtained in this way have a uniform resolution.

The ideal situation assumes that the camera is located in the centre of a sphere and that all the camera can see is the inside surface of the sphere. The sphere panoramic image must be transformed to fit onto the surface of the cylinder. Image 1 shows how this transformation is performed.

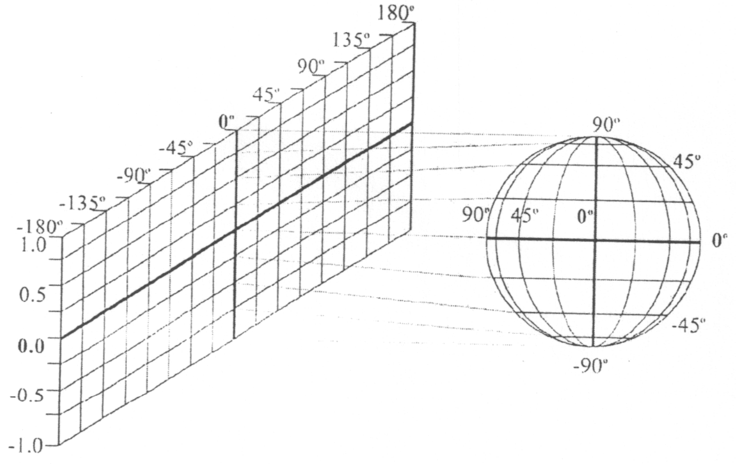


Figure 1: Projection of the images on the sphere onto the cylindrical surface.

Every image taken is a projection of the part of the sphere onto the plane whose normal vector is parallel to the camera's optical line. In sphere coordinates every point on the sphere surface can be represented as $p(\varphi, \vartheta)$, where φ is angle in horizontal direction ($\varphi \in [-\pi, \pi]$) and ϑ is angle in vertical direction measured from the sphere horizon ($\vartheta \in [-\pi/2, \pi/2]$). Every point on the cylinder surface can be represented as $u(\psi, v)$, where ψ is the angle in horizontal direction ($\varphi \in [-\pi, \pi]$) and v is the elevation. To transform the image on the sphere surface into the cylindrical panoramic image every point $p(\varphi, \vartheta)$ is transformed into the corresponding point $u(\psi, v)$ on the cylinder surface as follows:

$$\psi = \varphi$$

$$v = \frac{\vartheta}{\pi/2} R$$

R represents the radius of the sphere and can be set to 1.

In order not to loose the resolution in vertical direction the elevation v is mapped from ϑ as the vertical arc length between the corresponding point on the sphere surface and the point on the horizon of the sphere with the same horizontal angle φ. Elevation v is now in range of $[-1, 1]$. The horizontal resolution is decreasing from maximum at the elevation 0 to the minimum at the elevations 1 and -1. This horizontal resolution decrement can be formulated as the function of vertical angle ϑ this way:

$$HorRes = MaxHorRes \cdot \cos(\vartheta)$$

As can be seen that the resolution at the elevation of -1 and 1 (the north and the south pole of the sphere) is 0, this means that the north and south poles of the sphere are transformed into the top-most and bottom-most line in the cylindrical panoramic image.

The number of images needed to generate a panoramic image is $N_h \times N_v$ where N_h is the number of images taken in each horizontal plane and the N_v is the number of horizontal scans. For example, our camera's horizontal and vertical view angles are 20^o and 15^o, respectively, and we want to produce the panoramic image which would cover 360^o in the horizontal direction and 90^o in the vertical direction. The minimum number of images taken for that would be $(360/20) \times (90/15) = 108$.

To achieve smooth lines and avoid the need of any geometric transformation of images a brute-force scanning approach can be used. In the brute-scanning approach only a few center columns are taken from each image since the distortion increases from the center vertical line outward and only the center column has no distortion at all. Therefore, the brute-force scanning approach increases the number of images significantly. The number of center columns taken from each image is a compromise between quality and time that we need for scanning the whole panoramic image. The properties of the scene that we are scanning should be considered also (scene dynamics, structure, light sources, etc.).

Even if the camera direction information is available, as in our case, we still need some additional parameters to perform fast panoramic image construction without need to search for translation vector between two consecutive images. We need to know the horizontal and vertical view angles of the camera lens. By knowing those two parameters we can perform automatical image composition without the need to calculate the relative position between two consecutive images from images them selves. For every camera direction we know the precise position of the captured image within the whole panoramic image.

2.3 Camera calibration

To be able to automatically register images directly from knowing the camera viewing direction the camera lens horizontal and vertical view angles are required. We have developed an algorithm that calculates those two camera parameters and is designed to work with cameras where zoom settings and other internal camera parameters are unknown. The algorithm is based on the accuracy of the pan-tilt unit on which the camera is mounted. The basic idea of the algorithm is to calculate the translation between two images while the camera has been rotated in the horizontal or vertical direction. When the exact angle of rotation is known, the image angles can be calculated.

The complete calibration algorithm for the one axes consists of the following steps:

1. Position the pan-tilt unit to the base position for the calibration.
2. Get the reference image.
3. Turn the pan-tilt unit for the very small angle in direction which is beeing calibrated.
4. Get the image from this position.
5. Calculate the translation between those two images.
6. Calculate the first raw approximation of the camera view angle.
7. Turn the pan-tilt unit to the calculated position that should corespond to some predefined offset (for example 1/4 of the image).
8. Get the image from this position.
9. Calculate the translation between this and the reference image.
10. Calculate the corrected view angle.

11. If the view angle correction is small enough or some constant number of steps has been performed then finish otherwise go to step 7.

The resolution of the pan-tilt unit is 185.1428 arc seconds per pan or tilt position, that is, one pan or tilt position means 0.0514 degrees in the real world.

For calculation of the translation between two images the combination of two algorithms is used [9]. Basic incremental-motion estimator can obtain estimates to motion given that frame-to-frame displacements are small. The precision and range of the estimator are enhanced through coarse-fine motion tracking within a pyramid structure.

	α	β	α calculus		β calculus	
Near target	33,73	24,06	Average:	33,730	Average:	22,666
	33,73	22,22	St.Dev:	0,000	St.Dev:	0,863
	33,73	21,80				
	33,73	22,42				
	33,73	22,83				
Far target	34,35	26,13	Average:	34,350	Average:	26,130
	34,35	26,13	St.Dev:	0,000	St.Dev:	0,000
	34,35	26,13				
	34,35	26,13				
	34,35	26,13				

Table 1: Algorithm results - α is estimated horizontal angle of camera and β is estimated vertical view angle of the camera. Near target was an object wood patern with clean vertical lines located approximately 40 cm from camera, while the Far target was aproximately 4 meters away from camera.

2.3.1 Camera calibration results

Using this camera calibration algorithm few things should be noted.

First the selection of the area on which calibration is performed is very important. For example: estimation of the horizontal camera view angle would fail when using the scene with parallel horizontal lines. This can be seen in table 1 in estimation of β - vertical view angle for the near target. The calibraton was performed on the wood patern with clean vertical lines. In horizontal direction the estimation was done well. Since this algorhitm performs astimation of the horizontal and the vertical camera view angles separately it enables the selection of different area (viewing direction) for each direction.

Second, the distance from camera to objects on the scene on which calibration is performed is important. Since usual cameras do not have telecentric optics the view angles are changed when focusing to the objects with different distance from the camera. Therefore the best solution is to perform the camera view angles estimation directly on the scene which is to be scaned. In table 1 this change of view angle is quite obvious when comparing the estimated view angles between far and near target

2.4 Results

On our equipment the scanning of panoramic pictures in the brute-force scanning manner (without geometric transformation) takes from a few seconds (building the panorama out of complete, non-overlapping images) to about 3 minutes (taking only the 5 vertical lines in the center of each image) resulting in coarser or finer vertical image joints. Fig. 2 shows a panoramic picture taken in a single horizontal scan with a fish eye lens.

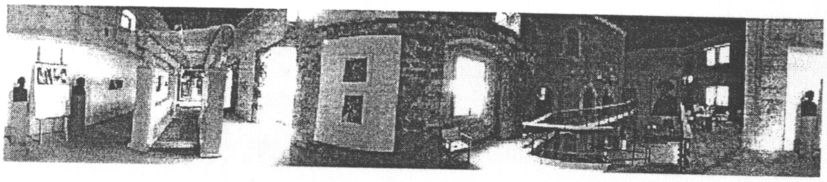
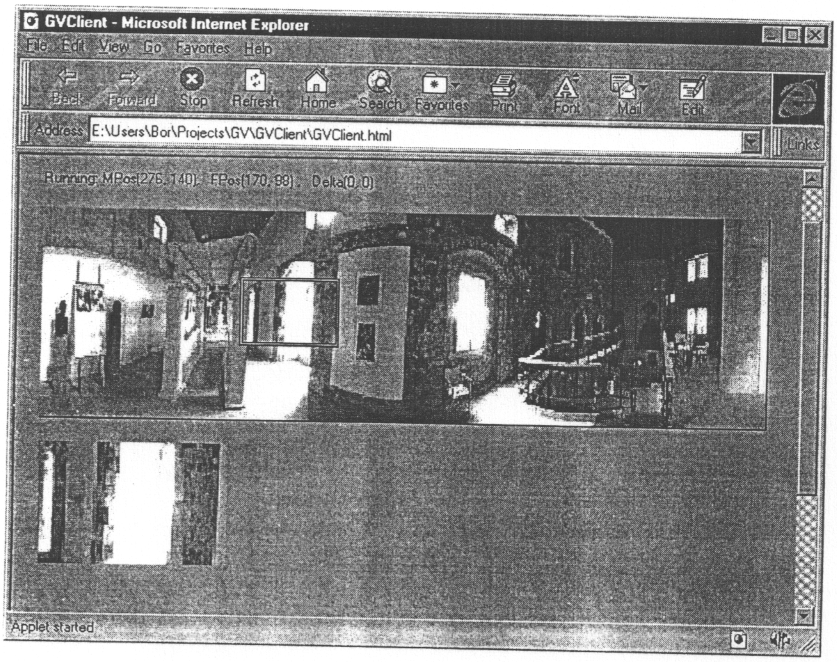


Figure 2: 360^o panoramic image generated with only one horizontal scan using a wide angle lens

2.5 Integration of the panoramic view into the IVS

GlobalView interface (Fig. 3) is a combination of a static panoramic view and live images received from IVS. Live images arriving from IVS are transformed from spherical to cylinder coordinates and superimposed on the corresponding position in the panoramic view.

At system startup, the panoramic image is generated first by scanning the surroundings of the camera. When a user starts to interact with the system he is shown the whole panoramic image. A rectangular frame in the panoramic image indicates the current direction of the camera. Inside this attention window the live video image is seamlessly superimposed onto the panoramic image. By means of a mouse the user can move the attention window and in this way control the direction of the camera. When the attention window is at the desired location the user prompts this to the



E:\Users\Bor\Projects\GV\GVClient\GVClient.html

Running: MPos(276, 140), FPos(170, 98), Delta(0, 0)

Figure 3: GlobalView interface of IVS. The rectangle in the panoramic image indicates the attention window which in turn indicates 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 an enlarged live video image is shown.

system. The position of attention window within the panoramic image is then mathematically transformed into the physical coordinates of the next camera direction and the appropriate command is issued to the camera.

The whole interaction with the camera is carried out through the attention window. Moving this attention window results in moving the camera. At any time, only one user can move the camera since allowing more than one user to interact with it could be confusing.

To allow fair time sharing every user can control the camera for few minutes and after that, the control is passed to the next user waiting in the queue. Other users can only observe the remote area. Their control of the attention window is disabled. Only the first user in the queue – the active user can move the attention window. Of course in remote survailance applications the operator is usualy only one and this multi-operator feature is generaly not required.

Since the panoramic image is rather large in comparison to the live video image and the attention window small, a separate window for live video can be present to make a more detailed observation possible. In this way the attention window acts like a magnifier glass over the remote location.

The client-side user interface was written in Java (Java applet) which is run within any web browser that supports Java. This approach enables platform independece which is almost a must in these days. With minor changes, it can be run on any Java virtual machine.

At present, only one focus rectangle is present since only one camera is available. In general, several cameras could be controlled by the same number of attention windows. A combination of cameras with different focal length would be possible. Zoom lenses could be controlled by resizing the attention window. An interesting combination would be a live panoramic image attainable by a system such as the Omnicam [6] and the IVS system offering a combination of peripheral, low resolution image, and a high resolution focal image. Such an interface would allow a better overview of remote areas without a complicated user interaction.

3 Generation of visual observation summaries

When using the IVS we sometimes wish to see what was happening at a remote location in the past. For example, one would want to check who entered a building in some specified time interval, or just to keep track of the number of people on a specific location.

To enable this kind of functionality the live video frames from the remote location should be saved in some appropriate form together with the time stamp and the information about the camera direction. Since some parts of the observed environment might not change frequently (i.e. ceilings) there is no need to save many basically identical images. Recorded pictures are compared by some statistical measure to the already saved images when the camera was pointed in the same direction. If the difference is below a certain threshold, only a dummy image (time stamp and a pointer to the previous image) is saved. This way the significant amount of disk space could be saved.

The system could operate in two modes: in active or passive mode. When operating in passive mode the control over the camera would be completely in the hands of an operator who would control the direction of the camera. The system's only job would be the recording of the user actions and saving of the video images.

When operating in active mode, the system should autonomously perform continuous observation of the areas of interest. These areas of interest could be predefined (like doors, windows, paintings on the walls, etc.) or could be extracted automatically. Automatic extraction of the areas of interest could be done by finding the areas of high level of change. A possible way to do this is first to scan the entire area a few times and find the areas in the panoramic image with the highest degree of change. Change or difference between two (sub)images can be computed in a number of different ways. We experiment now with fairly simple pixel-based statistical methods. In this way, a priority list of the areas of interest could be generated. The system would check the locations higher on the list more often. If new areas of large change would arise the priority list could be updated dynamically. Of course, the entire area could be defined as an area of top interest and the system would then continuously scan the entire area. Different intelligent schemes for visual surveillance using IVS are still under consideration. We are integrating into the IVS system a simple motion detection method which would enable the camera to automatically track a moving object.

In addition, a tool for visual report generation which will allow custom queries is under construction. The basic feature of the system will be the playback facility which will paste images on the appropriate place in the panoramic image in chronological order. Different options for this playback will be available:

- selecting the speed of the playback,
- selecting the time frame,
- playing back only those images in which a change occurred,

- playing back only images which are in a specific subframe in the panoramic image,
- tracking of selected objects over time.

For the generation of visual summary reports we are using a SQL database which enables image base queries and reports. The Informix Universal Server seems to be a good solution since it is a powerful SQL database which can be fully customized with so called DataBlades. DataBlade is a set of functions which can be applied on the data within a database. This enables definition of user defined filters which can be used as image evaluation functions in queries and reports. It has a built-in web support so clients could request and see the results of different kinds of customized queries within their favourite WEB browser.

4 Conclusion

IVS is a system which enables live video transmission over the Internet. With the GlobalView interface, which generates a panoramic image of the whole environment, the observer gains a better understanding of the observed location and a more intuitive control of the camera.

Video-conferencing and remote surveillance are examples of applications that would certainly benefit from such an interface.

Our future work is directed to integration of visual information from several cameras (i.e. relay-tracking) and visually-based teleoperation of a mobile platform.

References

[1] Webcam32 Gallery
`http://www.kolban.com/webcam32/forms/gallery_read.hts`

[2] Tommy's List of Live Cam Worldwide,
`http://www.rt66.com/ ozone/cam.html`

[3] B. Prihavec, A. Lapajne, and F. Solina. Active video observation over Internet. In B. Zajc and F. Solina, editors, *Proceedings Fifth Electrotechnical and Computer Science Conference ERK'96*, Vol. B, pages 117–120, Portorož, Slovenia, September 1996. IEEE Region 8, IEEE Slovenia Section.

[4] R. Szeliski. Video Mosaics for Virtual Environments. *IEEE Computer Graphics and Applications*, 16(2):22-30, 1996.

[5] F. Hamit. New video and still cameras provide a global roaming viewpoint. *Advanced Imaging*, March 1997, pp. 50–52.

[6] Shree K. Nayar. Omnidirectional Video Camera. Proc. of DARPA Image Understanding Workshop, New Orleans, May 1997.

[7] Y. Yagi, S. Kawato and S. Tsuji. Real-time omnidirectional image sensor (COPIS) for vision-guided navigation. *IEEE Trans. on Robotics and Automation*, 10(1):11–22, 1994

[8] J. Hong, X. Tan, B. Pinette, R. Weiss, E. M. Rieseman. Image-based homing. *Proc. IEEE Int. Conf. on Robotics and Automation 1991*, pp. 620–625, 1991

[9] J. R. Bergen, P. J. Burt, R. Hingorani and S. Peleg. Computing Two Motions from Thre Frames.

[10] Jožef Plečnik exhibition at Prague Castle, `http://razor.fri.uni-lj.si:8080/Plecnik/`

[11] ROTAS-TENET, `http://razor.fri.uni-lj.si:8080/Rotas/`

[12] Netropolis – Cyborg's eye, artinternet installation project, European Cultural Month Ljubljana 1997, `(http://razor.fri.uni-lj.si:8080/Netropolis-ECML97)`