Spherical Panoramas for Pan-Tilt Camera Motion Compensation in Space-Variant Images

Filiberto Pla and V. Javier Traver

Dep. de Llenguatges i Sistemes Informàtics
Universitat Jaume I,
E12080 Castelló, Spain
{pla,vtraver}@uji.es
http://vision.uji.es

Abstract. In active vision scenarios, the motion of the observer induces an apparent motion in the image plane. One approach for camera motion compensation is the use of panoramic images, representing the scene at the different positions of the camera. In this work, an approach to build spherical panoramic views from a pan-tilt camera is described, which is based on background updating techniques. Interestingly, panoramic representations allow motion detection and analysis to be performed independently from camera movements. This makes easier the detection of moving objects in active tracking tasks. Additionally, the advantages of combining spherical panoramas with log-polar images is discussed. An example of target segmentation by background difference is shown, which shows the effectiveness of panoramic representations in active vision systems.

1 Introduction

One of the problems to be solved in active vision systems, where the camera is moving, is to overcome the effect of *apparent motion* due to egomotion [6, 3]. This is the case of systems based on a pan-tilt camera performing known rotational movements. Active monocular target tracking with a pant-tilt head is a valid framework for some visual surveillance applications, like people or vehicle monitoring, where the system focuses its attention on certain target.

One way to achieve a representation in which camera motion is compensated consists of building a panoramic view [9]. A panoramic view acts as a reference image, which has compensated the effect of camera motion from all images taken by the camera into a single image representation. This approach has been used in video coding techniques [7] to extract moving objects from an image sequence.

The approach presented here has been inspired in works such as [1], where image mosaics built from a pan-tilt camera motion have been used to perform motion segmentation. Unlike this and similar works, a more adequate model of panoramic representation is proposed, to deal both, with a pan-tilt system and with the possibility of using a non-uniform image sampling sensor, like a log-polar imaging system [8].

M.T. Escrig Monferrer and F. Toledo Lobo (Eds.): CCIA 2002, LNAI 2504, pp. 375–387, 2002.

An accurate way to perform registration to build the panorama, based on a motion estimation technique previously developed in [5], is also proposed. Previous works on building panoramic images have focused on constructing image mosaics for visualization applications, where accuracy is not as important as it is in moving target segmentation and motion analysis tasks. Figure-ground segmentation is here done using an approach reported by [2], which is more independent from tuning parameters.

Next section describes the proposed panoramic model, pointing out its properties and the relation with non-uniform sampled images. Section 2 shows an example of how these panoramic representations can be used for motion detection and analysis. Section 3 presents the results of some experiments carried out using the described techniques. Finally, some conclusions from the present work and the way we can use it in future work are drawn in Sect. 5.

2 Building Panoramic Views

2.1 Notation

There are several types of panoramic representation to deal with image information in a scene. The kind most often used is an image plane projection onto a reference image frame [7,4]. Panoramic views (or image mosaics) generated from camera pan-tilt movements are also possible [1,2].

Other types of panoramic representations are cylindrical and spherical panoramas [9]. *Cylindrical* panoramas are quite adequate for building panoramas from pan camera movements, because given the camera focal length, we only have to bring out the pan angle rotated for the camera to match the image contents with the panoramic representation. *Spherical* panoramas could be a more natural way of building panoramic views generated from pan-tilt movements. In a spherical panorama, world coordinates of a point $p = (x, y, z)$ are mapped into 2D spherical coordinates (θ, ϕ), which represent the direction of point p in the world coordinates. These spherical coordinates are calculated from the 3D world coordinates as

$$\begin{cases} \theta = \arctan\left(\sqrt{x^2+y^2}/z\right), & \theta \in [0, \pi] \\ \phi = \arctan\left(y/x\right) \qquad\qquad, & \phi \in [-\pi, \pi] \end{cases} .$$

If a pin-hole camera model is assumed (Fig. 1), it is possible to compute $f(\theta, \phi)$, corresponding to a world point $p = (x, y, z)$ as

$$\begin{cases} \theta = \arctan\left(\sqrt{{x'}^2+{y'}^2}/f\right) \\ \phi = \arctan(y'/x') \end{cases} , \tag{1}$$

where (x', y') are the image coordinates, and f the focal length of the camera.

In the case of a log-polar image $I(u, v)$, and under an ideal mathematical model, we have that $u = \log\left(\sqrt{x'^2+y'^2}\right)$ and $v = \arctan(y'/x')$. Therefore, the spherical coordinates can be computed as

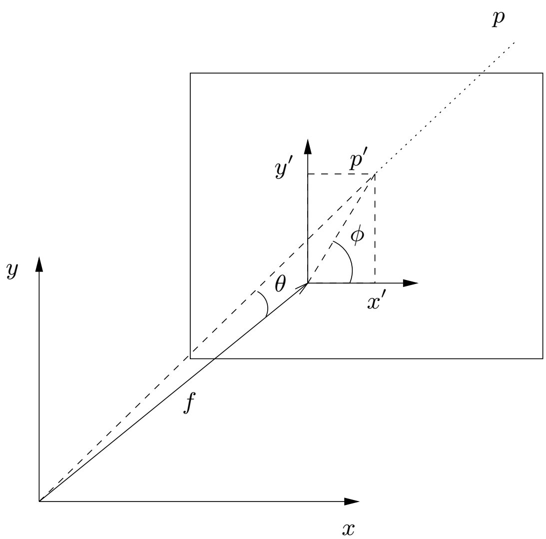

Fig. 1. Pin-hole camera model and spherical coordinates of a world point

$$\begin{cases} \theta = \arctan(e^u/f) \\ \phi = v \end{cases} . \tag{2}$$

When the camera rotates a pan angle of α radians about y axis, and a tilt angle of β radians about x axis (Fig. 2), the (θ_r, ϕ_r) spherical coordinates of p with respect to the new camera position are related to the coordinates (θ, ϕ) of the initial camera position as

$$\begin{pmatrix} \sin\theta_r \cos\phi_r \\ \sin\theta_r \sin\phi_r \\ \cos\theta_r \end{pmatrix} = \boldsymbol{A}(\alpha, \beta) \cdot \boldsymbol{B}(\theta, \phi) , \tag{3}$$

with

$$\boldsymbol{A}(\alpha, \beta) = \begin{pmatrix} \cos\alpha & \sin\alpha \sin\beta & \sin\alpha \cos\beta \\ 0 & \cos\alpha & -\sin\beta \\ -\sin\alpha & \cos\alpha \sin\beta & \cos\alpha \cos\beta \end{pmatrix} ,$$

and

$$\boldsymbol{B}(\theta, \phi) = \begin{pmatrix} \sin\theta \cos\phi \\ \sin\theta \sin\phi \\ \cos\theta \end{pmatrix} .$$

Thus, by knowing the pan and tilt angles (α, β) of the camera movement, a back-mapping can be performed from the spherical coordinates (θ, ϕ) of the image points, in a given camera position, with respect to the spherical coordinates of the initial coordinate system, which represents the panoramic representation.

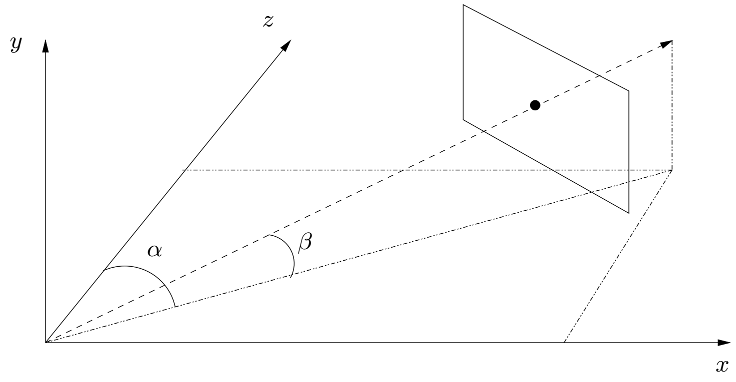

Fig. 2. Pan-tilt movement of a camera with respect to the reference coordinate system

Algorithm 1 Building a panoramic background image from a set of camera motions

Initialize the panoramic image $P(\theta, \phi)$ with ∞
Rotate camera a certain pan-tilt angles (α_i, β_i)
for all (θ, ϕ) in the panoramic image **do**
  Compute the corresponding (θ_r, ϕ_r) in the rotated camera axis by using the Equation (3).
  Compute the current image plane coordinates using the inverse relations either of (1) or of (2), depending on using either cartesian or log-polar images, that is:

$$\begin{cases} x' = f \tan\theta \cos\phi \\ y' = f \tan\theta \sin\phi \end{cases} \qquad \begin{cases} u = \log(f \tan\theta) \\ v = \phi \end{cases}$$

  if $P(\theta, \phi) = \infty$ **then**
    $P(\theta, \phi) \leftarrow I(x', y')$
  else
    $P(\theta, \phi) \leftarrow (1 - \lambda) P(\theta, \phi) + \lambda I(x', y'), \lambda \in [0, 1]$
  end if
end for

2.2 Algorithm

An accurate computation of the pan and tilt angles is important to perform a good mapping of the images taken at different camera positions and, in turn, to achieve a good accuracy in the panoramic representation. To that end, some kind of pan-tilt camera calibration is needed [10].

To build a panoramic background image from a set of different camera rotations (α_i, β_i) with respect to the initial camera position, Algorithm 1 is proposed.

The parameter λ is used in a simple auto-regressive filter to integrate the background information from the different views. In all cases, for log-polar images, $I(u, v)$ has to be used instead of $I(x', y')$.

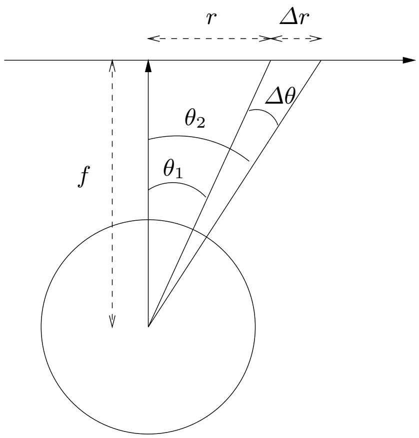

Fig. 3. Relation between the panoramic image and the image plane resolution

The heaviest computational effort in this procedure lies in computing the (θ_r, ϕ_r) coordinates in a rotated coordinate system after some pan-tilt camera movement (i.e., the operations described in Equation (3)). At this moment, this operation has to be performed for every pixel in the image. Alternative ways to perform this coordinates transform are possible (e.g., using quaternions), and should be investigated to increase the efficiency.

Calibration of the intrinsic camera parameters (f) can be done by means of some standard calibration algorithm [11]. The possible values of (x', y') or (u, v) for all predefined values of the panoramic image (θ, ϕ) can be precomputed and kept in LUTs.

The pan and tilt angles (α_i, β_i) rotated by the camera can be obtained more accurately after a movement by these steps:

1. Read the encoder pan-tilt positions of the camera motors.
2. Use these pan and tilt angles as initialization values to the translational motion estimation between the reference and the current image, by using a motion estimation algorithm [5].
3. Find out the initial translational input as $(x_0, y_0) = (x_r\alpha_i, y_r\beta_i)$, with x_r and y_r being constants worked out by the calibration process.
4. Approximate the pan and tilt angles by $(\alpha_i, \beta_i) = (x_0/x_r, y_0/y_r)$.

2.3 Image Resolution Issues

The resolution or size of the panoramic image can be chosen according to the desired resolution in the image plane. To take this decision, it has to be taken into account that, given a spherical panoramic representation, the projection of the panoramic information (stored in uniform resolution in θ, ϕ) into an image plane in a certain orientation, does not provide a uniform resolution in the x' and y' axes in the image plane. In fact, we can approximate (see Fig. 3) the

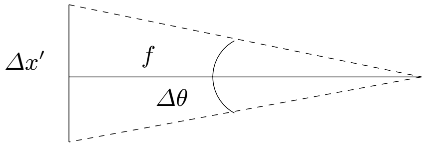

Fig. 4. Relation between spherical image resolution and image plane resolution

relation between the width of a pixel in the panoramic image $\Delta\theta$, and the width Δr of the area (pixel width) covered by a bin in the panoramic image into the image plane as

$$\Delta r = f \cdot \frac{\Delta\theta}{1+\theta_1^2}$$

since

$$\begin{aligned} \tan(\theta_1) &= r/f, \\ \tan(\theta_2) &= (r+\Delta_r)/f \\ \Delta r &= f(\tan(\theta_2) - \tan(\theta_1)), \\ \theta_2 &= \theta_1 + \Delta\theta \\ \tan(\theta_1 + \Delta\theta) &= \tan(\theta_1) + \Delta\theta/(1+\theta_1^2) \end{aligned}$$

Therefore, a *uniform* sampling in the spherical representation approximately becomes a *quadratic* sampling in the image plane, which means that resolution decreases in a quadratic way when we move away from the image center. This means that resources are wasted if cartesian images are used as a source to build spherical panoramic images. And this is so because the resolution of the image plane far away from the image center (the optical axis) is not used when this peripheral information is projected onto the surface of the sphere representing the panoramic image.

On the contrary, building spherical panoramas from log-polar images makes some sense, as it is illustrated in the following observations:

1. The fact of mapping the image plane information into the panoramic representation takes advantage of the decreasing resolution occurring in log-polar images far from the image center. It is true, however, that resolution in log-polar images decays exponentially, rather than quadratically.
2. Similarly, the back-projected information from the panoramic view to a log-polar image will not suffer from the decrease of the projected image resolution.
3. High-resolution spherical panoramic views could be built from different log-polar images, integrating the high fovea resolution at different views in a single representation.
4. There is a straightforward relation between the pixel coordinates in log-polar images and the corresponding spherical coordinates in the panoramic image (see Equation 2).

To have a spherical panorama with enough resolution to take advantage of the image resolution of the images the panoramic view is built from, if we would like to build a panoramic that covers half the 3D space, that is $\theta \in [0, \pi/2]$ and $\phi \in [-\pi, \pi]$, the following relation has to be taken into account (Fig. 4):

$$\begin{aligned} \Delta\theta &= 2\arctan(\Delta x'/2f) \\ \pi/(2\Delta\theta) &= M \end{aligned}$$

with M being the number of bins along θ in P.

3 Motion Detection in Spherical Panoramas

Given that a panoramic view is a representation built in a incremental way from a sequence of images taken with different positions of the camera, the idea of building the panoramic image that represents the background, could be mixed with the process of detecting changes in the source image with respect to the panoramic representation when we are performing the mapping.

In other words, while updating the panoramic representation, which represents the background, some procedure could be done while doing the updating to extract information about what is moving with respect to the panoramic background. Thus, if the panoramic background is supposed to be static and free of the apparent movement created by the camera motion, any change in the present image when mapping it into the panoramic representation will be due either to a different motion present in the images or to a change in the scene illumination with respect to the camera.

A straightforward way to see the above mentioned effect is to use a difference-based approach to detect motion changes. In this case, the difference operation will be performed between the panoramic image representation $P(\theta, \phi)$ and the present image $I(x', y')$ taken at a given camera pan-tilt position (α, β) with respect to the reference coordinate system of the panoramic representation.

The algorithm used to perform this image difference or background subtraction, is based on the approach used by [2]. Therefore, the procedure described in Algorithm 2 is used to perform the panoramic background image updating and the detection of possible moving pixels in the current image.

$V(\theta, \phi)$ is a variance image where the evolution of the variance of the gray levels of a pixel in the panoramic view is kept and updated, integrating the actual variance (or standard deviation) of the pixel in the current image with respect to its mapping in the panoramic representation.

The binary image $T(x', y')$ resulting from the process will represent the possible moving pixels due to moving objects with respect to the static background, the latter being represented by the panoramic image. The procedure described above has the advantage of having no *a priori* thresholds to set up, thus getting rid of tuning parameters.

Other types of motion analysis could be done using the information of the current image taken by the camera. For instance, one can use log-polar images rather than cartesian images, and compare the contents of these images with the

Algorithm 2 Panoramic image updating and motion detection

Given an image $I(x', y')$ taken from a certain pan-tilt (α, β) camera movement
for all (θ, ϕ) in the panoramic image **do**
  Compute the corresponding (θ_r, ϕ_r) in the rotated camera axis by using the Equation (3.
  Compute the current image plane coordinates using the following expressions, depending on using either cartesian or log-polar images:

$$\begin{cases} x' = f \tan\theta \cos\phi \\ y' = f \tan\theta \sin\phi \end{cases} \qquad \begin{cases} u = \log(f \tan\theta) \\ v = \phi \end{cases}$$

  if $P(\theta, \phi) = \infty$ **then**
    $P(\theta, \phi) \leftarrow I(x', y')$
    $V(\theta, \phi) \leftarrow \sigma_0$ {some initial value}
  else if $|I(x', y') - P(\theta, \phi)| < V(\theta, \phi)$ **then**
    $P(\theta, \phi) \leftarrow (1 - \lambda)P(\theta, \phi) + \lambda I(x', y'),\ \lambda \in [0, 1]$
    $V(\theta, \phi) \leftarrow (1 - \lambda)V(\theta, \phi) + \lambda |I(x', y') - P(\theta, \phi)|$
    $T(x', y') \leftarrow 0$
  else
    $T(x', y') \leftarrow 1$
  end if
end for

Fig. 5. From left to right, frames 1, 5 and 17 in the sequence

current state of the panoramic representation to segment the target and estimate its motion parameters, with the advantage of having compensated the apparent motion of pan-tilt camera movement.

4 Experimental Results

A spherical panorama has been built from a real sequence of images taken from a pan movement, using a Sony EVI-G21 camera. In this case, a sequence of 17 images was taken. The pan angle was computed by the procedure described in Sect. 2, using also the algorithm reported in [5] to estimate more accurately the rotation performed, in order to calculate the mapping and updating of the panoramic view.

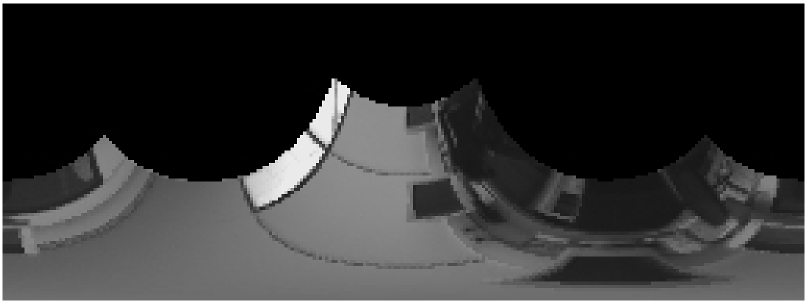

Fig. 6. Spherical panoramic view built from sequence in Fig. 5

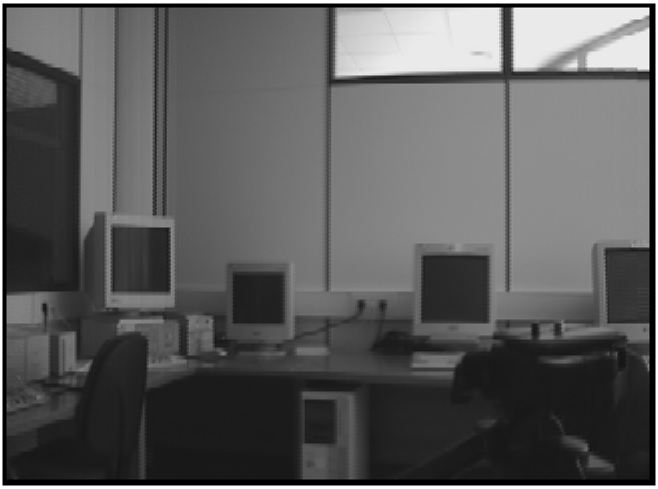

Fig. 7. Panoramic view projected in the image plane of the first image in the sequence

Figure 5 shows some frames in the sequence, and Fig. 6 shows the spherical panoramic representation built. Rows in Fig. 6, image denote the θ angle, while columns denote the ϕ angle. Figure 7 represents the same panoramic image as that in Fig. 6, but represented in the image plane corresponding to the initial image in the sequence, in order to appreciate the quality of the mapping. A second experiment has been performed consisting of pasting on the sequence of images of Fig. 5, a target region that undergoes a movement of scaling, rotation and translation with respect to the image plane. Figure 8 shows three images of that sequence with an X-shaped target.

Therefore, if two subsequent images in the sequence were compared, two different movements will be detected, the one corresponding to the target and the one corresponding to the apparent motion of the pan movement of the camera. On the other hand, Fig. 9 shows the result of applying the algorithm described in Sect. 2 for simultaneous background updating and motion detection. In these

Fig. 8. From left to right, frames 1, 5 and 15 in the sequence with the pasted X-shaped target

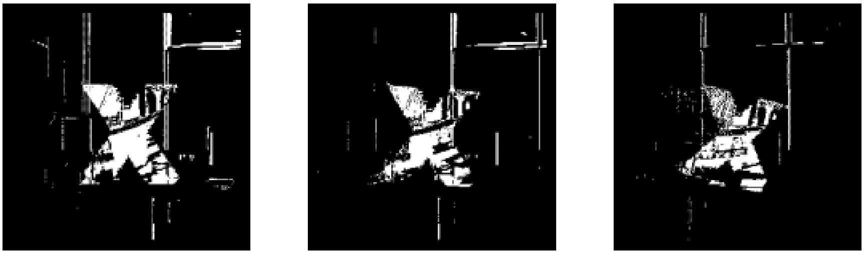

Fig. 9. Moving and grey level changing pixels between the images in the sequence and the panoramic background image. From left to right, the moving pixels corresponding to frames, 5 and 15 in the sequence

images, we can see how, apart from the obvious detection of pixels at borders of objects because of the effect of occluding pixels from view to view, the main area of the moving target has been detected, while the background has been considered with no moving changes, due to the effect of camera motion compensation in the panoramic representation.

Figure 10 shows the same images in the sequence in Fig. 5, but with cartesian images reconstructed from log-polar images sized 64×128. The corresponding panoramic image projected in a cartesian image plane referred to the first image in the sequence is shown in Fig. 11. Note that, because of the rightward pan movement; the resolution in the middle of the image has increased when moving this direction, due to the effect of mixing the information of the fovea through the sequence.

Figure 12 shows the result of the algorithm for image change detection at some frames in the sequence. Note that, the results are quite similar to those in Fig. 9 but with a "smoothing" effect due to the lower resolution far from the image center. Therefore, these results provide an idea of the potential of this technique for motion analysis where camera motion is compensated in a pan-tilt camera. Moreover, the use of spherical panoramic images are a suitable approach to take advantage of space-variant images, like log-polar images, because of the

Fig. 10. From left to right, log-polar image 1, 5 and 17 in the sequence

Fig. 11. Panoramic view projected in the image plane of the first image in the sequence built from log-polar images

effects of the projection of the image plane information in the panoramic image, as commented in the previous section.

5 Conclusions

Camera motion compensation in a pan-tilt camera system can be achieved by building panoramic representations of a scene with a static background. The approach presented here describes a way of building spherical panoramas from a sequence of frames taken from a pan-tilt camera.

Spherical panoramas offer some properties that could be exploited by using log-polar images because: (i) this panoramic representations produces an effect of non-uniform image resolution when projecting its contents into an image plane in a given camera direction; and (ii) there exists a straightforward relation between spherical and log-polar coordinates.

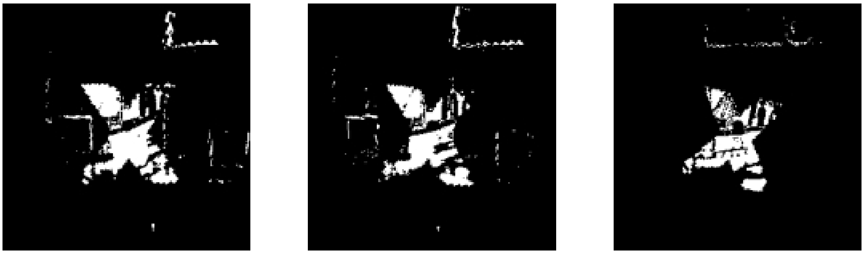

Fig. 12. Moving and grey level changing pixels between the log-polar images in the sequence and the panoramic background image. From left to right, the moving pixels corresponding to frames 1, 5 and 15 in the sequence

The results of the experiments carried out show that panoramic representations can be used for the detection and analysis of targets that move with respect to a static background. An example of image difference technique has been shown, to illustrate how this process could be implemented, by fusing the panoramic image representation information with the contents of the current image taken from a given pan-tilt direction.

Future work could be done directly using the proposed approach to detect moving regions in a pan-tilt system for target monitoring, and combine it with log-polar image representation, in the way described in this work.

References

1. K.S. Bhat, M. Saptharishi, and P.K. Khosla. Motion detection and segmentation using image mosaics. In *IEEE International Conference on Multimedia and Expo*, volume 3, pages 1577–1580, 2000.
2. A.R.J. Fran ois. *Semantic, interactive manipulation of visual data.* PhD thesis, University of Southern California, 2000.
3. J.E. Ha and I.S. Kweon. Robust direct motion estimation considering discontinuity. *Pattern Recognition Letter*, 21(11):999–1011, 2000.
4. K. Mase and H. Nishira. Computing the field-of-view of a stitched panorama to create fov sensitive virtual environments. In *International Conference on Pattern Recognition, ICPR '96*, volume I, 1996.
5. R. Montoliu and F. Pla. Multiple parametric motion model estimation and segmentation. In *IEEE Intl. Conf. on Image Processing (ICIP)*, volume II, pages 933–936, Thessaloniki, Greece, October 2001.
6. J.M. Odobez and Bouthemy P. Detection of multiple moving objects using multiscale mrf with camera motion compensation. In *International Conference on Image Processing, ICIP'94*, volume II, pages 257–261, 1994.
7. F. Odone, A. Fusiello, and E. Trucco. Robust motion segmentation for content-based video coding. In *6th RIAO (Recherche d'Informations Assist e par Ordinateur) International Conference*, Paris, France, 2000.
8. Fernando Pardo-Carpio. *Sensor Retínico Espacio Variante Basado en Tecnología CMOS.* PhD thesis, Dept. Informàtica i Electrònica, Universitat de València, September 1997.

9. R. Szeliski and H. Shum. Creating full view panoramic image mosaics and environment maps. In *Proceedings of SIGGRAPH*, pages 251–258, 1997.
10. V. J. Traver and F. Pla. Motion estimation-based self-calibration of a log-polar pan-tilt camera. In *Visualization, Imaging and Image Processing Conf.*, September 2002. (Accepted for VIIP'02).
11. R.Y. Tsai. A versatile camera calibration technique for high accuracy 3D machine vision metrology using off-the-self TV camera and lenses. *IEEE Journal of Robotics and Automation*, RA-3(4):323–344, 1987.