# FUNCTIONALITIES FOR MAPPING 2D IMAGES AND 3D WORLD OBJECTS IN A MULTICAMERA SYSTEM*

*Òscar Garcia, Josep R. Casas*
UPC – Technical University of Catalonia
Image Processing Group

## ABSTRACT

We present four functionalities intended to improve the ability of image detection and tracking algorithms to understand a scene in a multicamera system. The redundancy of several available projections of any 3D object onto different cameras might ease video analysis tasks. When some prior information about the 3D object or any of its projections is known, geometric constraints can help to restrict search areas in the images under analysis. The functionalities presented also tackle the problem of selecting the best camera at any time, or computing projected areas of 3D objects in images.

## 1. INTRODUCTION

In a 'multicamera system', a number of different cameras placed in known positions observe the objects in the scene from different views. The existing redundancy across several projections of the same 3D object onto different cameras can be exploited for video analysis tasks. The main objective of this paper is to describe several functionalities useful for image analysis algorithms. The geometric constraints of a multicamera system will make it possible to considerably restrict search areas for object detection, tracking and analysis, by using properly calibrated cameras and if some prior information about the 3D object is known, such as the position in one image or the approximate 3D position in the room.

The paper is organized as follows. Section 2 briefly describes the scenario for our experiments. Section 3 deals with the process of mapping 2D image coordinates and 3D world coordinates in a multicamera system. The main problems arising in such process are pointed out. These two sections are necessary to understand section 4, which constitutes the central part of this paper, and describes the four functionalities presented. Some experimental results for each one of the functionalities under study are shown in section 5 and, finally, conclusions are discussed in section 6.

## 2. THE SMART ROOM SCENARIO

The algorithms and functionalities presented in this paper have been tested in the Smart Room at UPC. The size of the room is 5,25 m x 4,20 m and it has five fixed cameras installed: four cameras are placed on the corners of the room; and a fifth camera is placed on the ceiling. With this camera layout, most of the points in the room are viewed from at least 3 cameras.

An absolute coordinate system, adapted to the room and common to all the available cameras, has been considered in order to carry out the experiments. Once the cameras are completely calibrated, all 3D world coordinates will be referred to this absolute coordinate system.

## 3. MAPPING 2D IMAGES AND 3D OBJECTS

If we intend to benefit from the intrinsic geometric constraints of a multicamera system, two important pre-requisites must be satisfied:

- Camera calibration: It allows extracting metric information from 2D images, by relating pixel coordinates with 3D back-projected rays. A planar chess board pattern and Zhang's [1] calibration algorithm have been used in our case.
- Projection and reconstruction algorithms: which allow to bidirectionally map 2D image coordinates of different projections of the same 3D object with its 3D world coordinates. Some precise and robust algorithms have been developed [2].

Let us look further into the problem of computing the 3D coordinates of a world point given the coordinates of its projections onto several cameras. If the involved cameras are properly calibrated, the **back-projected rays** for those projections (i.e. the sets of points in the 3D world which map to the same point in the image) may be computed [3]. At least two cameras are needed in order to compute the 3D world point from its projections. If only one camera is available, any point along the back-projected ray would be a potential 3D candidate. On the contrary, if two or more projections of the same 3D point are known on different cameras, their back-projected rays will meet in a single 3D point. In real applications, however, those rays will not intersect due to the following two main reasons:

- **Calibration errors**: Real camera calibration parameters will be affected by slight errors due to

* *This material is based upon work partly supported by the IST programme of the EU through the NoE IST-2000-32795 SCHEMA*

difficulties in the detection of calibration points (in the chess board images) or inaccuracies in the lens distortion model. These errors can be minimized by adding redundancy to the problem. Several reconstruction methods are presented and compared in [2]. A method for the joint estimation of minimum reprojection error with outlier rejection is proposed to obtain a more robust and precise estimate when more than two cameras are available, even if one of them is poorly calibrated.

- **Detection errors**: Real world objects (arbitrary 3D volumes) can look quite unlike in different views. Occlusion with other objects, light intensity changes, projection sizes, or the inexistence of characteristic points –with punctual features easy to identify across images– might be some reasons for this problem.

Consequently, accurately detecting the different projections of the same 3D point across available images is not an easy task. The functionalities presented in this paper will help us to improve the results of object detection, tracking and image analysis procedures, by combining the available information in a calibrated multicamera system. More concisely, if we had detected the projection of the desired object in one camera –the one with the view that makes detection easier–, the search area in a second camera could be considerably reduced by means of the functionalities proposed in section 4.

## 4. FUNCTIONALITIES DESCRIPTION

Four functionalities intended to improve detection effectiveness in a multicamera system are presented in this section: epipolar lines, best camera selection, sphere projection and oriented circle projection. Experimental application results are presented in next section.

Briefly speaking, epipolar lines allow reducing the search area in a second camera to a neighborhood of a segment. Additionally, the best camera selection functionality could provide the closest camera, or the best oriented camera with respect to a given object. And finally, sphere and oriented circle projection will allow us to relate simple 3D objects with the sizes and shapes of their projections on all the available cameras.

### 4.1. Epipolar lines

Let us imagine that the image coordinates of the projection of a given object on one camera -say camera 1- are known in advance, but the projection of the same object is unknown for camera 2. From a geometric point of view the epipolar line is defined as the projection onto camera 2 of the back-projected ray in camera 1.

Epipolar lines in a multicamera environment allow to reduce the search area from the entire image to just a neighbourhood of the epipolar line [2]. Furthermore, if we combine epipolar information with ground truth information in a known stable environment (such as the Smart Room described in section 2), we can reduce the search area even more. For example, if the object that we are looking for is a human head, the likelihood of being between 1 and 2 meters above the floor is high. This additional height restriction on the 3D back-projected ray can be translated into an additional restriction on the epipolar line. Thus, instead of searching along the whole epipolar line on camera 2, we can restrict the search area to only a neighbourhood of a segment of the epipolar line, as it has been done in [4] in a different context. Epipolar line equations can be directly computed from calibration parameters of the implied cameras [3].

### 4.2. Best camera selection

Some of the most basic functionalities in a multicamera system are perhaps those related to the selection of the best camera for a given image analysis task. We should select the closest camera to get the highest possible size and resolution of the object under analysis, or the best oriented camera if frontality is critical (as it uses to be in face recognition tasks). If the 3D object position is known and the cameras are correctly calibrated, camera selection can be performed according to different criteria:

- *Closest camera selection*: From all available cameras that can see a given 3D point, we choose the one whose center of projection (COP) is closest to the desired 3D world position ('closeness' measured by 3D Euclidean distance). This functionality is useful to analyze small objects, for example. Furthermore, distance relationships between 3D objects and cameras are also useful to disambiguate occlusions.
- *Most centered camera selection*: Among all the available cameras that can see a given 3D point, we choose the one for which the projection of the 3D world object is closest to the center of the image (2D euclidean distance). This is helpful when significant radial distortion affects the periphery of the images.
- *Best oriented camera selection*: From all available cameras that can see a given 3D point, we choose the one which is better aligned with an oriented object. We define an oriented object by a world point and an orientation vector. The best camera is considered to be the one that can see the object and minimizes the dot product between the orientation vector and camera's projection ray director vector. This functionality deals with objects' orientation and may be useful in face identification or gesture recognition applications, where frontal views are required.

### 4.3. Sphere projection

Being able to project 3D spheres onto 2D images allows relating volumes in space with projected areas in images. While we can probably know beforehand the approximate size of the object we are looking for, the same object can be imaged as a larger or smaller spot, depending on its relative position with respect to the camera.

We may think of different applications as, for example, 'How many pixels occupies the projection of a human head, depending on its position in the room?' Or 'How large will be the uncertainty area (in pixels) of a moving object in a certain camera, if we have an upper bound of its velocity?' We can simply transform that upper bound in velocity into an upper bound in distance (radius from the last known position, in the previous frame) by means of the known time between frames. Thus, by projecting a sphere with that given center and radius, we would be able to restrict the search area to the projection, which allows us to characterize the relationship between volume and image area adapted to every camera, depending on the relative position between the object and the camera.

Given a 3D point (center of the sphere) and a radius, the equation of the projected sphere (an ellipse in general) can be computed by means of the camera calibration parameters [3]. From its equations, a number of simple geometric computations yield major semiaxis of the ellipse [5]. Thus, the ellipse can be efficiently drawn [2].

### 4.4. Oriented circle projection

While the projection of spheres outlined in previous section can give an idea of how 3D real volumes and image areas relate, it could be desirable to add a notion of orientation in some cases. For example, let us consider that we would like to know how large the projection of a human face is in a given camera, assuming that we know the 3D position of that face and its orientation. Roughly speaking, we are now obviously interested in knowing the area of a "face" (oriented circle) in the image, and not the area of a "head" (sphere).

In order to completely describe an oriented circle in 3D space, the following 6 parameters have been used:

- 3 parameters for the 3D coordinates of the center.
- 3 parameters for the orientation vector, perpendicular to the plane containing the circle. The modulus of this vector is equal to the radius of the circle.

Given those six parameters, the projection of the circle can be efficiently drawn on the camera projection by means of the camera calibration parameters [2].

## 5. EXPERIMENTAL RESULTS

The functionalities described in this paper have been implemented and tested in the Smart Room scenario mentioned in section 2. Some of these tests will be briefly presented here, and they intend to prove the potential usefulness of these functionalities when combined with other detection or tracking algorithms.

### 5.1. Epipolar lines

A test with Ground Truth Points (GTP) has been performed to evaluate the functionality of the epipolar lines. 10 Ground Truth Points (16 x 16-cm. squares with an inscribed cross) have been placed at different positions in the room, with different distortions, heights, poses, and distances to the different cameras of the room.

From their imaged coordinates, the epipolar lines corresponding to the GTP's visible in camera 1 have been computed and represented in the other cameras. The resulting distances from the imaged GTP's to their respective epipolar lines has not exceeded 8 pixels in any camera, and their mean value was about 3 pixels. We must indicate here that, for maximum coverage of the room, the five cameras are equipped with wide angle lenses (70º–150º), yielding significant distortion in the periphery of the images.

### 5.2. Best camera selection

Although all the camera selection criteria described have been implemented and tested, we present here a specific test for the best oriented camera functionality, as we think that this is the most useful for different image analysis algorithms.

A video sequence of a moving bar has been synchronously recorded from all the available cameras. This bar is equivalent to an orientation in the room, which can be known in each frame by computing the 3D coordinates of its extreme points. The 3D coordinates have been found by manually marking their image projections in the different cameras. As the bar moved in the room, it pointed at different cameras, so that the best oriented camera could be easily recognized from the images. The best camera (according to visual estimations from the recorded images) was correctly selected along the whole video sequence. In order to fully appreciate the correctness of the best camera selection functionality, a VRML 3D animation has been constructed(*). The room and the table are represented, and a cylinder is fitted to the computed 3D coordinates of the bar extreme points. The positions of the cameras have also been computed from the extrinsic parameters. During the animation, the

(*) *The authors would like to thank Jordi Salvador for VRML display.*

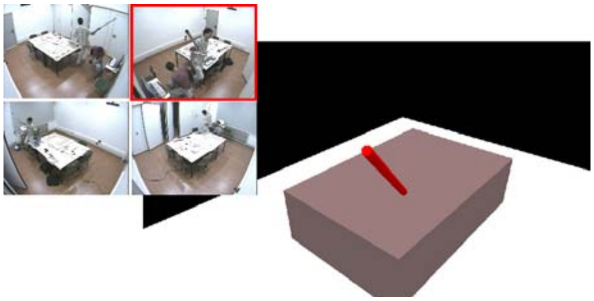

*Figure 1. Selection of best oriented camera with an oriented bar.*

*Figure 2. Sphere projection over GTPs*

best oriented camera (according to cylinder orientation) is automatically selected. For a sample frame, see Figure 1.

### 5.3. Sphere projection

Two different series of tests have been carried out: sphere projection over GTP's and in the oriented bar test images, which are described below.

*5.3.1. Sphere projection over GTP's*

A sphere of radius $8\sqrt{2}$ cm centered in the computed 3D coordinates of the GTP has been projected onto the five cameras. The length of its radius is important, because GTP's are 16x16cm squares, so they should be inscribed in the projection of the sphere. The resulting images (Figure 2) prove the validity of projected spheres' sizes.

*5.3.2. Sphere projection in the Oriented Bar Test*

This test consisted in projecting a sphere of radius 20 cm (shown in red in Figure 3) centered at the top extreme point of the bar, in cameras 1 to 4.

### 5.4. Oriented circle projection

As with the sphere projection functionality, two different series of tests have been carried out for oriented circles:

*5.4.1. Oriented circle projection over GTP's*

A circle of radius $8\sqrt{2}$ cm centered in the computed 3D coordinates of every GTP, has been projected onto the five cameras. Two tests of this kind were performed: one with (0,0,1) as orientation vector –pointing to the ceiling–, and (1,0,0) –pointing to one of the long walls–. As in the sphere projection test, the resulting images (available from [2]) show a good relationship between the sizes of the GTP's projections and their associated circles, as well as with their orientation.

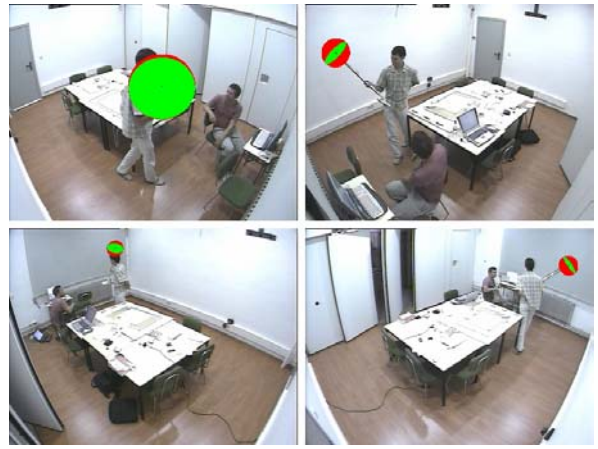

*Figure 3. Sphere and oriented circle projection in the oriented bar test.*

*5.4.2. Oriented circle projection in the Oriented Bar Test*

This test consists in projecting a circle of radius 20 cm (shown in green in Figure 3) centered at the top extreme point of the bar in cameras 1 to 4. The orientation of the circle plane is perpendicular to the bar. Significant coherence of distances to the cameras and the orientation of the bar with the projected areas were achieved.

## 6. CONCLUSIONS

We have presented in this paper four different functionalities for mapping 2D images and 3D world objects in a multicamera system, targeted at easing video analysis tasts. A set of tests have been implemented and tested for all the functionalities. Some of the tests carried out have been briefly presented here in order to prove the potential interest of these functionalities when assisting image detection and tracking algorithms.

## 7. REFERENCES


[1] Zhengyou Zhang. "A flexible new technique for camera calibration". Technical report, Microsoft Research, August 2002.

[2] Òscar Garcia. "Mapping 2D images and 3D objects in a multicamera system". Master's Thesis, UPC, October 2004. (http://gps-tsc.upc.es/imatge/_JosepRamon/PFC/reports)

[3] Richard Hartley and Andrew Zisserman. *Multiple view geometry in computer vision*. Cambridge Univ. Press, 2000.

[4] K. Shafique, Mubarak Shah, *A Noniterative Greedy Algorithm for Multiframe Point Correspondence*. IEEE Trans on PAMI, Vol.27, no 1, January 2005, pp. 51-65.

[5] Sebastià Xambó. *Geometria*. Edicions UPC, 1997.