Computer Vision for determination of Fridge Contents *

Fredrik Färnström, Björn Johansson and Kalle Åström
Centre for Mathematical Sciences
Lund University

Abstract

The topic of this paper is a computer vision system for a refrigerator. A number of cameras are placed in the fridge. The system then extracts and processes information about the contents of the fridge. As a user manipulates items in the fridge, a motion-detecting system selects key-frames from the video streams. Images before and after actions are kept and analyzed. A camera mounted below the shelf is used to obtain an image of the footprints of the objects on the shelf. By analyzing difference images of the footprint, it is possible to determine the type of action that has been performed (insertion, movement, removal). Furthermore, the side view and the footprint are used to estimate the 3D shape of the object within a specific class of object shapes. As long as the true shape lies within or is sufficiently close to this class, this fast and novel method is successful.

1. Introduction

The topic for this paper is a fully automated computer vision system used in a refrigerator. The purpose of the system is to extract information about the contents in the fridge and communicate this to a remote user. The work is one part of a project in collaboration with Electrolux, a large manufacturer of refrigerators. For further details and references see [1, 2, 3].

The objective of this work is to detect events, to segment the scene and to build a 3D model of the contents of the fridge for visualization, as well as extract 3D information and texture to be used for identification. The identification problem will not be considered here.

One central idea in this paper is that the footprint, i.e. the bottom surface of objects in the fridge, contains very valuable information about its contents. In order to obtain these footprints, it will be assumed that the shelves of the fridge are semi-transparent, in this way the objects are visible for a camera located below a shelf. All objects in the fridge have to be placed directly on the shelf, and may not be placed on top of other objects. In addition they have to be inserted so that the entire object is visible for the camera. They may then be moved behind other objects.

This vision based system does not require modifications of the objects which are to be placed in the fridge, for example by the attachment of tags. Neither does the user have to change the behavior when manipulating the contents of the fridge, this is the case for systems using barcodes.

2. System Approach

This section describes the proposed system. The focus of the work was to build a complete automatic system. Due to time restrictions simple solutions have been chosen to many of the image processing problems we came across.

2..1 Overview

A number of cameras are placed at a fixed position in the fridge, so they can take pictures of the objects of the fridge. There is one image taken from below and one taken from above each shelf. The shelf is prepared with a plastic film so that the footprint of the objects clearly stands out in the image taken from below. The objects in the fridge are assumed to be standing on the shelf, not on top of each other.

We assume to have a model of the current contents of the fridge. The system then detects a change in the fridge, e.g. insertion, and the model is then updated.

The first stage of this procedure is the *motion detector*. The motion detector compares images from the same camera taken at different time-instants, to detect when there is any motion in front of the cameras. One or more cameras can be used for this. The processing at this stage is simple but requires a video flow from the cameras, making it an ideal candidate for hardware implementation. A few images are then forwarded to the next stages of processing.

Useful information can be achieved from an image of the *footprint* of the objects standing on a shelf in the fridge. From these, it is possible to determine what the user has done with the contents of the fridge, for example

*This work has been supported by, Electrolux, and the Swedish Research Council, project 221-2000-476.

if an object has been inserted or moved. Depending on the type of action, the system moves, removes, or adds an object to the current model.

When a new object is inserted the information from the footprints works as an aid for determining the three-dimensional shape of the inserted object. An allowed 3D shape is adapted to the footprint and a contour image of the object. The footprint also provides the position of the object.

This three-dimensional information and the footprint information may together with the texture of the object be supplied as input to an identification sub-system. This part of the problem will not be considered here.

Finally the contents of the fridge is presented to the user. This may be done with two-dimensional images, three-dimensional models which may be viewed from different directions, or with a simple list of objects. This latter form is useful since it may easily be transferred to wireless mobile devices, e.g. cell phones.

A schematic overview of the system can be seen in Figure 1. In the following sections the different parts of the system will be described in detail.

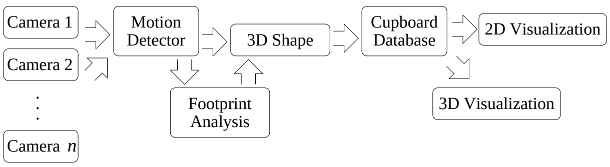

Figure 1: Overview of the system solution. Images are captured by the cameras. These are then processed and 3D information is extracted. A database keeps track of objects currently in the fridge. The contents of the database can be visualized by the user in different ways.

2..2 Motion detection

The first part of the system is the *motion detector*. The purpose of the motion detector is to detect when changes occur in the fridge. When this happens one image is captured by each of the cameras. The images are used for further processing.

Let $A(x,y) = [A_r(x,y)\ A_g(x,y)\ A_b(x,y)]$ and $B(x,y) = [B_r(x,y)\ B_g(x,y)\ B_b(x,y)]$ be two color images captured with a camera. Define $E(A,B)$ to be a measure of the total difference between images A and B, as

$$E(A,B) = \sum_{x,y} |A(x,y) - B(x,y)|^2. \qquad (1)$$

Although this is only one of many potential measures of difference between images, it is simple and good enough for our purpose. We say that there is motion in front of the cameras when the difference defined by (1) of two following images is larger than a threshold. An action is defined to have occurred when there is a sequence of no motion, motion and then no motion from any of the cameras.

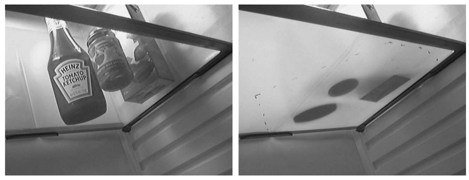

Figure 2: The left image shows an image of some objects from below the transparent glass shelf. The right image shows the same objects on a shelf covered with a plastic fi lm.

2..3 Footprint View

The footprint view image is an image showing the bottom surface of every object placed on a shelf, see Figure 2. It is rectified to correct metric by a homography specific for each camera and shelf pair. This homography is estimated in advance and only once.It is however possible to adaptively reestimate camera calibration on-line. By processing these images it is possible to achieve important information about the objects, e.g. for detecting and classifying actions. They are used to determine the position and for certain objects the orientation, and they are used in the reconstruction of the 3D shape as well. In addition the footprint may be used in the identification process.

In order to decide what type of action occurred we analyze the footprint images. In the current implementation we define four different types of basic actions:

- *Insertion*. A new object has been put onto the shelf.
- *Removal*. An object has been removed from the shelf.
- *Movement*. An object on the shelf has been moved (translated and/or rotated).
- *No action*. Nothing has changed, the two images show the same objects at the same positions.

By thresholding the difference between footprint views before and after an action with two different thresholds we get an image showing inserted, removed and indifferent regions, cf. Figure 3. We make the classification by comparing the area of these. Denote by A_I and A_R the area of the inserted and removed regions respectively. The classification is done as follows:

- If $A_R < A_0$ and $A_I < A_0$ then no action has been performed,

- else if $A_I \geq A_0$ and $\frac{A_I}{A_R} > d$ then an object has been inserted,
- else if $A_R \geq A_0$ and $\frac{A_R}{A_I} > d$ then an object has been removed.
- otherwise, an object has been moved.

Here, A_0 is a threshold. If an object has been moved, then $A_I \approx A_R$ and $\frac{A_I}{A_R} \approx 1$. Therefore, d works as a threshold for classifying a movement from an insertion or a removal.

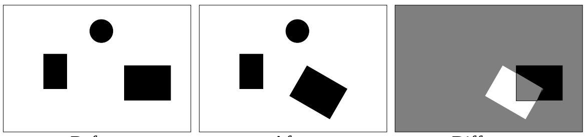

Figure 3: The figure illustrates the footprint image before (left) and after (middle) a movement. The resulting difference image is shown to the right.

This procedure fails if objects are put on top of each other and the current implementation only handles actions on single objects.

2..4 Handling actions

Depending on the type of detected action, the response of the system varies as discussed below.

Removal

If the action is decided to be a removal, the object corresponding to the largest removed region in the difference of the footprint images is deleted from the model.

Movement

In the case of a movement, the object in the model corresponding to the largest removed region is moved to the center of mass of the largest inserted region. The orientation of the object is also updated by determining the central axis of the footprint. If there is an overlap between the regions, cf Figure 3, we have to use the original footprint images as well, in order to decide which object has been moved and where to. A disadvantage analyzing only the movement in the footprint views is that the new orientation of e.g. a cylinder cannot be determined. There is also an ambiguity of 180 degrees in rectangular objects.

Insertion

We determine the position of the inserted object by analyzing the footprint images in the same manner as above. In order to determine the 3D shape of the object we also use the difference between a side view of the object before and after the action. This image is used to estimate the contour of the inserted object, cf Figure 4. In order to estimate the contour, the inserted object must not be occluded by any other object. Knowing the projection matrices for the cameras, we adopt a shape from a predefined class of objects to the footprint- and contour image.

Figure 4: The two images to the left show the shelf before and after insertion of the object. The third image from the left shows the difference image. The rightmost image is a thresholded image.

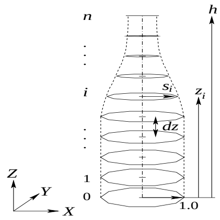

Figure 5: The figure shows how a bottle can be represented by the heights z_i, in steps of dz, and corresponding scale factors s_i relative to the footprint polygon.

The class of allowed objects is objects with uniform cross sections. This class includes many of the most common objects in a fridge, e.g. cylinders with central axis orthogonal to the shelf, boxes, cones with the top on the central axis, bottles, etc. Spherical objects (apple) are approximated reasonably within the class, however e.g. cylinders lying down, does not fit at all.

The parameters that have to be determined are the scales of the footprint for different heights above the shelf, c.f. Figure 5. They are estimated by trying different scales at every height to find the largest scale for which the scaled footprint is projected inside the contour. When the scales are determined for every height they are filtered to achieve a smoother appearance.

2..5 Visualization

In order to visualize the content in the fridge the 3D model was implemented as a texture-mapped polygon in OpenGL. The front textures were used to texture both the front and the back of the models. This procedure gives a complete textured model that can be viewed from any angle with a reasonable result.

3. Examples

Below are some frames from a movie showing a user putting in objects in a fridge. A few images of the resulting 3D reconstruction are given above.

4. Conclusion

We have presented an automatically vision system for determining and visualize the contents in a fridge. The system detects a change in the fridge and update the current 3D model by moving, removing or inserting objects.

The focus of the work sofar has been to build a complete system, for the future it would be desirable to implement more sophisticated algorithms for some of the basic image processing problems. Another interesting continuation is to mount the cameras in the door and use the image sequence produced when opening and closing the door. Having an approximative 3D model and knowing the motion up to one parameter, there are great possibilities to improve the model.

The refrigerator has to have prepared shelves and there are some restrictions on the user, e.g. objects must be inserted fully visible and standing on the shelf. Is spite of this the system is very flexible compared to non-vision solutions based on e.g. tags or bar-codes.

References

[1] F. Färnström. Computer vision for fridge content determination. Master's thesis, Centre for Mathematical Sciences, Lund Institute of Technology, Sweden, 2001.

[2] D. Sarmiala. Kameraplacering i intelligenta kylskåp. Master's thesis, Dept. of Geodesy and Photogrammetry, Royal Institute of Technology, Sweden, 2001. To be published.

[3] J. Sundin. Identification of food packages through image analysis. Master's thesis, Centre for Image Analysis, Uppsala University, Sweden, 2001. To be published.