Expanding Irregular Graph Pyramid
for an Approaching Object*

Luis A. Mateos, Dan Shao, and Walter G. Kropatsch

Vienna University of Technology,
Pattern Recognition and Image Processing Group,
Favoritenstr. 9/1832, A-1040 Vienna, AUSTRIA
{lam,shao,krw}@prip.tuwien.ac.at
http://www.prip.tuwien.ac.at/

Abstract. This paper focus on one of the major problems in model-based object tracking, the problem of how to dynamically update the model to adapt changes in the structure and appearance of the target object. We adopt Irregular Graph Pyramids to hierarchically represent the topological structure of a rigid moving object with multiresolution, making it possible to add new details observed from an approaching object by expanding the pyramid.

Key words: irregular graph pyramid; adaptive representation; object structure; tracking model;

1   Introduction

One of the major problems for model-based object tracking is: how to dynamically update the model to accommodate the object appearance and structure changes due to the changes in the surrounding conditions against which the tracked object is observed [1].

By reviewing the previous works for model updating, a method for adjusting features while tracking is presented by Collins and Liu [2]. Their hypothesis is that best discriminating features between object and background are also best for tracking the object. This method uses the immediate previous frame as the training frame and the current as the test frame for the foreground and background classification. In [3] the template is first updated with the image at the current template location. To eliminate drift, this updated template is then aligned with the first template to give the final update.

However, these methods update their models by getting new information of the target object from resampled input image frames, with a fixed resolution. The properties of the Irregular Graph Pyramid and its applications in image processing motivated us to use the advantages of its topological/structural hierarchy features in tracking. With its hierarchical feature, we are able to represent

---

the moving object in multi resolution, and incrementally update it by adding new details from the approaching target object.

Each level of the pyramid is represented by a graph which embeds the topological structure of the object at certain resolution. Such graphs are built of vertices and spatial edges. In the vertex; attributes like size, color and position of the corresponding pixels (region) can be stored. The spatial edges are used to specify the spatial relationships (adjacency, border) between the vertex (regions). Tracking methods using structural information often employ graphs to describe the object structure. Locally, features describe the object details; globally, the relations between features encode the object structure.

By exploiting the spatial and temporal structure of the scene [4], Artner and Lopez improve the performance of object tracking in video sequences. They present a structural tracking approach which is robust to different types of occlusions, as the structural information is able to differ objects by their different structural types.

Less work has been done for tracking with graph pyramid. [5] presents a method of tracking objects through occlusions using graph pyramid. They apply graph matching to find the matching between the vertices in current image frame with the ones in previous image frames. Instead of doing the graph matching which is known as computationally expensive, we propose a top - down pyramid reconstruction approach to avoid graph matching.

Major contributions of this paper are 1) Our model hierarchically represents a moving object with multi resolution; 2) For an approaching object, we encode the new details by expanding the pyramid structure; 3) Compared to other pyramid tracking methods, we use a top - down pyramid reconstruction approach instead of bottom-up recomputing the graph pyramid for each frame. In such way, computational costs would be reduced.

Organization of paper In section 2 we recall the concept of Irregular Graph Pyramid (IGP). In section 3 we describe the process of tracking with IGP. Section 4 describes a concept of Adatpive Zoom in for tracking the approaching object. Section 5 finalize with conclusion and open questions.

2   Recall of Irregular Graph Pyramid

An irregular pyramid combines graph structures with hierarchies. Each level is a graph describing the image with various resolutions by contracting the graph from the level below. Specifically, contraction is a process to take the attributes of all children as input and then compute the parent's attribute as output, removes the edges from input graph while simultaneously merging together the vertices it used to connect [6].

For this paper we are considering combinatorial maps. A combinatorial map is a topological model which allows to represent subdivided objects as planar graphs. A 2D combinatorial map is defined by a triplet $M = (D, \sigma, \alpha)$ where D is a finite set of darts, σ is a permutation on D and α is an involution on D

without fixed point [7]. For each dart, σ gives the next dart by turning around the vertex v in the positive orientation (clockwise); For each dart, α gives the other dart of the same edge e. There are always two darts corresponding to a same edge, α allows to retrieve edges e, and σ allows to retrieve the vertex v.

Taking a simplified image of a cup as example, we build the base graph as the input image, where each vertex represent a pixel in the input image. Then use the contraction methods to build the irregular pyramid. Such approach would lead to a pyramidal structures like the following figure:
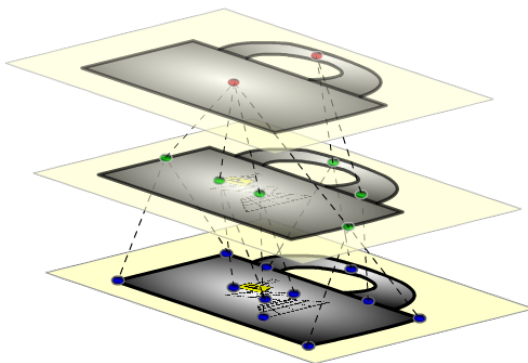

Fig. 1. Irregular Graph Pyramid.

Level 0: The base level of the pyramid consists in a geometry description of the underlying image (here a simplified image of a cup).

Level 1: The second level of the pyramid, simpler boundaries are abstracted from base level (like the handle and the logo of the cup).

Level 2: Adjacent parts of the cup are grouped in order to represent compound abstract objects.

In irregular pyramids, during the building process of pyramid, the adaptive contraction of the structure preserves its topological structure.

3 Tracking with Irregular Graph Pyramid

In this paper we only consider tracking linear object movement. Translation, scaling and rotation are the basic types of linear object movement. In this section, we describe how to track the translation movement using graph pyramid. In the following section, we focus on tracking the scaling movement (approaching object). Rotation is not covered by this paper. However, in our other works, we present the concept of topological completion to reconstruct the closed surface of a 3D object for tracking a rotating 3D object.

For initialization, a pyramid is bottom-up built for the target object from a video frame. The target object is represented at different levels of resolution using

a graph for each level. We build the base graph as the input image, where each vertex represent a pixel. We save the coordinates of pixels in the attributes of the vertices in base graph. And discriminability can be computed locally similar to Lowe's SIFT descriptor [8].

In the contraction process, the vertices with most discriminative feature survive. The coordinates of the surviving vertices are preserved in the attributes of the parents. For the non surviving vertices, coordinates are saved in the contracted edges for the purpose of later pyramid reconstruction.

In such way, the pyramid apex encodes the most discriminative feature of the target object as well the coordinate of the most discriminative feature. For instance of a cup with yellow logo on the surface, the apex encodes the yellow logo and the coordinates of the logo.

For a rigid object, the structure of the pyramid is invariant to any geometric transformation. By estimating the location of the logo (which is abstracted as the apex), we are able to locate the whole pyramid. All the vertices can be accessed efficiently from the apex by following the parent - children path. The construction of the pyramid is a bottom - up process while the reconstruction from the apex is a top - down process. In such way we can reconstruct the whole pyramid by only locating the apex point.
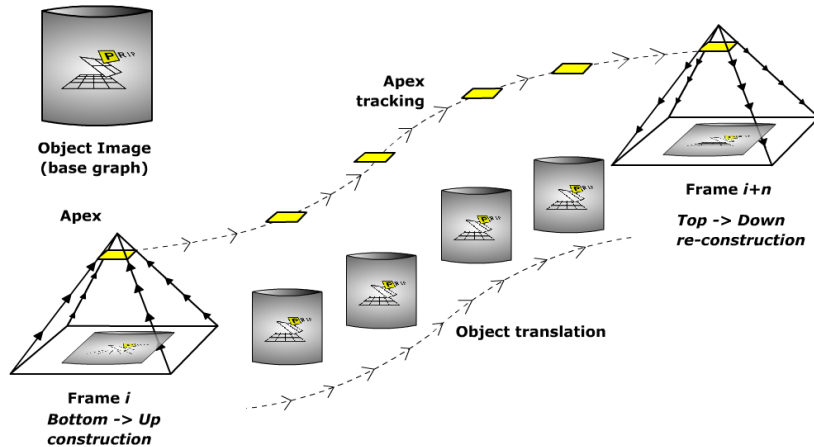

Fig. 2. Tracking with Irregular Graph Pyramid.

An example of the tracking process is shown Fig. 2. At frame i the object is detected and its pyramid is initially built. The apex encodes the most discriminative feature of the object (the yellow logo). In the next frames $i + n$ the object moves, we detect the location of the yellow logo by motion estimation method. Once the apex coordinate (coordinate of yellow logo) is known, we can reconstruct the pyramid until we retrieve all the vertices and their coordinates

of base graph. As the base graph of pyramid present the image of the target object, the target object at frame $i + n$ is tracked.

4   Adaptive Zoom - In

Model-based object tracking intend to keep a clear and detailed description of the object by gathering its latest descriptors available, these descriptors describe elementary characteristics of the object such as the color, the shape, the texture, among others. In order to do this, the model must be able to adapt the object changes. If an object changes its details for a given reason then the model must have such a robust adaptive system that keeps a clear identification and location of this tracking object.

Every frame is different, so new relevant pieces of information or features from the object may appear due to different reasons such as appearance or pose changes. We are treating the current frame as the latest source of information, containing the ultimate object descriptors. This new information may include changes in its internal or structural descriptors. Traditional model based representations are reliable and robust in scenes where the object is consistent, no expansion of details, no considerable illumination changes. In this section we present a case and the proposed solution using an adaptive zoom - in; if the object gets closer to the camera, the distance camera-object changes and we will obtain a bigger picture (higher resolution) of the object and more detailed descriptors will appear.

From the irregular graph pyramid perspective, this new image can be seen as a proportional projection of the original base graph which includes more detailed descriptors. The pyramid will expand one level below the current base graph, this new base graph encodes both structures due to a higher resolution of the object. And the problem is to find the vertical connection between the new base graph and the existing upper part of the pyramid.
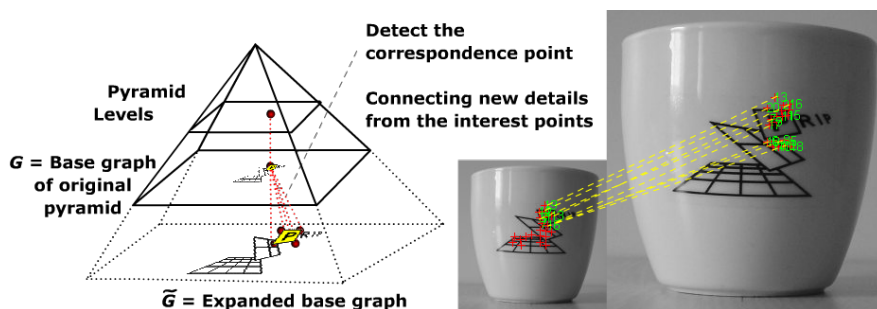

Fig. 3. Pyramid base projection to a lower level.

Let $G = (V, E)$ be the graph at the base of the original pyramid and $\tilde{G} = \left(\tilde{V}, \tilde{E}\right)$ the new base graph.

Other than in most parameter optimization tasks, both the vertices of the lower level $\tilde{G}$ and the next upper level G are know as well as the reduction function. The only unknown is the contraction kernel K such that $G = \tilde{G}/K$.

$\tilde{G} = \left(\tilde{V}, \tilde{E}\right)$ corresponds to the image grid with higher resolution, which is considered as a 4-connected graph. For each vertex $v \in \tilde{V}$, a set of allowed candidate parents are computed by finding the closer correspondences. Then, a new parent is chosen from the set of allowed candidate parents. The vertex is relinked to the new parent and the graph structure and attributes of vertices are updated accordingly. This procedure is repeated until a stable configuration is reached. The new parent p_{new} is chosen for each vertex v such that the difference between $g(\tilde{v})$ and $g(p_{new})$ is minimized. The rules which determine the selection of a new parent for a vertex can be formulated as an energy minimization problem relinking [9]. An energy

$$E = \sum_{v \in \tilde{V}} n(v) \left[g(v) - g(p(v))\right] \tag{1}$$

can be defined. Here, $n(v)$ denotes the area of the receptive field of v, $g(v)$ denotes its gray level or local feature value and $p(v)$ denotes the parent assignment. As pointed out by Nacken [9], this relinking algorithm way destroy the connectivity of the receptive fields. Improve our method with Nacken's modification is one of our future directions.

By graph relinking, we put higher resolution image frame as new base graph, then define contraction kernel that produce the old base. Instead of bottom-up rebuilding pyramid by applying contraction kernel, we use the graph at the base of the original pyramid as fixed upper level graph, and try to define the expansion kernel. In such way, the original pyramid structure remains unchanged, we only attach a lower level base graph into this pyramid.

Assumption. Consider the speed of the moving object, we assume the approaching speed is not too fast. The current size of the target object can not exceed twice as the one in previous image frame, which means the maximum scaling factor can not exceed 2. Otherwise there might be a gap between the new base graph and the old base graph so that we have to insert extra levels to bridge the new base graph with the old base graph.

5   Conclusions

This paper presents a novel concept of using irregular graph pyramid to represent a moving object in multi resolution. This concept can be applied into many tracking applications. In this paper, we focus on how to track approaching object (scaling movement) by expanding the irregular graph pyramid. Considering the

computational costs, we propose a top-down pyramid reconstruction approach, instead of bottom - up rebuilding pyramid for the target object in each image frame. Our future work will concentrate on how to extend this tracking framework to track other object movement such as reflection, shear and non linear movement.

References

1. Feng Tang, Hai Tao: Probabilistic Object Tracking With Dynamic Attributed Relational Feature Graph. IEEE Trans. Circuits Syst. Video Techn. 18(8): 1064-1074 (2008)
2. Robert Collins, Yanxi Liu, and Marius Leordeanu, "On-Line Selection of Discriminative Tracking Features," IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol. 27, No. 10, October, 2005, pp. 1631 - 1643.
3. I. Matthews, T. Ishikawa, and S. Baker "The template Update Problem" IEEE Transaction Pattern Analysis and Machine Intelligence, Vol. 26, No. 6, June, 2004, pp. 810 - 815.
4. S. B. Lopez Marmol, N. M. Artner, M. Iglesias, W. Kropatsch, M. Clabian, and W. Burger. Improving Tracking Using Structure. In proceedings of Computer Vision Winter Workshop (CVWW 2008). Pages 69 - 76. 4. - 6. February 2008. Moravske Toplice, Slovenia.
5. Donatello Conte, Pasquale Foggia, Jean-Michel Jolion, Mario Vento: A graph-based, multi-resolution algorithm for tracking objects in presence of occlusions. Pattern Recognition 39(4): 562-572 (2006).
6. W.G. Kropatsch. Building irregular pyramids by dual-graph contraction. IEE Proceedings- Vision Image and Signal Processing, 142(6):366 - 374, 1995.
7. Brun, L. and Kropatsch, W.G. (1999a). Dual Contraction of Combinatorial Maps. Technical Report PRIP-TR-54 Institute f. Computer Aided Automation 183/2, Pattern Recognition and Image Processing Group, TU Wien, Austria.
8. David G. Lowe, "Object recognition from local scale-invariant features," International Conference on Computer Vision, Corfu, Greece (September 1999), pp. 1150-1157.
9. Peter F.M. Nacken. Image segmentation by connectivity preserving relinking in hierarchical graph structures. Pattern Recognition, 28(6):907-920, June 1995.