Video Pop-up: Monocular 3D Reconstruction of Dynamic Scenes

Chris Russell Rui Yu Lourdes Agapito

University College London

OVERVIEW

In this paper we propose an unsupervised approach to the challenging problem of simultaneously segmenting the scene into its constituent objects and reconstructing a 3D model of the scene. The strength of our approach comes from the ability to deal with real-world dynamic scenes and to handle seamlessly different types of motion: rigid, articulated and non-rigid. We formulate the problem as hierarchical graph-cut based segmentation where we decompose the whole scene into background and foreground objects and model the motion of non-rigid or articulated objects as a set of overlapping rigid parts.

MOTIVATION

Multibody sfm and non-rigid structure from motion NRSfM have seen sustained progress in dealing with dynamic scenes or creating vivid life-like reconstructions of deformable objects. However, they remain far behind their rigid counterparts. Multibody sfm approaches can segment the scene into multiple rigidly moving objects, however they cannot deal simultaneously with the presence of deformable or articulated objects in the scene.

Our aim is to offer a solution to the problem of scene reconstruction for real-world dynamic monocular videos that deals seamlessly with the presence of non-rigid, articulated or pure rigid motion.

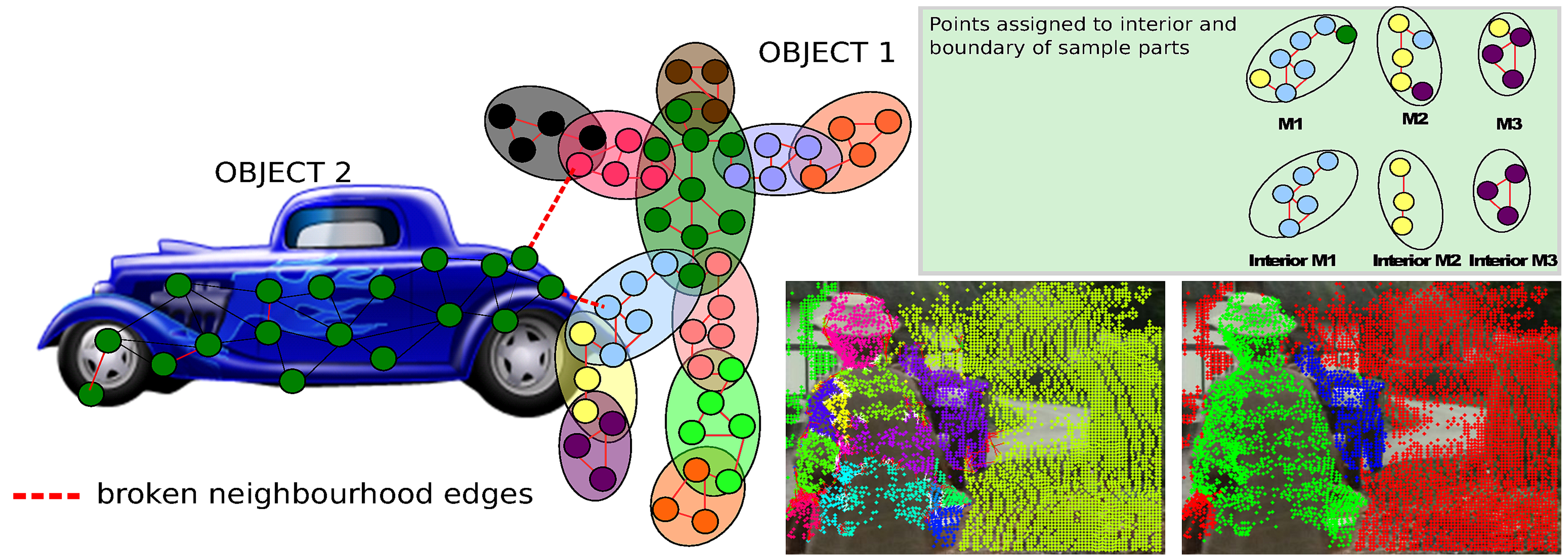
CONTRIBUTIONS

1. Our approach is able to adapt the topology of the neighbourhood graph by breaking edges where necessary to preserve boundaries between objects. In this way our approach can deal with an entire scene where objects might occlude one another and not just pre-segmented objects.
2. Our work results in a hierarchical approach to dynamic scene analysis. At the higher level of the hierarchy the scene is explained as a set of objects that are detached from the background and from each other. At the lower level of the hierarchy, each object can be explained as a set of overlapping parts that can model more complex motion.

REFERENCES

1. C. Russell, J. Fayad, and L. Agapito. Energy based multiple model fitting for non-rigid structure from motion. In CVPR, 2011.

SCENE RECONSTRUCTION WITH AN ADAPTIVE NEIGHBOURHOOD



ENERGY FORMULATION

$$C(x) = E_{data} + E_{edge\_break} + E_{sparse} + E_{mdl}$$
$$= \sum_{i\in\Gamma}\sum_{m\in x_i} U_i(m) + \sum_{i\in\Gamma}\sum_{j\in N_i} d_{i,j}\Delta(j\notin N'_i)$$
$$+ \sum_{m\neq n\in M}\Delta(\exists i: I_i = m, n\in x_i) + \mathrm{MDL}(x)$$

UNARY COSTS

$$U_i(m) = G_i(m) + P_i(m)$$
$$G_i(m) = \sum_{f<F} r^{-1}(u_i^{f+1^T} F_m^{f,f+1} u_i^f)^2$$
$$P_i(m) = \lambda_s \sum_{f\le F}(S_{I_f}(i) - \overline{S}_m)^2$$

The unary cost is the sum of two costs, i.e. rigidity term $G_i(m)$ and saliency term $P_i(m)$. The first term evaluates the cost of assigning tracks to models as the deviation from epipolar geometry across all pairs of consecutive frames. The second term computes a saliency score for each pixel in every frame and encourages tracks with similar saliency scores, to belong to the same model.

TOPOLOGICALLY ADAPTIVE NEIGHBORHOOD

$$\sum_{i\in T}\sum_{m:\exists j\in N_i \cap m=I_j} \min\Big(\sum_{j:I_j=m} d_{i,j}, U_i(m)\Big)$$

To separate connected objects from one another, we discard edges from the neighborhood with a per-edge cost.

OVERLAP SPARSITY TERM E_{sparse}

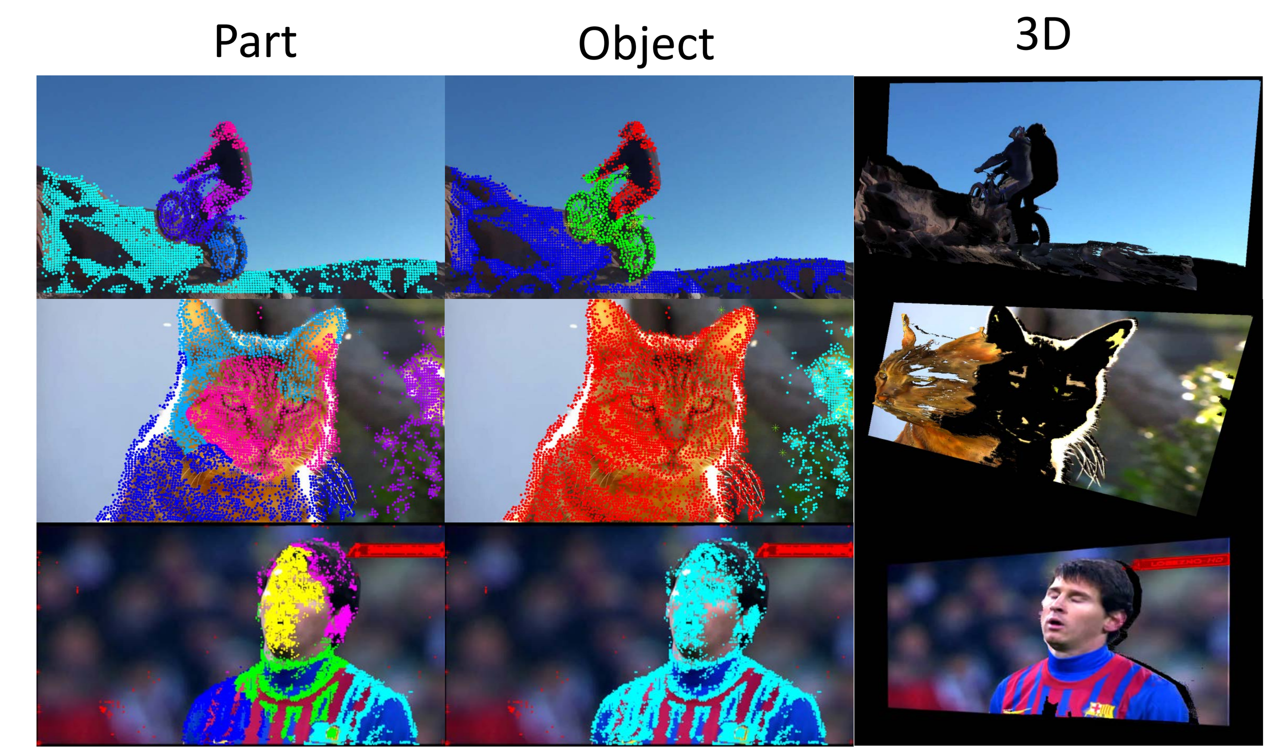
$$\sum_{m\neq n\in M}\Delta(\exists i: I_i = m, n\in x_i))$$

We introduce a novel sparsity prior that penalizes the total number of models that overlap and encourage regions with limit overlap to disconnect.

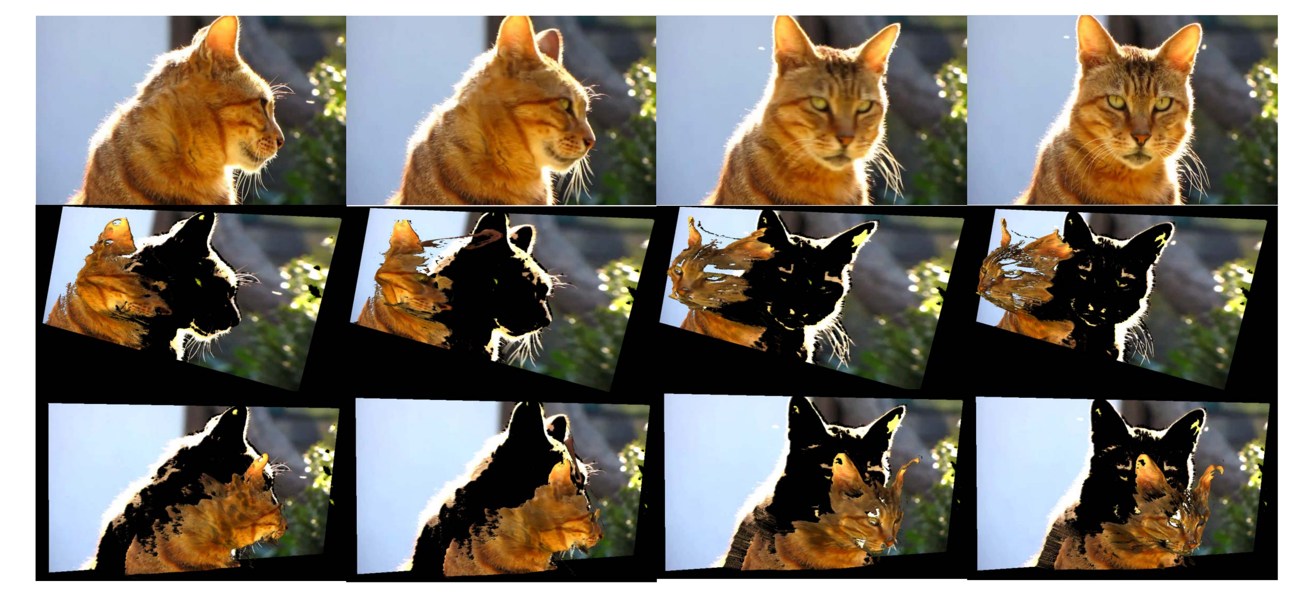
EFFICIENT OPTIMIZATION

1. Initialize with an excess of models by random sampling of feature tracks.
2. Optimize the energy using a hill-climbing approach alternating between fixing model para-mters and optimising the labelling of assigning tracks to a set of parts and fixing the labelling and optimize $\mathbf{F}_m$ for all models.
3. The overlapping cost and mdl cost is optimized using the same approach as [1], while a new alpha expansion is introduced to optimize edge breaking and sparsity cost.

EXPERIMENT RESULTS



Cat sequence



Motorbike sequence

Messi sequence(sparse results)