Image Interpretation Confusion Resolution by Collaboration

Ben Mathew, Akram Alghanmi, Marius Silaghi
Florida Institute of Technology
bmatthew2014@my.fit.edu, aalghanmi2019@my.fit.edu, msilaghi@fit.edu

Abstract

Visual scene understanding can benefit from inputs provided by multiple participants with their different perspectives, and a distributed version of a modified Waltz filtering enriched with modern AI inferences can potentially help accuracy and speed trade-offs by exploiting the simultaneous perspectives and logic. Speed is improved by the contribution of the implicit parallelisation in processing. Accuracy improvements are expected from updating constraints with novel and more powerful inferences that the participants can apply. Automatically understanding scenes is a highly relevant problem. Multiple robots communicate with one another to classify shapes of edges of an object. Local reasoning can reduce communication latency.

Introduction

The Distributed Enriched Waltz Filtering technique draws inspiration from the classic Waltz algorithm, which was initially proposed for solving the graph vision problem (Waltz 1975; Bessiere et al. 2005; Balafrej et al. 2014).

In scene understanding, the goal is to improve the accuracy of interpretation of objects or segments within an image or video frame. Traditional filtering techniques often focus on local data and relationships, but Distributed Enriched Waltz Filtering takes a more collaborative and global approach.

Work on line drawings interpretation by propagation of constraints was proposed in (Huffman 1971). The problem that was solved using the above technique is as follows: given a 2-D line drawing representing a collection of polyhedral blocks on a table, Guzman labeling tries to determine which faces (bounded regions in the drawing) go together as parts of the same object (Guzman 1968). An early theoretically and experimentally analysis of chaotic collaboration was contributed in (Arbab and Monfroy 2000) with respect to general constraints.

Problem Formulation

Our problem can be formalized as follows. We have a set of n participant agents (robots), $A = \{A_1, ..., A_n\}$. Each of them is surveying the scene for a set of m salient items of interest denoted:

The items (vertices in Waltz graphs) are associated with features that make the pairing between items from various agents probabilistic with the function $P_{ij}(x)$ specifying whether for agent i the item x actually corresponds to the item I_j.

Robots observe subsets of items (patterns of salient features corresponding to Waltz graph vertices) and maintain an interpretation of the scene as a set of edges between some of the items: $E \subseteq \{(i, j, \tau) | i, j \in I, \tau \in \Theta\}$ (Cieslewski, Choudhary, and Scaramuzza 2018; Mallya and Lazebnik 2015).

Each edge between two items i and j is labeled with a tag τ, from a set Θ, specifying one of the Waltz relations. The relations are $W = \{+, -, \rightarrow, \leftarrow\}$, where $+$ denotes a concave edge with planes being closer to the viewer than the edge itself, $-$ denotes a convex edge with planes being farther to the viewer than the edge itself, $\rightarrow$ with the lower plane being hidden behind the higher plane, and $\leftarrow$ with the higher plane being hidden behind the lower plane.

The agents exchange messages describing changes in their perception of the items and edges, changes that are due to additional inspection or modification of perspective. A message has the format $M = \langle \mathcal{I}_\mathcal{T}, \mathcal{E}_\mathcal{T} \rangle$, where $\mathcal{I}_\mathcal{T}$ is a set of items together with their observed instance features, while $\mathcal{E}_\mathcal{T}$ is a set of edges between items in $\mathcal{I}_\mathcal{T}$, as perceived at moment $\mathcal{T}$ (Khan and Al-Habsi 2020). The features on $\mathcal{I}_\mathcal{T}$ specify information such as absolute 3D position in space, a segment in space containing the item, and color. Features enable receiving participants to estimate the matching probability between items coming from different observers.

In environments without noise, ambiguous similarity between items, and imperfect information, the problem is to determine a consistent interpretation for all items and edges in the scene, as well as an interpretation for each edge facet as either a material area or an empty space (Gaschnig 1978). In case of uncertainty, noise, and ambiguity, an optimization process is obtained.

Distributed Star Topology Algorithm

on `Init(`R`): /* The sequence for each robot on start */`
  The robot takes an image (image with the potential vertices and edges extracted) ;
  Robot filters the edges interpretations using the Waltz filter for a single process;
  Sends the output $O_R = \{(i, j, l)|i, j \in I\}$ to the central supervisor.;
Algorithm 1: Initialization of Robots

In a first version, robots are connected in a star-topology with one of them as a supervisor. In the Algorithm 1 the robot then takes an image of the object under examination. Using the EnrichedWaltz filter inspired from (Richter and Roth 2018) the robot computes the edges as a single process. It then sends the output of this result back to a central supervisor.

def `Supervisor():`
  forever *(;;)* **:**
    Get the messages O_R from all robots R;
    Composes the available data: $O = \text{filter}_R\ (O_R)$;
    if *no change* **then**
      terminate;
    Send message Data(O) to all robots;
Algorithm 2: Algorithm Supervisor Star Topology

In the Algorithm 2 the Supervisor gets the messages of the current round from all the robots. The supervisor's role is to then aggregate the data it receives from all the agents. It sends the aggregated data back.

on `Data(`D`):`
  Project data D on relevant perspective and run EnrichedWaltz on the data D;
  Sends the output $O_R = \{(i, j, l)|i, j \in I\}$ to the central supervisor.;
Algorithm 3: Message handling by robots

In Algorithm 3 robots run an EnrichedWaltz filter on the data and send the output of the data to the central supervisor. The last two processes repeat until convergence, which is guaranteed by the monotonicity and finite space of possible EnrichedWaltz interpretations for edges.

Algorithm for a Mesh

Another approach is to use a mesh communication between equal robots. In the Algorithm 4 each robot in the Mesh Network gets the edges from the filtering Algorithm, using the labeling sent to its neighbors.

In Algorithm 5, the labelings from the Mesh Neighbors are integrated into the aggregate and the EnrichedWaltz filtering Algorithm is rerun to check if there is any inconsistency. The output is sent to all neighbors.

on `Init:`
  Take image I;
  $E = edges(I)$;
  Labeling $L = EnrichedWaltz(E)$;
  send Labeling(I) to mesh neighbors;
Algorithm 4: Init Mesh

on `Labeling(`D`):`
  Integrate D with local labeling L;
  Run EnrichedWaltz on L;
  if *there is any modification to* L **then**
    Sends the output $L = \{(i, j, l)|i, j \in I\}$ to all the neighbors.;
  end
Algorithm 5: Message Handling on Mesh

Lemma 1 *The Algorithm 5 converges to the minimal consistent labeling, and the processing eventually terminates when the set of robots is connected.*

Proof The termination is due to the fact that the possible labelings are discrete and finite, and EnrichedWaltz only reduces the set labels. Stability and termination is guaranteed by the property of strict monotonicity.

Experimental Results

For the case of perfect information and 4 robots viewing a cube, we have simulated the implementations of the discussed communication models.

So far we have only analyzed the obtained computation time in simulator. The summarized results are:

Number of Edges	Supervisor Model with 4 robots (msecs)	Mesh Model with 4 robots(m/secs)
2	0.023	0.056
4	0.033	0.044
8	0.046	0.048
16	0.058	0.046
32	0.064	0.042
64	0.078	0.038
128	0.088	0.036
256	0.100	0.034
512	0.140	0.022
1024	0.230	0.021

Conclusion

We introduce a framework model for robots that collaborate to refine understanding of scenes by Distributed EnrichedWaltz Filtering. We have studied a comparison between two collaboration models. Improvements in speed and accuracy are enabled. Actual experiments are being run with the above frameworks and preliminary results will be presented with our poster.

References

Arbab, F., and Monfroy, E. 2000. Distributed splitting of constraint satisfaction problems. In *International Conference on Coordination Languages and Models*, 115–132. Springer.

Balafrej, A.; Bessiere, C.; Bouyakhf, E. H.; and Trombettoni, G. 2014. Adaptive singleton-based consistencies. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 28.

Bessiere, C.; Régin, J.-C.; Yap, R. H.; and Zhang, Y. 2005. An optimal coarse-grained arc consistency algorithm. *Artificial Intelligence* 165(2):165–185.

Cieslewski, T.; Choudhary, S.; and Scaramuzza, D. 2018. Data-efficient decentralized visual slam. In *2018 IEEE international conference on robotics and automation (ICRA)*, 2466–2473. IEEE.

Gaschnig, J. 1978. Experimental case studies of backtrack vs. waltz-type vs. new algorithms for satisficing assignment problems. In *Proceedings of the Second Canadian Conference on Artificial Intelligence*, 268–277.

Guzman, A. 1968. Computer recognition of three-dimensional objects in a visual scene. *PhD thesis, MIT AI-Lab.*

Huffman, D. A. 1971. Impossible objects as nonsense sentences. *Machine intelligence* 6:295–323.

Khan, A. I., and Al-Habsi, S. 2020. Machine learning in computer vision. *Procedia Computer Science* 167:1444–1451.

Mallya, A., and Lazebnik, S. 2015. Learning informative edge maps for indoor scene layout prediction. In *Proceedings of the IEEE international conference on computer vision*, 936–944.

Richter, S. R., and Roth, S. 2018. Matryoshka networks: Predicting 3d geometry via nested shape layers. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1936–1944.

Waltz, D. 1975. Understanding line drawings of scenes with shadows.