Semantic Computer Vision tools for mission-critical visual communications over bandwidth limited networks

Laura López Fuentes

University of the Balearic Islands, Crta. Valldemossa km. 7.5, Palma, Spain
Computer Vision Center, Autonomous University of Barcelona, Barcelona, Spain
AnsuR Technologies, Martin Linges Vei 25, 1364 Fornebu, Norway
l.lopez@uib.es

1 Information

- Title: Semantic Computer Vision tools for mission-critical visual communications over bandwidth limited networks
- Supervisor 1: Andrew D. Bagdanov, University of Florence
- Supervisor 2: Harald Skinnemoen, AnsuR Technologies
- Supervisor 3: Joost van de Weijer, Computer Vision Center
- Supervisor 4: Manuel González Hidalgo, University of the Balearic Islands
- Institution: University of the Balearic Islands, Dept. Mathematics and Computer Science. Ctra. Valldemossa km 7.5, 07122- Palma, Illes Balears, Spain.
- Starting date: 1st September 2015
- Estimated duration: 4 years

2 Summary

Context: This project is carried out in a collaboration agreement between AnsuR Technologies [**?**], the University of the Balearic Islands and the Autonomous University of Barcelona, in what is called an Industrial PhD. This type of PhDs aim to contribute to a practical problem inside a company, in this case AnsuR. Hence, before introducing the thesis proposal it is important to have some background about the hosting company.

AnsuR's efforts have been primarily focused on the development of novel transmission protocols at the upper layers (from IP up to Application layers) for optimal delivery of audio-visual content over bandwidth limited channels.

In the search for further optimization, it has been noticed that user experience in mission-critical communications has very specific needs that have not been yet explored. Namely, information content of audio-visual data that is relevant to the decision-maker might be a rather minimal part of what is actually transmitted. This means that content awareness would potentially help matching the (relevant mission-critical) information to be transmitted to the bandwidth limitation. This observation opens up the natural research direction of identifying ways that

Actas de la XVII Conferencia de la Asociación Española para la Inteligencia Artificial, pp. 965-969

consider semantic aspects for transmission optimization. In this work, computer vision will interact with interactive communications and geo-spatial processing in order to explore the substantial gains that are possible in communications efficiency.

Main goal: Our proposed work's primary objective is to research in the direction of what is known as "Semantic Information Theory". Specifically, the crucial purpose of the proposed PhD Thesis is to develop semantic computer vision tools that are relevant to the purpose of AnsuR's core technology.

The focus is communication of visual situational awareness in mission-critical situations, where lives or security could be at risk. A main objective is to communicate visual content that helps real-time understanding of critical situations in remote locations, following an Observe-Decide-Act paradigm.

As part of the decision process, analysis of the relevance and impact of observations must be done, so that communications network resources can be focused on primarily data with effect on operations. Such analysis is challenging, since the relevance of an image depends on the context images are used in, and specially if a similar image has already been received and considered relevant.

When sending visual data over limited communication networks, we use interactive communications technologies. Analysts at a receiver location will initially see a smaller photo/video version, but can interactively request details as needed at a cost that must be weighted against the value of the data.

Large amounts of images at some point becomes challenging to manage for humans. Therefore, computer vision enters the game to help understand and communicate relevant content. It can operate partly at the sender with access to the original data, and partly at the receiver with initially very reduced quality versions.

Sub-goals:

- Perform computer vision research into how the information semantic aspects modify traditional entropy-based information paradigms in mission critical communication networks.
- Develop a computer vision framework for exploring potential theoretical and practical gains of the suggested disruptive approach to information and communications in bandwidth-limited for mission critical networking.
- Applicability for a Situational Awareness tool: A tool to share what is going on, where - when it happens. Situational Awareness is a fundamental foundation for decision-making; in particular in any emerging situation where safety, security, lives, environment and other values are at stake. Images, both as photos and video, including satellite-based observations, as well as maps and geographical information systems (GIS), form an essential component in decision support.

3 Methodology, Workplan and Milestones

While keeping an open eye to other possibilities that may arise during the completion of the thesis, in this project we have established the following milestones:

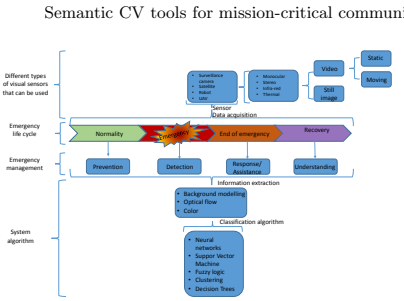

Fig. 1: Graphical overview of the survey's structure.

– Study of state-of-the-art algorithms developed in the computer vision framework to help prevent or identify emergencies or assist emergency responders during the emergency. Focusing on two key scenarios: surveillance systems that capture large amounts of visual information where much of the data is irrelevant and crowdsourcing of visual content through large numbers of users who send few observations each and where the relevance of the data is intended. This study has been already addressed and is being recompiled in the form of a survey, entitled "Computer Vision in Emergency Situations". The survey recompiles recent papers on state-of-the-art computer vision algorithms applied to emergency situations. Following the diagram from Figure **??**, the survey is structured in the following way: first the algorithms are grouped depending on the phase of the emergency in which they are used, then a classification of the algorithms is done depending on the sensor used for the data acquisition and finally depending on the feature extraction algorithms and classification algorithms used. The survey is expected to be submitted to a journal during the first year of the thesis.

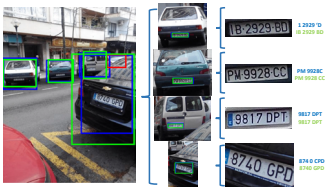

Fig. 2: Example of recognition results. Blue represents the output given by the algorithm on the different steps, green represents the ground truth with "Easy" difficulty and red the ground truth with "Hard" difficulty.

– Investigation and determination of the potential of deep Convolutional Neural Networks (CNN) for semantic analysis in support of visual situational awareness in emergency and critical situations. During this investigation, an active information seeking model has been already developed. This model identifies promising regions in low resolution imagery and progressively requests higher resolution regions on which to perform recognition of higher semantic quality. From this general framework, a car recognition system via the identification of its license plate has been developed. The system can save up to one order of magnitude of bandwidth when compared to other compression algorithms such as JPEG [**?**], while sacrificing little in terms of recognition performance. Some qualitative results of this application are shown on Figure **??**. In this example, car detection is performed on a low resolution image, the detected cars are then requested at a higher resolution on which license plate detection is performed. Finally the regions containing the detected license plates are requested at higher resolution to proceed to the license plate reading. The detection algorithms consist on two networks, based on the VGG_CNN_M_1024 network [**?**], fine-tuned on a self compiled and annotated dataset of cars in high resolution. The training is done using the Fast R-CNN framework [**?**]. The lecture of the license plate is performed with Tesseract [**?**], a popular optical character recognition engine. All the investigation and experiments have already been carried out and the work is being prepared to be submitted to a conference or journal from the field together with the car dataset that has been created.
– Study of the use of active learning techniques to obviate the need for expensive data annotation and usage of deep features to create an application that dynamically adapts to the user's needs. A first part of the thesis has been devoted to solve specific problems useful in emergency situations, such as the identification of cars. However it is infeasible to create a specific solution for each visual content that may be useful in an emergency situation. In order to solve this problem an unsupervised or semi-supervised model that could learn from the user's actions will be studied.
– Integration of the computer vision tools developed, which help understand the situation and focus communications bandwidth on relevant content, in AnsuR's crowdsourcing application. This work will improve the optimization of bandwidth and time resources during emergencies or critical situations, leading to a final application that can help save lives and protect society.

4 Relevance

This thesis is going to be developed in the context of an important company in the field of visual communications in emergency situations. Hence, this project will focus on two main problems: situation awareness from visual content for emergency situations and bandwidth optimization by sending only semantic relevant data for the application at task. In an emergency situation where lives may be at stake, time is a powerful and limited resource that is of crucial importance because a few minutes can make a difference between saving a person

or not. Therefore, automatic situational awareness determination through the visual content that is sent through the system is crucial to help emergency responders focus on relevant information and optimize resources. It is well known that in this kind of situations, bandwidth limitations are a general problem that can put at risk the precision and speed of the communication. This problem is addressed through the semantic active information seeking solution. Moreover, this solution is very novel in the field of computer vision where bandwidth limitation problems had only been addressed from the compression point of view, not giving any relevance to the semantic information of the image.

Automatically selecting semantic relevant data from the incoming visual content can help to speed up the response after an emergency. Moreover, saving big amounts of bandwidth is of crucial interest since it is precisely in these situations when the increase in call volume can cause the networks to collapse. Ensuring a fast and robust communication in crucial situations can help to optimize resources during the response to an emergency and therefore help to save lives.

To ensure that the investigation and work being developed throughout this thesis is of relevance and applicability to the company as well as of scientific relevance, the thesis is partially developed in the company's installations and partially at the University. Thanks to the close work with the company, all the work being carried out can be easily integrated in the company's system and application. The fast integration of the algorithms on the final system is of great relevance both for the users who can benefit from new features and also for the thesis since users can give relevant feedback and information about their particular needs. Among the partners and clients from AnsuR stand out important institutions such as the United Nations, the civil protection from Catalonia and Madrid and the National Bank from Norway. These institutions have shown a great interest on the work that has been carried out in the first year of the thesis.

References

1. Smith, R. "An overview of the Tesseract OCR engine". In Proc. of 9th International Conference on Document Analysis and Recognition, pp. 629-633, 2007.
2. Felzenszwalb, P. F., Girshick, R. B., McAllester, D., Ramanan, D. (2010). "Object detection with discriminatively trained part-based models". IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32(9), pp. 1627-1645, 2010.
3. Wallace, G. K. "The JPEG still picture compression standard". IEEE Transactions on Consumer Electronics, vol. 38(1), pp. xviii-xxxiv, 1992.
4. AnsuR Technologies URL: http://www.ansur.no/
5. K. Chatfield, K. Simonyan, A. Vedaldi, A. Zisserman. "Return of the Devil in the Details: Delving Deep into Convolutional Nets".British Machine Vision Conference, 2014 (arXiv ref. cs1405.3531)