# Author's Response to the Reviewers' Comments

for the thesis manuscript entitled:

*Fuzzy Reasoning for Multimedia Semantic Interpretation: Fuzzy Ontology Based Model for Video Indexing*

November $16^{th}$, 2016

I would like to thank the reviewers for their interest in my work and their relevant and helpful comments that will greatly improve and velarize the thesis manuscript. Thus, I tried to do my best to respond to the raised points.

The reviewers have brought up good comments and the opportunity to clarify our research objectives and results. In this document, I have checked all the comments provided by the reviewers and have made necessary changes according to their recommendations.

Yours sincerely, Mohamed ZARKA.

## Answers to reviewer: Professor Fakhri Karray

**Comment 1:** There are several typos and editorial mistakes and strongly recommend the candidate to carefully proofread the thesis.

**Answer 1:** We apologize for the typos and the editorial mistakes. They were considered and corrected.

**Comment 2:** The candidate made very good literature review and has highlighted recent work in the field. Possibly more would have been proposed especially recent approaches highlighted in the literature dealing with big data based video clip repositories indexing and search using on-line learning and domain adaptation. This is a very promising research direction and would wish to have the candidate mention it as future potential direction.

**Answer 2:** We thank the reviewer for his proposal.

In our dissertation, we focused mainly on an ontology-based framework to enhance semantic indexing abilities. Our aim was to prove that such a framework could enhance the accuracy of a semantic interpretation for a given content. We have conducted then some works on how to extract valuable knowledge to be populated within the proposed ontology, but we haven't contributed in the ontology content evolution.

As the reviewer pointed out, we think that domain adaptation could be a very interesting research direction to be considered. In fact, many research works are proving that deep neural networks have the ability to compute domain invariant features [Donahue et al., 2014, Yosinski et al., 2014, Liu et al., 2016, Kumar et al., 2016]. Then, and when dealing with big data based video contents, it could be interesting to consider deep neural networks when the distribution of the training data is different from the test one. Thus, we think that the proposed abduction engine (defined in the second contribution) could be improved by extracting cross-domain knowledge.

In the section 7.2, we added a paragraph to highlight that the domain adaptation is a track that we should follow in our future work.

**Comment 3:** The examiner would have wished to see more powerful performance metric such as recall performance and computation requirement (algorithmic time performance).

**Answer 3:** Mainly, we conducted four experimentations within three benchmarks:

- TrecVid2010 (for the contribution $C_1$),
- ImageClef2012 (for the contribution $C_2$),
- and ImageClef2015 (for the contribution $C_3$).

As discussed in section 2.3, each benchmark selects particular metrics to compute participants ranks. Generally, the Map (*Mean Average Precision*) is the most used metric.

As regards to the computation requirement, we have considered such a performance aspect only for the conducted experimentations within ImageClef2012 and ImageClef2015. In fact, we have not discussed the computation requirement for the TrecVid2010 preliminary conducted experimentation. In the latter, we used a small number of rules with a lightweight deduction engine.

**Comment 4:** Moreover, it is very important when dealing with large data set to apply distributed/cloud computing environment to handle real-time classification and detection of new concepts/relationship. This requires a certain type of the algorithm structure, not mentioned in details in this work. These are major issue on their own, and it is not expected the candidate to deal with all of them here. He simply needs to highlight them though in the final draft.

**Answer 4:** We highlighted some cloud/distributed based research works in section 3.4.2.

**Comment 5:** A lot of research work is being produced these days within the framework of big data and machine learning. Some recent approaches based on machine learning have been proven to be very powerful and the candidate needs to refer to this work. Moreover, these approaches come specifically applied to certain type of domains or classes of events (sports, movies, ads, ...). Hybrid approaches using existing solid techniques for concept detection coupled with machine learning algorithms to deal with concept relationship (context) is being proven very powerful. Candidate should mention these approaches in his references.

**Answer 5:** We thank the reviewer for his proposal.

In fact, and in recent literature, a growing number of research work are being exploring the effectiveness of deep neural networks based methods in multimedia analysis, in particular Deep Convolutional Neural Networks (CNN) [Girshick et al., 2014, Sainath et al., 2015, Jiang, 2015, Tong et al., 2015, Wu et al., 2015, Druzhkov and Kustikova, 2016].

We highlighted such interesting approaches in section 3.4.3, and as a potential future research direction in section 7.2.

**Comment 6:** I have some issues with the publication record (in terms of quantity and quality) and strongly suggest that the candidate should really work hard on publishing his work in well-known venues pertinent to the field.

**Answer 6:** We thank you for pointing this out.

In terms of publication quality, we published our second contribution within a well known journal (Springer MTAP with impact factor of 1.331). Actually, I am drafting a second journal paper that deals with our third contribution in order to highlight the obtained results. Yet, I target the following two journals for the submission: *ACM Multimedia Computing, Communications, and Applications* (Impact Factor of 2.465), or *Springer International Journal of Computer Vision* (Impact Factor of 4.270).

In terms of publication quantity, I would like to notice that handling large amount of data was a real hard task with the use of very classical computing machines. I will try to do my best to increase my publication rate. Thank you again for this comment.

As for industrial development, we prepared a patent (to be submitted within INNORPI: *National Institute of Standardization and Industrial Property*). This patent is entitled *"Dispositif d'Enrichissement Sémantique en utilisant des ONTOlogies (DESONTO) pour l'amélioration de l'indexation des contenus multimédias"*, and it details the semantic enhancement for multimedia retrieval systems through the use of fuzzy ontologies (as described in our second contribution C2). This patent is currently under review. Furthermore, we are preparing a second patent about the scalability aspect. This patent is entitled *"Annotation Sémantique Rapide des Images en utilisant des Ontologies (ASRIO) pour l'indexation des contenus visuels."*, and it proposes an alleviated computing task for semantic concept detection within multimedia contents. This patent is being drafted.

## Answers to Reviewer: Professor Sami Faiz

**Comment 1:** Le chapitre 4 est dédié à la première contribution du candidat: Mise en place d'un modèle à base de connaissances pour l'indexation de la vidéo. Le candidat a su, dans un contexte très technique, apporter une contribution à trois étapes. Les mises en œuvre et les résultats sont indéniables. En effet, l'outil d'annotation assistée et collaborative de la vidéo réalisé a fait l'objet de plusieurs expérimentations dans le cadre, notamment, de compétition internationale TrecViD2010. Nous aurions juste aimé avoir une discussion au cas où les connaissances pouvant être elles mêmes floues.

**Answer 1:** La représentation des connaissances dans une ontologie consiste à définir des faits et des assertions. Ces derniers définissent soit un individu $a$ comme étant une instance d'un concept $C$ ($\langle C(a)\rangle$), soit deux individues $a$ et $b$ qui sont reliés par un rôle $R$ ($\langle R(a,b)\rangle$) [Grau et al., 2008].

La logique floue est également employée pour représenter qu'un individu $a$ est une instance d'un concept $C$ avec un certain degré d'appartenance $n$ ($\langle C(a) \geq n\rangle$ avec $n \in [0,1]$), ou deux individus $a$ et $b$ sont reliés par un rôle $R$ avec un certain degré d'appartenance $n'$ ($\langle R(a,b) \geq n'\rangle$ avec $n' \in [0,1]$) [Stoilos et al., 2007, Stoilos et al., 2010, Horrocks et al., 2011]. La logique floue permet ainsi de manipuler l'incertitude au niveau des connaissances représentées.

Dans le cadre de notre travail de recherche, nous avons opté d'utiliser des bases de connaissances afin d'améliorer la qualité sémantique des systèmes d'indexation. Dans le chapitre 4, nous avons proposé une première modélisation pour un système d'enrichissement sémantique à base d'ontologies et de leurs capacités d'abduction et de déduction. Cette première modélisation a été mise en place par deux étapes :

- Dans la première étape (la section 4.2), nous avons proposé d'utiliser l'ontologie LSCOM [Kennedy and Hauptmann, 2006]. En effet, cette ontologie représente des concepts sémantiques (ex. *personne, président, voiture, immeuble, ...*) et des relations de généralisation entre ces concepts sémantiques (ex. *président* ***is a*** *personne*). Ainsi, les connaissances utilisées dans cette première partie ne sont pas floues ;
- dans la deuxième partie (la section 4.3), nous avons présenté une ontologie qui permet de manipuler des relations entres concepts sémantiques dans le cadre d'un contexte sémantiques particulier. De plus, ces relations sont floues. Nous avons proposé ainsi deux qualificateurs flous : *Strong* et *Weak* pour qualifier une relation entre deux concepts comme, respectivement, forte ou faible.

Ainsi, les connaissances qui ont été traité dans le chapitre 4 et dans la section 4.2 ont été extraites à partir de l'ontologie LSCOM, donc des connaissances non floues. Mais à

partir de la section 4.3, nous avons commencé à utiliser des connaissances floues. Avec l'expérimentation que nous avons mené sur la base de teste de TrecVid2010, nous avons obtenu des résultats qui montrent que l'utilisation de l'ontologie floue que nous avons proposé $O^f$, permet en effet d'améliorer la qualité d'un système d'indexation de la vidéo.

Étant donné que nous n'avons pas bien souligné dans notre manuscrit l'apport de l'utilisation des connaissances floues dans l'amélioration de la qualité d'indexation, nous avons apporté une modification à la section 4.3.5.

## References


[Donahue et al., 2014] Donahue, J., Jia, Y., Vinyals, O., Hoffman, J., Zhang, N., Tzeng, E., and Darrell, T. (2014). Decaf: A deep convolutional activation feature for generic visual recognition. In *ICML*, pages 647–655.

[Druzhkov and Kustikova, 2016] Druzhkov, P. N. and Kustikova, V. D. (2016). A survey of deep learning methods and software tools for image classification and object detection. *Pattern Recognition and Image Analysis*, 26(1):9–15.

[Girshick et al., 2014] Girshick, R., Donahue, J., Darrell, T., and Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 580–587.

[Grau et al., 2008] Grau, B. C., Horrocks, I., Motik, B., Parsia, B., Patel-Schneider, P., and Sattler, U. (2008). {OWL} 2: The next step for {OWL}. *Web Semantics: Science, Services and Agents on the World Wide Web*, 6(4):309 – 322. Semantic Web Challenge 2006/2007.

[Horrocks et al., 2011] Horrocks, I., Pan, J. Z., Stamou, G. B., Stoilos, G., and Tzouvaras, V. (2011). Reasoning with very expressive fuzzy description logics. *CoRR*, abs/1111.0039.

[Jiang, 2015] Jiang, Y. G. (2015). Categorizing big video data on the web: Challenges and opportunities. In *Multimedia Big Data (BigMM), 2015 IEEE International Conference on*, pages 13–15.

[Kennedy and Hauptmann, 2006] Kennedy, L. and Hauptmann, A. (2006). Lscom lexicon definitions and annotations (version 1.0). Technical report, Columbia University.

[Kumar et al., 2016] Kumar, S., Gao, X., and Welch, I. (2016). Learning under data shift for domain adaptation: A model-based co-clustering transfer learning solution. In *Pacific Rim Knowledge Acquisition Workshop*, pages 43–54. Springer.

[Liu et al., 2016] Liu, G., Yan, Y., Subramanian, R., Song, J., Lu, G., and Sebe, N. (2016). Active domain adaptation with noisy labels for multimedia analysis. *World Wide Web*, 19(2):199–215.

[Sainath et al., 2015] Sainath, T. N., Kingsbury, B., Saon, G., Soltau, H., rahman Mohamed, A., Dahl, G., and Ramabhadran, B. (2015). Deep convolutional neural networks for large-scale speech tasks. *Neural Networks*, 64:39 – 48. Special Issue on "Deep Learning of Representations".

[Stoilos et al., 2010] Stoilos, G., Stamou, G. B., and Pan, J. Z. (2010). Fuzzy extensions of OWL: logical properties and reduction to fuzzy description logics. *Int. J. Approx. Reasoning*, 51(6):656–679.

[Stoilos et al., 2007] Stoilos, G., Stamou, G. B., Pan, J. Z., Tzouvaras, V., and Horrocks, I. (2007). Reasoning with very expressive fuzzy description logics. *J. Artif. Intell. Res.(JAIR)*, 30:273–320.

[Tong et al., 2015] Tong, W., Song, L., Yang, X., Qu, H., and Xie, R. (2015). Cnn-based shot boundary detection and video annotation. In *2015 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting*, pages 1–5.

[Wu et al., 2015] Wu, Q., Zhang, H., Liu, S., and Cao, X. (2015). Multimedia analysis with deep learning. In *Multimedia Big Data (BigMM), 2015 IEEE International Conference on*, pages 20–23.

[Yosinski et al., 2014] Yosinski, J., Clune, J., Bengio, Y., and Lipson, H. (2014). How transferable are features in deep neural networks? In *Advances in neural information processing systems*, pages 3320–3328.