ENHANCED IMAGE APPLICATIONS FOR HIGH DATA RATE HF CHANNEL

C. Carincotte[1], R. Elmostadi[2], K. Hagihara[3]
[1] Multitel
Mons, Belgium
carincotte@multitel.be
[2] Thales Communications
Colombes, France
rachid.elmostadi@fr.thalesgroup.com
[3] U.C.Louvain
Louvain-La-Neuve, Belgium
kaori.hagihara@uclouvain.be

ABSTRACT

The goal of this paper is to present innovative image applications that will be available on future high data rate HF transmissions. Fitting in the considerations and studies carried out currently on future HF transmissions, for instance within NATO groups, these new image applications are currently under development in the EDA HDR-HF project, which aims at developing new communication technologies to offer a high bit data rate transmission system for data and multimedia military applications over HF channels. In this paper, we present the work currently on-going on these new multimedia applications, mainly focusing on JPEG 2000 and JPIP image transmission, and on image enhanced analysis such as automatic object detection, and image registration and 3D rendering.

Keywords: HF transmission, Image applications, JPEG2000/JPIP transmission, Automatic object detection, Image registration, 3D rendering.

INTRODUCTION

With the advent of new technologies such as H.264 or JPEG2000, the possibilities of data compression, and thus data transmission, have been considerably widened in the last decade. In addition, the desired military applications of the future clearly reflect a growing operational need in terms of data and multimedia transmission, such as image or video transmission over HF. In this context, the HDR-HF project [1] aims at developing new technology to offer a high bit data rate transmission system for data and multimedia military applications over heterogeneous networks channel, and specifically over HF network. In particular, the system currently under development will enable the transmission of audio messages, images and data transmissions with a strong robustness towards the error prone transmission channel. In this paper, we present the work currently on-going on these new multimedia applications, mainly focusing on JPEG 2000 and JPIP image transmission, and on image enhanced analysis such as automatic object detection, and image registration and 3D rendering.

The remaining of the paper is organized as follows; first, next section introduces the multimedia system within the overall HDR-HF functional architecture. We then present the multimedia sub-system in more details, from proposed image services/applications to sub-system architecture. Following sections are then devoted to the four corresponding scenarios in terms of applications, namely progressive/interactive image transmission, automatic object detection in image, and image registration and 3D rendering. Conclusions and perspectives are last drawn in the last section.

HDR-HF ARCHITECURE OVERVIEW

The HDR-HF functional architecture, sketched in Figure 1, shows the separation of the system into four different sub-systems, with very different roles and modalities. The *multimedia sub-system* (the two extremities of the chain) consist of a multimedia server including image, audio or data acquisition and analysis at the server side, and the corresponding decoding and display operations at the multimedia client side. In between one find the *IP gateway* responsible for routing the messages, then the *HF modem*, which is in charge of performing the adaptation to the radio medium, and last, the *radio module* itself, responsible for over high data rate HF transmission. It worth pointing out that this paper does not intend to provide a detailed description of the whole chain, but will focus on the multimedia server/client; interested readers may consult [2] for a detailed description of the whole HDR-HF system and sub-systems.

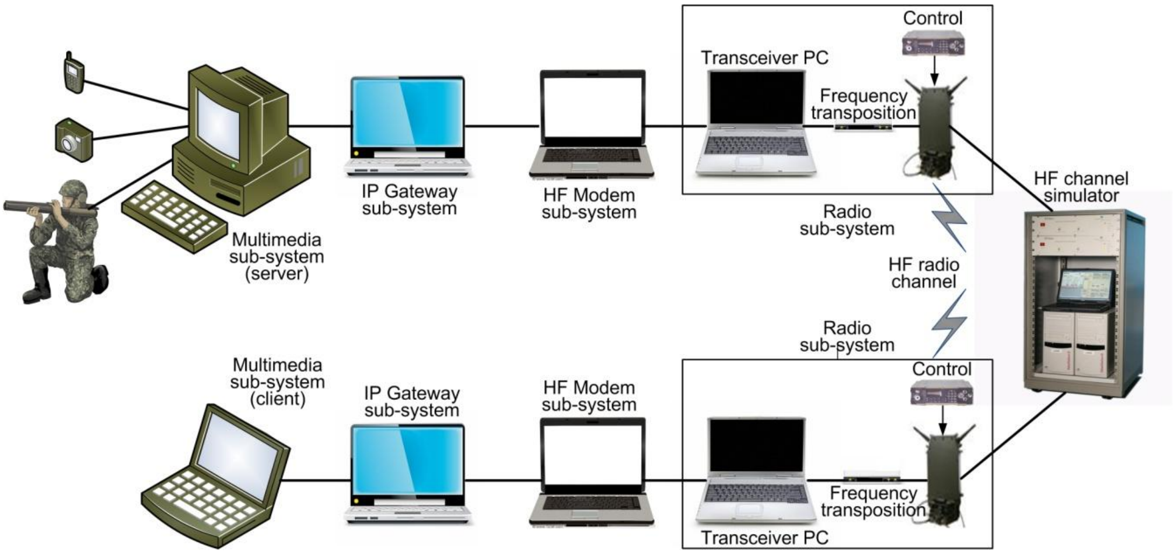


Figure 1 : HDR-HF functional architecture.

MULTIMEDIA SUB-SYSTEM DESCRIPTION

Main goals of HDR-HF at the multimedia application level are twofold;

1. to enable the robust processing, compression and transmission of image data, which often require throughputs of tens of Kbits/s at the transmission layer;
2. to offer new multimedia applications, with enhanced image services, such as automatic analysis of image for object detection, image registration and 3D rendering.

Enhanced image services/applications

At the image services/applications level, three main scenarios have been defined;

- Progressive/Interactive image transmission; the goal of this scenario is to transmit images (or part of them) progressively from a server to clients, benefiting from JPEG2000 encoding, which allows progressive image transmission, and JPIP, which allows interactive image transmission. In addition, at the server side, images can be manually annotated with their regions of interest (ROIs). Then, ROIs data is firstly streamed in high resolution to allow a first high resolution display of relevant parts of images, while the remaining of the image is streamed in low resolution and is progressively updated.

- Automatic object detection in image; following the progressive image transmission, the goal of this scenario is to transmit relevant/interesting regions of images, but using an automatic detection of these ROIs. Practically speaking, at the server side, images are automatically processed using specific image analysis applications related to the ROIs content, so as to detect and encode these ROIs. Then, image data of those ROIs is firstly streamed in high resolution, while the remaining is streamed in low resolution and is progressively updated. In this context, one algorithm has been developed to allow the automatic detection of airplanes in aerial images. While not presented in this study, it worth noting that a second algorithm has also been investigated, to allow the automatic detection of person in infra-red images.
- Image registration and 3D rendering; in order to provide more usability of transmitted images over JPIP protocol, these two scenarios are designed. Image registration is to compare an image transmitted over JPIP with an image in local storage along with one coordinate system. We consider a situation in which the server and the clients share a certain image in their own storage, and the server sends an image to the client to update the image during a mission. In this context, even if both images are not acquired from exactly the same point of view, with the same resolution or quality, and by the same modality, they must be aligned/registered to be compared, and to allow the update of the old image. 3D rendering is to browse images transmitted over JPIP three dimensionally using elevation maps. Following the nature of JPIP, i.e. the quality or the resolution of the transmitted image refines progressively, the texture on the 3D reconstruction refines progressively as well. These two sub-modules are developed to be optimal to JPIP transmission module.

Multimedia sub-system architecture

The image sub-modules are described in Figure 2 for the server part and Figure 3 for the client part. The server architecture is the most important since most of the processing takes place at the server side, and several blocks have to be synchronized. Depending on the scenarios, external data is used for the processing, and metadata is produced, in the form of a XML file, which is then further embedded in a JP2 file.

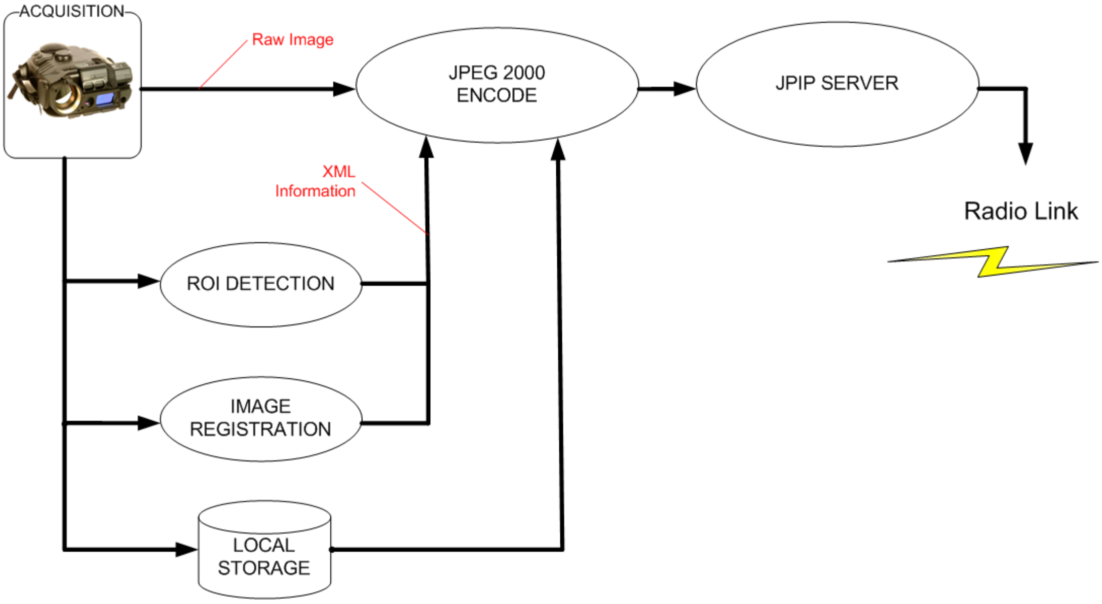


Figure 2 : Server architecture for the image application.

The client architecture is more straightforward, and involves mainly the decoding of the JPEG2000 file, and the extraction of the XML information, to be used by the Graphical User Interface (GUI). User interaction is handled through the GUI and the JPIP client.

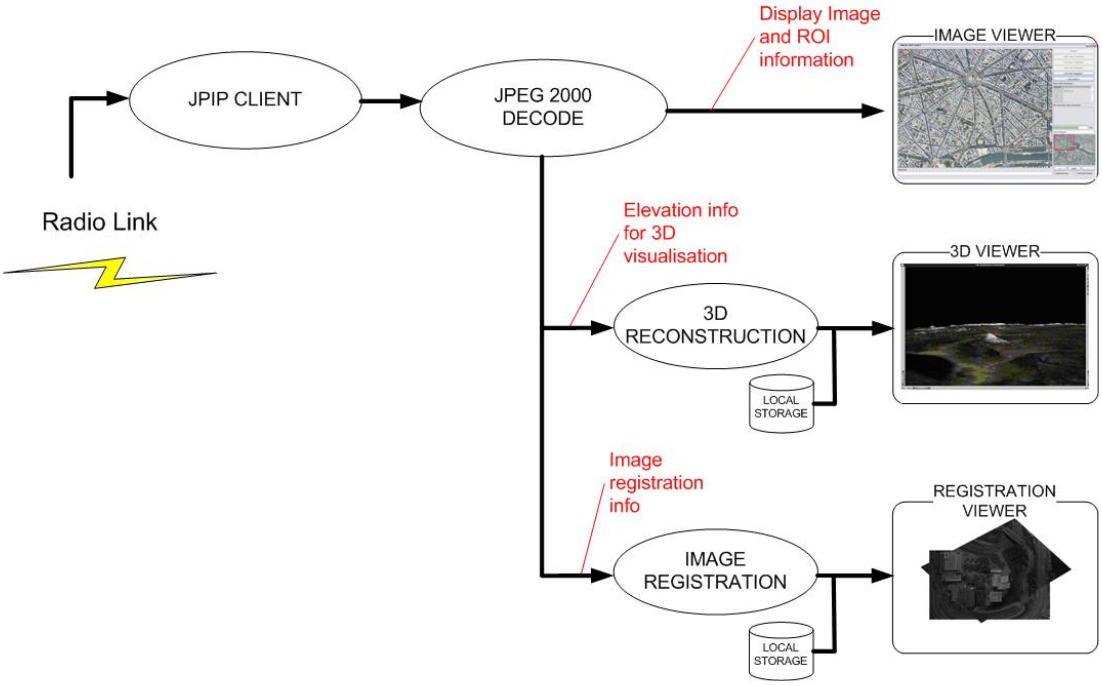


Figure 3 : Client architecture for the image application.

PROGRESSIVE/INTERACTIVE IMAGE TRANSMISSION

At the transmission stage, the main challenge is to ensure ad-hoc bandwidth allocation and robust transmission over the error-prone HF channel. The project therefore worked on improving the JPEG 2000 Interactive Protocol (JPEG2000 part 9 – JPIP [3, 4, 5]).

This protocol, as standardized in ISO/IEC 15444-9 and ITU-T T.808 [4], defines the communication syntax between a server and a client and provides a progressive and interactive way of transmitting images, without the need of waiting for the complete image transfer, also integrating the state of the art and the most efficient image codec. Based on the scalability properties of JPEG2000, JPIP protocol allows getting access to images from the field. The client can request the server to send specific Region of Interest to avoid complete image transmission, and/or to prioritize portions of images. Only data necessary to reconstruct the required part of an image are sent in order to limit the bandwidth and to access quickly to the interesting data. The delivery of image data over a network (with TCP or UDP protocol) is incremental by quality, resolution and spatial region, which allows stopping transmission when the necessary level of image data information has been received.

The following figure shows a transmission of an aerial image (2340*2304) at the rate of 10 kb/s; the image is transferred in resolution increment order. Figure 5 shows the GUI of the client and the server and the exchange between them.

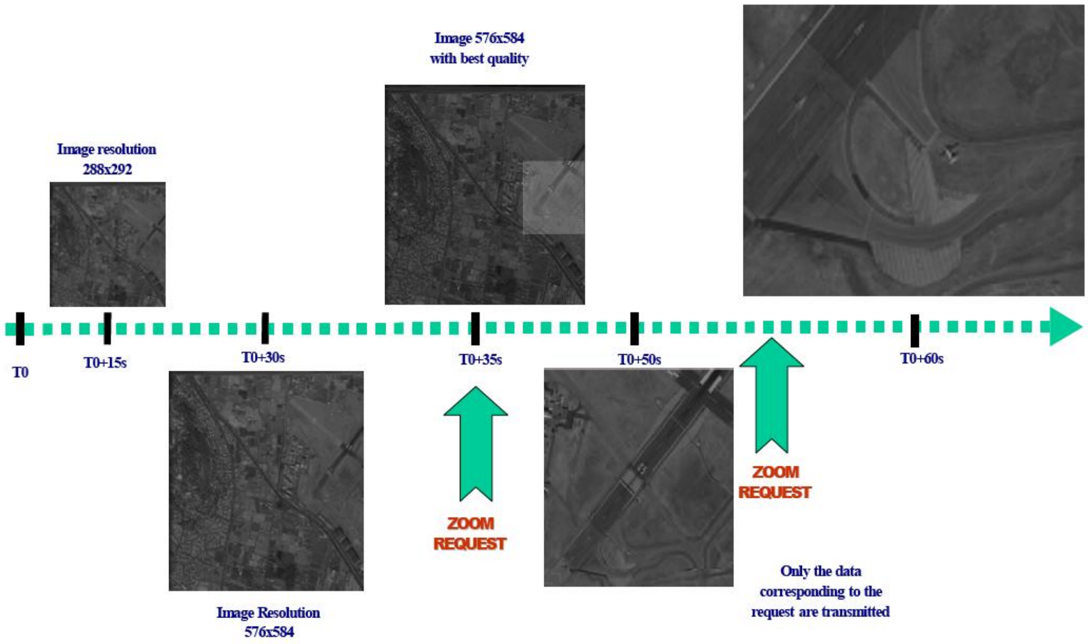
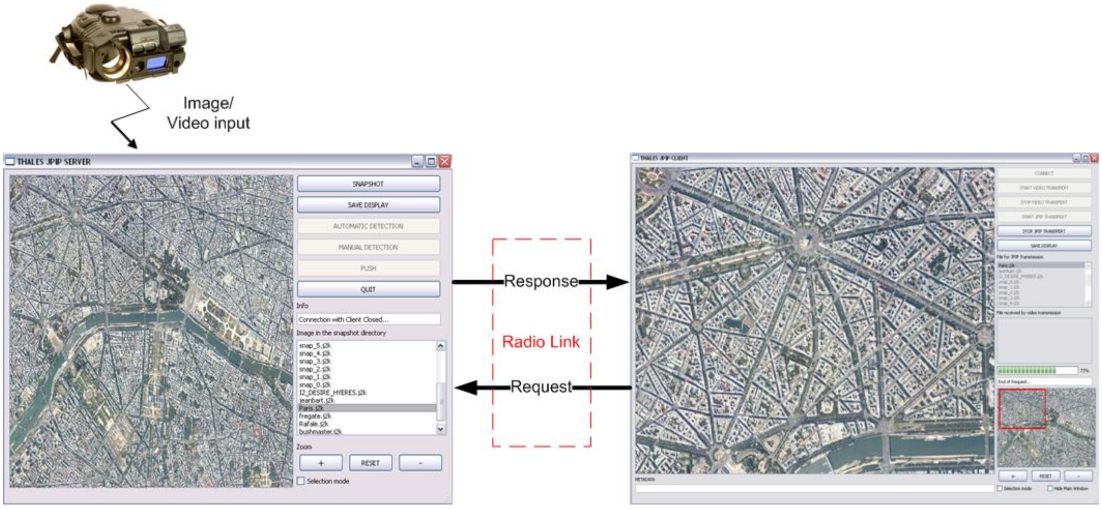


Figure 4 : Example of JPIP Transmission**.**

At the server side, the application is connected to a camera and can take snapshot. The snapshot directory contains JPEG2000 image and snapshots taken by the camera. Additional metadata (geo-location, position of ROI, etc.) can be transferred together with the image (or alone) through the JPIP transmission protocol, for better understanding and exchange of images. The server can then either select manually or automatically a specific ROI in the image and put this information in the metadata of the JPEG2000 file. This permits to inform the client for the presence of a specific region of interest. This region of interest is transmitted before the other part of the image. At the client side, the application receives after the connection, all the file name existing in the server side. By selecting one file, the application receives the stream and displays it in the GUI. The client can make request (mouse selection, zoom …) to the server to receive only the selected area.



Connection with Client Closed...

Figure 5 : JPIP Server and JPIP Client GUI.

AUTOMATIC ROI DETECTION IN IMAGE

In the context of the project, the automatic ROI detection which is currently under study is the detection of airplanes from aerial images. Indeed, few scientific literature exist on this specific task, but main ones are dedicated to military applications and quite recent [6, 7]. Chalmond et al. [6] propose a two-step detection procedure, first detecting salient structures by extracting local edges using Gabor filters at multiple scales, and propose a method to estimate the appropriate scale. The contrast and scale information are then combined in a probabilistic model, classifying each pixel into object/non-object. Perrotton [7, 8] builds on the Viola and Jones method and modifies the Haar features so that it is no longer a difference of sum between two regions, but a distance between histograms of the two Haar-regions. Perrotton used the soft cascade of Bourdev et al [9], which reduces the cascade to a single boosted classifier, where the stages correspond to partial sums in the perceptron. The orientation problem is solved by using an unsupervised method to form clusters of appearances/orientations and using a single cascade instead of orientation-specific classifiers.

In this study, we investigate the use of a cascade of classifiers using Haar features, as introduced by Viola and Jones, and propose to study the rotation invariance of the detector for airplanes. In more details, the structure of the cascade and the learning algorithm are the original ones, while the used features are the extended set of Haar-like features proposed by [10]. Given a feature set (here Haar-like features) and a training set of positive and negative images, a variant of AdaBoost is used both to select a small set of features and train the classifier AdaBoost learning algorithm is used to boost the classification performance of a simple (sometimes called weak) learning algorithm. For each feature, the weak learner determines the optimal threshold classification function, such that the minimum number of examples is misclassified. A weak classifier $h_j(x)$ thus consists of a featuref_j, a threshold θ_j and a parity p_j indicating the direction of the inequality sign:

$$h_j(x) = \begin{cases} 1 \text{ if } p_j\, f_j(\mathrm{x}) < p_j\, \theta_j \\ 0 \text{ otherwise} \end{cases}$$

Table 1 provides a summary of the boosting process.

Table 1 : The AdaBoost algorithm for classifier learning (Viola and Jones, 2001).

- Given example images $(x_1, y_1), \dots, (x_n, y_n)$ where $y_i = 0, 1$ for negative and positive examples respectively.
- Initialize weights $w_{1,i} = \frac{1}{2l}, \frac{1}{2l}$ for $y_i = 0, 1$ respectively, where m and l are the number of negatives and positives respectively.
- For $t = 1, \dots, T$
  1. Normalize the weights, $w_{t,i} \rightarrow \frac{w_{t,i}}{\sum_{j=1}^{n} w_{t,i}}$ so that w_t is a probability distribution.
  2. For each feature j, train a classifier h_j which is restricted to using a single feature. The error is evaluated with respect to w_t, $\in_j = \sum_i w_i |h_j(x_i) - y_i|$.
  3. Choose the classifier h_t, with the lowest error $\in_t$.
  4. Update the weights:

$$w_{t+1,i} = w_{t,i} \beta_t^{1-e_i}$$

where $e_i = 0$ if example x_i is classified correctly, $e_i = 1$ otherwise,

and $\beta_t = \frac{\in_t}{1-\in_t}$.

- The final strong classifier is:

$$\mathrm{h}(x) = \begin{cases} 1 \displaystyle\sum_{t=1}^{T} \alpha_t\, h_t(x) \geq \frac{1}{2} \sum_{t=1}^{T} \alpha_t \\ 0 \text{ otherwise} \end{cases}$$

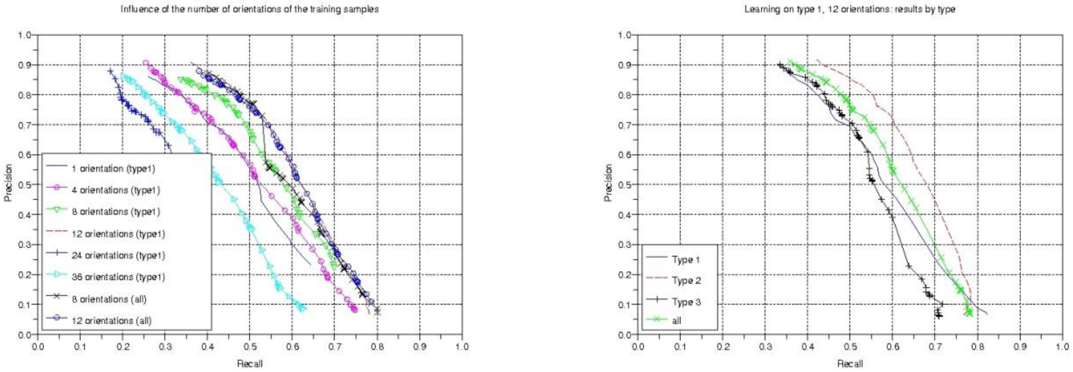
where $\alpha_t = \log \frac{1}{\beta_t}$.

The dataset we built for the experiment is composed of approximately 1800 images collected from Google Earth, at various airports locations and from the dataset of planes in flight collected by Google Earth enthusiasts[1]. These images contain a total of 6213 airplanes, ranging in size from 20x20 pixels up to 220x200 pixels. The images themselves are medium-sized, approximately 1600x1100 pixels. Data is collected from all over the world, on various backgrounds (fields, mountains, sea, urban, desert…) and at various elevations (from parked planes to planes in flight at unknown altitude). Shapes also vary a lot, from small general aviation planes and gliders, to large regular commercial airplanes,

---

[1] Google Earth Community: All Aircraft in Flight, April 2010.

as well as military airplanes. Since data in Google earth is multi-source, image quality is not uniform and ranges from excellent to poor. Shadows and ghosts (due to sensor acquisition artifacts) are also present. We believe It worth noting that this dataset is significantly harder than those previously used in [6, 7], and requires significant work to achieve good performances. Indeed, this study confirms the conclusions of [8], contrary to what was stated by [11] for faces, i.e. that a single cascade is able to learn many orientations for specific objects, and in pour case for airplanes. Moreover, learning a single type only does not prevent from detecting the others with same or even better results than for the learned type (see Figure 6-(b)), but the split of the training set into different types does not improve the results over the classifiers that uses all types (see Fig. Figure 6-(a)).



(a) Multi-orientation classifier performances for different number of orientations.

(b) Multi-orientation performances for object type.

Figure 6 : Detection results for different configurations of the algorithms.

A visual inspection (see Figure 7) shows that the false alarms are mainly due to man-made structures, such as painted lines on airports tarmacs, roads intersections, buildings. Non-detection may be due to poor contrast, low resolution, or complex backgrounds.

Figure 7 : Illustrations of detection results (green bounding boxes correspond to good detections, red ones to false alarms, and blue circles to missed detections).

These obtained results can still be improved, e.g. using higher detection rate target in the cascade. The optimal number of neighbors (a detection is considered valid only if it has k neighbors), the best performance is obtained using 8 orientations and all types of planes, for a scale factor of 1.01, k = 13, resulting in a precision of 0.75 and recall of 0.63. The detection time is below 1 second per image for a scale factor of 1.2 on a 2.8 GHz laptop.

IMAGE REGISTRATION

An optimum approach to realize the image registration scenario is to compute the alignment factors between images at the server side, and to send the alignment factors in addition to the image. This way, the client does not need to do heavy computations to find the alignment factors, and the server can distribute the image alignment factors such as transformation parameters, and reference image ID. Such information can be transmitted as metadata-bin message over the JPIP.

In spite of many papers proposing image registration methods, there are no universal methods applicable to all registration tasks due to diversities of images (e.g. degradations and feature densities) [12]. Intensity-based methods are efficient if images do not contain complex structures since we compute the matching level from the entire images. Many algorithms use metrics such as mean squares, normalized correlation, and mutual information [13, 14]. On the other hand, feature-based methods are efficient if images contain complex structures. These methods generally extract salient regions from both images, and find corresponding ones by template matching, contour matching [15], or curvature matching, etc. Our approach, therefore, is to supply modules based on a variety of algorithms and apply one of them or a couple of them one after the other, depending on image contexts. Hereafter, we present one of our validated algorithms with its implementation results. Notice, we target images found in the domain of computer vision and remote sensing for the military use. Thus we consider true alignments matches roughly with linear transformations. The transformation from a pixel $\boldsymbol{p}$ in the first image to $\boldsymbol{p'}$ in the second image can be represented as following:

$$\boldsymbol{p'} = \boldsymbol{Hp}$$

where $\boldsymbol{H}$ is 3*3 homography matrix, and $\boldsymbol{p}$ and $\boldsymbol{p'}$ are homogeneous vectors.

Kanazawa and Kanatani proposed a method for detecting global consistent point matches between two images, which leads good image registration results [16]. They define the “confidence” of the “soft constraints”, which preserve the global consistency, to all potential matches. The confidence is progressively updated by “mean-field approximation”. Finally, the “hard” epipolar constraint is imposed by RANSAC.

We exploited a pair of cropped images (510x1200) from band 2 and 15 of a hyper-spectral image on this method (See Figure 8-(a) and (b)). In order to simulate sequential shots from a moving camera, slight level (one degree) of rotational distortion is applied to one of the cropped image.

We detected 300 feature points separately using the Harris operator. Figure 8-(c) is the “optical flow” of the initial candidate matches based on local normalized correlations. This scene has many periodic patterns, and the simple template matching by only local correlations produces many mistakes. Figure 8-(d), (e), and (f) shows the refinement after imposing spatial consistency, plus global smoothness, and finally plus the epipolar constraint. We obtained 63 matches at the end, and computed a homography matrix for the image alignment. The difference image against the ground truth (See Figure 8-(h)) contains white dots at edges in the overlap area, meaning non accuracy at the sub pixel level. However, it is sufficient at the pixel level.

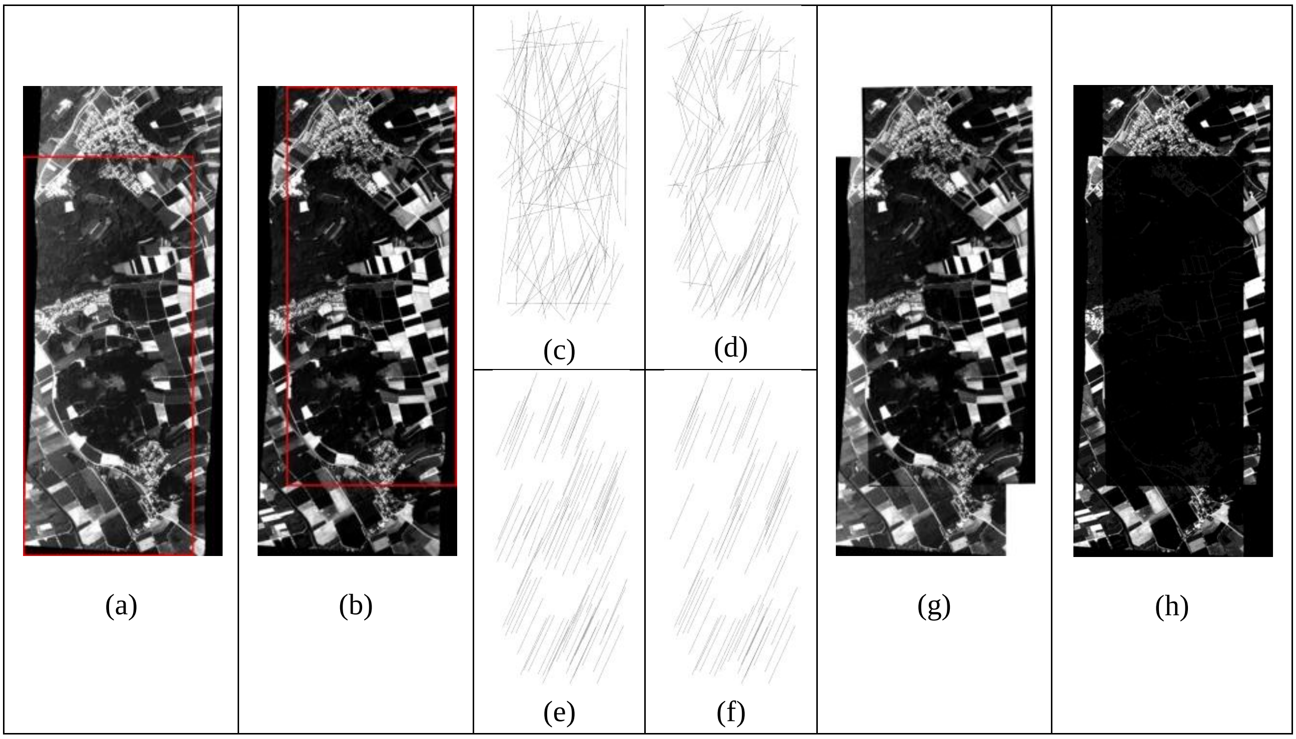


Figure 8 : Multi-modal image registration results.
(a) and (b): band 2 and 15 images, red rectangles correspond to cropped areas,
(c) to (f): Refinement of optical flow, (g) Mosaic of band 15 and band 2,
(h) Difference against the ground truth.

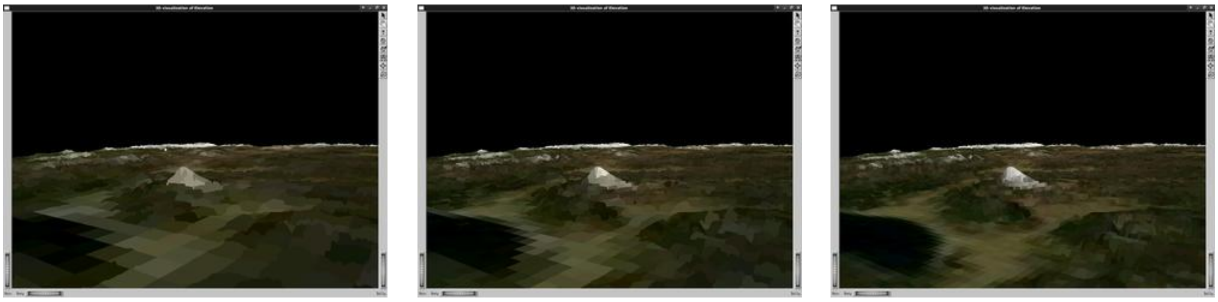
3D RENDERING

3D rendering functions as an optional viewer of images sent by JPIP in this context. Reconstructed structure is as simple as representing every image pixel in 3D position. The elevation/depth of the image is considered to be available initially or by pre-processing (c.f. image registration, stereo vision). Except the output representation, our approach is common to the one proposed for image registration scenario, since we have to align the transmitted image and the elevation map. We developed a 3D rendering module which meets the specific character of the image reception and reconstruction through JPIP protocol. As the quality or resolution of an image is progressively improved, its 3D representation refines progressively too.

More technically, the 3D rendering module is called by JPIP client and stays active as long as the viewer is open. First, this module performs image alignment between the elevation map and the image from JPIP client as described above. Then, a 3D mesh surface is reconstructed from elevation map and the image is mapped onto this surface. Every time JPIP client refines the image, the mapping image is also refined right after. We implemented this module using the topography and colour maps of the earth from a NASA project VISIBLE EARTH (see Figure 9).

Figure 10 shows the 3D rendering output through the OpenInventor library 3D viewer. By nature, the viewer supports rotation of the 3D textured surface and zooming in and out to the central point. We could recognize the refinements of the texture on the reconstruction.

Figure 9 : Colour (Left) and Topography (Right) maps.



Resolution: 64x64 Resolution: 128x128 Resolution: 256x256

Figure 10 : 3D rendering view at coarse to fine resolution levels of texture.

CONCLUSION

In this paper, we presented the innovative image applications/services available for high data rate transmissions over HF channels. The concerned work mainly focus on JPEG2000 and JPIP image transmission, and on image enhanced analysis such as automatic airplanes detection from aerial images, satellite image registration and related 3D rendering.

Regarding the JPEG2000 and JPIP image transmission, the proposed application allows to perform some progressive image transmission over the HF channel, and allows the client to make request (selection, zoom …) to the server to receive only the selected area. As for the detection of airplanes from aerial images, it allows to process, at the server side, images coming from various sources to detect airplanes, and to firstly streamed them to the client in high resolution, while the remaining of the image is streamed in low resolution, and progressively updated using the classical JPEG2000/JPIP processes. As for the registration application, it allows to compare a newly sent image with an image already stored locally at the client side, to update only new data coming from the new image. Last, the 3D rendering application allows constructing a 3D mesh surface from an elevation map and an image mapped onto this surface.

Next steps focus on the integration of the overall HDR-HF architecture (currently in progress) and the first on-air transmissions for assessment of the overall image data transmission in real scale (planned for the end of 2011).

ACKNOWLEDGEMENTS

This work has been carried out thanks to the B-0448-IAP4-GC HDR-HF project, which is partially funded by the Walloon region and the German and French Ministries of Defence, and contracted by the European Defence Agency (EDA), in the framework of the IAP4 Captech. The consortium is composed of four partners, from three countries: THALES Communications France (TCFR) and THALES Communications Belgium (TCB), the Fraunhofer Institute for Integrated Circuits (FHG IIS), MULTITEL and the Université Catholique de Louvain (UCL). The authors want to acknowledge all the HDR-HF partners for their contribution to the studies presented in this paper.

BIBLIOGRAPHY

[1] Contract No B-0448-IAP4-GC covering the Ad Hoc Research & Technology Project HDR-HF, partially funded by the European Defence Agency, in the framework of the IAP4 Captech.

[2] I. Icart, R. Elmostadi, J.-Y. Bernier, R. Koch, “A Very High Data Rate System for IP communications over HF links”, in *Military Communications and Information Systems Conference*, Amsterdam, The Netherlands, October 2011.

[3] A. Skodras, C. Christopoulos, and T. Ebrahimi, "The JPEG 2000 Still Image Compression Standard ", in *IEEE Signal Processing Magazine*, pp. 36-58, September 2001.

[4] ISO/IEC 15444, Information Technology, JPEG2000 Image Coding System - Part 9: Interactive tools, APIs and protocols, 2003.

[5] D.S. Taubman and R.Prandolini, "Architecture, Philosophy and Performance of JPIP: Internet Protocol Standard for JPEG2000", in *Visual Communications and Image Processing*, 2003.

[6] B. Chalmond, B. Francesconi, and S. Herbin, “Using hidden scale for salient object detection,” in *IEEE Trans. on Image Processing*, vol. 15, pp. 2644–2656, 2006.

[7] X. Perrotton, “Automatic object detection in digital images - Application to aerial images”, *Ph.D. Thesis, EADS Innovation Works and Télécom ParisTech*, 2009.

[8] X. Perrotton, M. Sturzel, and M. Roux, “Implicit hierarchical boosting for multi-view object detection,” in *IEEE Conference on Computer Vision and Pattern Recognition*, June 2010.

[9] L. Bourdev and J. Brandt, “Robust object detection via soft cascade,” in *IEEE Conference on Computer Vision and Pattern Recognition*, 2005, vol. 2, pp. 236–243.

[10] R. Lienhart and J. Maydt, “An extended set of haar-like features for rapid object detection,” in IEEE Conference on Image Processing, 2002.

[11] M.J. Jones and P. Viola, “Fast multi-view face detection,” in *IEEE Conference on Computer Vision and Pattern Recognition*, 2003.

[12] B. Zitova and J. Flusser. “Image registration methods: a survey”, in *Image and Vision Computing*, Vol. 21, No. 11, pp. 977-1000, October 2003.

[13] P.Viola and W.M.Wells III. “Alignment by maximization of mutual information”, in *International Journal of Computer Vision*, Vol. 24, N°2, pp. 137-154, 1997.

[14] D.Mattes, D.R.Haynor, H.Vesselle, T.K.Lewellen, and W.Eubank, “PET-CT image registration in the chest using free-form deformations”, in *IEEE Trans. On Medical Imaging*, 22(1):120-128, 2003.

[15] H.Li, B.S.Manjunath, and S.K.Mitra. A contour-based approach to multisensor image registration, in *IEEE Trans.on Image processing*, 4(3):320-334, 1995

[16] Yasushi Kanazawa and Kenichi Kanatani, “Robust Image Matching Preserving Global Consistency”, in *Proc. 6th Asian Conference on Computer Vision*, pp.1128—1133, 2004.