ADAPTATION DES IMAGES ET DES VIDEOS POUR DES UTILISATEURS MULTIPLES DANS DES ENVIRONNEMENTS HETEROGENES

par

Lotfi Aouissi

Mémoire présenté au Département d'informatique
en vue de l'obtention du grade de maître ès sciences (M.Sc.)

FACULTÉ DES SCIENCES
UNIVERSITÉ DE SHERBROOKE

Sherbrooke, Québec, Canada, janvier 2008

Pagination non continue
mais complet tel quel.

Your file Votre référence
ISBN: 978-0-494-49446-2
Our file Notre référence
ISBN: 978-0-494-49446-2

Le 7 octobre 2008

le jury a accepté le mémoire de M. Lotfi Aouissi dans sa version finale.

Membres du jury

M. Djemel Ziou
Directeur
Département d'informatique

M. Abdelhamid Benchakroun
Membre
Département d'informatique

M. François Dubeau
Président-rapporteur
Département de mathématiques

A mon cher père et ma chère mère.

A mes frères.

A tous mes amis.

SOMMAIRE

La dernière décennie a connu l'émergence de l'utilisation des équipements mobiles comme les assistants personnels et les téléphones, ainsi que la prolifération des réseaux personnels favorisée par le développement considérable dans les technologies de communications. D'autre part, l'information véhiculée a travers le World Wide Web devient de plus en plus visuelle (images et videos) grâce à la numérisation. Afin de permettre à tous les usagers un accès universel à cette information visuelle dans un environnement caractérisé par la diversité des équipements et l'hétérogénéité des réseaux, il devient nécessaire d'adapter les documents multimédia. L'adaptation consiste à appliquer une ou plusieurs transformations sur un document multimédia. Dans ce cadre, plusieurs travaux ont été élaborés en partant de différentes formulations. Nous pensons qu'un système d'adaptation efficace doit choisir les traitements nécessaires à appliquer sur un document visuel afin de maximiser la satisfaction de l'usager. Il doit considérer conjointement les caractéristiques de cet usager ainsi que les performances de son équipement, la qualité de sa connexion et les conditions de son environnement. La majorité des travaux réalisés dans ce domaine n'ont traité que des cas limités, par exemple ajuster une video pour la capacité d'un reseau donné. Dans la présente recherche, nous proposons une solution globale obtenue à l'aide d'un modèle probabiliste qui utilise les traitements des images et des videos et l'extraction des caractéristiques des contenus.

REMERCIEMENTS

Je tiens premièrement à remercier mon directeur de recherche, le Professeur Djemel Ziou pour sa grande disponibilité, ses précieux conseils et la confiance qui m'a accordée.
Je remercie également mes collègues du laboratoire MOIVRE (MOdélisation en Imagerie, Vision et RÉseaux de neurones) *Zahir*, *Ali*, *Touati*, *Nizar*, *Mohand*, *Sabri*, *Samy* et *Ouael* pour leur soutien et l'agréable ambiance de travail. Je tiens à remercier aussi mes collègues *Ahmed* et *Sidali* pour leur soutien.
Je tiens à remercier les laboratoires universitaires Bell et le Conseil de Recherches en Sciences Naturelles et en Génie du Canada (CRSNG) pour leur soutien.
Mes remerciements s'adressent à tous les membres du Département d'informatique de l'Université de Sherbrooke pour leur disponibilité et leur contribution à ma formation.
Enfin, j'adresse un remerciement particulier pour toute ma famille ainsi que toutes les personnes qui me sont proches pour leur soutien.

Table des matières

Image and Video Adaptation for Multiple Users in Heterogeneous Environments

Lotfi Aouissi, Djemel Ziou.

January 12, 2008

Abstract

In this paper, we introduce a new framework to achieve universal multimedia access by adapting image and video content. To give a global formulation able to deal simultaneously with the different factors that influence the adaptation process, the description of these factors and their mutual relationships is inspired by part 7 of the MPEG-21 standard, entitled Digital Item Adaptation (DIA). The selection of the optimal adaptation operation is based on the prediction of utility value. The utility value reflects the goodness of the adaptation with respect to the nature of the content, user preferences, and system constraints. We use the history of previous adaptation operations to process new user requests on the assumption that an adaptation operator process different contents with similar features for the same user and the same system constraints will yield similar associated utility values. To achieve prediction, we therefore classify the adaptation situations using Bayesian multinomial logistic regression, allowing the system to select the adaptation operator with the maximal predicted utility value. Results obtained from experiments with real users demonstrate the validity of our model.

Introduction

Nous assistons présentement à une prolifération de dispositifs portables munis de caméra, d'écran, et de capacité de communication à travers des réseaux sans fil. Il en résulte une utilisation massive de ces équipements pour fabriquer, gérer et exploiter des documents visuels telles que les images et les videos. Par exemple, la quantité des documents non-textuels constitue 96% du World Wide Web [1].

En conséquence, une nouvelle problématique commence à paraître, en effet comment peut on permettre à tous les usagers un accés aux documents visuels de façon transparente à la diversité des équipements et à l'hétérogénéité des réseaux, ce concept connu sous le nom d'accès universel aux documents multimédia [2] [3] constitue l'objectif principal de l'adaptation des documents visuels. L'adaptation est nécessaire à cause de la diversité des équipements, des spécificités perceptibles des usagers, et des limites des réseaux de communications. Par exemple comment permettre à un usager ayant une acuité visuelle limitée de visionner une séquence vidéo sur un assistant personnel qui est connecté par le biais d'une connexion téléphonique à faible bande passante ? La nécessité d'assurer cet accès universel est primordiale dans plusieurs applications comme la médecine, la coordination opérationnelle lors d'une crise, la vidéo conférence et l'industrie de l'audiovisuel.

L'objectif principal de notre travail concerne la conception d'un système d'adaptation de documents visuels capable de répondre aux besoins des usagers selon leurs spécificités et leurs environnements d'usage.

L'adaptation soulève plusieurs problèmes. D'abord, l'identification des différents facteurs qui peuvent influencer le processus d'adaptation comme les caractéristiques de l'usager, la nature du contenu visuel, la bande passante disponible dans le réseau et les caractéristiques du dispositif utilisé par cet usager. Aussi, il faut déterminer la meilleure façon de considérer conjointement ces facteurs. En général, la plupart des systèmes d'adaptation existants se focalisent sur des cas particuliers d'adaptation comme contrôler le "bit rate" d'une vidéo en fonction des fluctuations de la connexion réseau disponible. La plupart des travaux précédents ne prennent pas en considération les besoins et les caractéristiques de l'usager [6], [7], [9]. Pour élaborer une formulation générale qui englobe tous les facteurs relatifs à l'adaptation ainsi que leurs descriptions, nous nous sommes inspirés de la partie 7 du standard MPEG21, intitulée Digital Item Adaptation [4], [5].

Dans le processus de l'adaptation, le contenu visuel subit plusieurs transformations par l'application des opérateurs d'adaptation. Un opérateur est une séquence de plusieurs algorithmes de traitement d'images et de videos numériques, comme la réduction de la résolution spatiale et la compression. Ces opérateurs modifient l'image ou la vidéo afin que les usagers puissent utiliser le contenu visuel dans des conditions satisfaisantes et en respectant les contraintes imposées par les caractéristiques de leurs équipements et leurs connexions respectives. La performance d'un système d'adaptation réside dans sa capacité à sélectionner l'opérateur adéquat à appliquer sur le contenu visuel pour chaque requête. Afin de prendre cette décision pour chaque situation d'adaptation, nous devons comparer les performances des différents opérateurs disponibles. Pour cela, nous reformulons le concept de fonction d'utilité [8] qui combine les critères suivants : la satisfaction des préférences et besoins visuels de l'usager, la fidélité au document original, le temps et le coût nécessaires pour que l'usager exploite le contenu visuel. Cette fonction décrit pour chaque opérateur d'adaptation la relation entre les caractéristiques de l'usager, les contraintes existantes et la valeur d'utilité résultante lorsqu'on applique cet opérateur. Nous impliquons les usagers dans le processus de l'adaptation durant une phase initiale. Ceci en collectant des notes reflétant leurs degrés de satisfaction pour chaque opération d'adaptation qu'on effectue. Pour la prédiction de la valeur d'utilité, nous définissons un modèle probabiliste qui utilise l'information présente dans l'historique des opérations d'adaptation précédentes pour répondre aux nouvelles requêtes des usagers. L'idée consiste à séparer les variables influençant l'adaptation en classes. A chaque classe sont associés les opérateurs qui s'exécutent avec succès. Pour valider ce modèle, nous avons effectué des expérimentations avec des utilisateurs réels. Dans le reste de ce mémoire nous détaillons le modèle proposé pour la selection automatique des opérateurs d'adaptation des images et des videos.

Adaptation des images et des videos pour des utilisateurs multiples dans des environnements hétérogènes

Dans ce chapitre, nous présentons le travail "*Image and Video Adaptation for Multiple Users in Heterogeneous Environments*". Ce travail concerne la sélection automatique des traitements nécessaires pour l'adaptation d'une image ou une vidéo. Cette sélection dépend de plusieurs facteurs que sont la nature et les caractéristiques du document visuel à adapter, les besoins de l'usager, les caractéristiques de son équipement, les performances de sa connexion et les conditions de son environnement naturel. La majorité des travaux existants ont abordé ce problème d'une façon partielle, c'est-à-dire ils se focalisent sur une tâche spécifique dans l'adaptation en considérant un sous ensemble des facteurs cités. Dans ce chapitre, nous donnons une formulation globale qui prend en considération les différents facteurs d'adaptation en se référant au standard MPEG21. Afin qu'on puisse sélectionner les opérations adéquates, nous utilisons le concept de la fonction d'utilité qui nous permet de comparer entre les performances des différentes opérations dans chaque situation d'adaptation. Nous définissons la fonction d'utilité par une formulation qui va permettre d'inclure plusieurs critères de qualité envisageables. Notre approche se base sur l'usager, ses besoins et ses préférences, parce que l'évaluation des documents adaptés est effectuée par des utilisateurs durant une phase initiale. Afin d'éviter de calculer cette fonction d'une façon analytique qui risque de ne pas refléter la satisfaction de l'usager à cause de la nature subjective de son jugement, nous proposons un modèle probabiliste basé sur la classification par régression bayesienne multinomiale logistique pour la prédiction de la valeur d'utilité. Ainsi, le modèle répond aux nouvelles requêtes des usagers en exploitant l'historique des opérations d'adaptation effectuées et l'évaluation fournie par les usagers.

L'idée de base du problème a été proposée par le Professeur Djemel Ziou et les recherches nécessaires à la modélisation ainsi qu'à la résolution des équations mathématiques et la validation des algorithmes ont été réalisées sous sa direction.

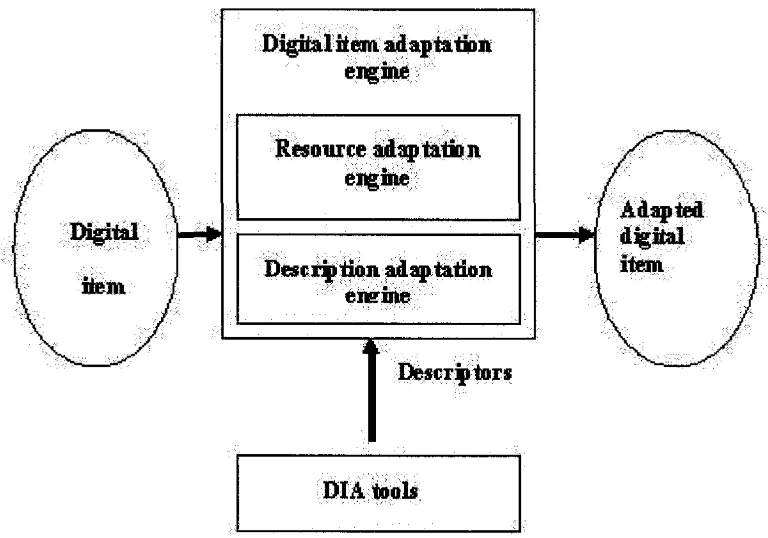

Figure 1: General architecture of MPEG21-DIA. Tools: Usage Environment Description (UED), Bitstream Syntax Description Language (BSDL), Adaptation QoS, Universal Constraints Description (UCD).

In [10], Vetro, one of the founding members of the MPEG21 standard, proposes a unified conceptual framework and technology taxonomy to handle the adaptation process while fitting with the general architecture of MPEG21-DIA. A set of definitions is introduced, beginning with the entity which is the atomic unit of media that can be subjected to modification or transformation and can be processed by an adaptation operator. Several entities exist on different levels, e.g., pixel, region, image, frame, or group of frames (GOF) in a compressed bitstream. The second definition establishes the adaptation operators formed by the set of image and video processing and coding algorithms such as spatial resolution modification in an image. Each operator can be executed using a single algorithm or a sequence of algorithms. We can define two operators with identical sets of algorithms but with different parameters. The set of operators form the adaptation space. The third definition concerns resources. Image or video is processed by adaptation operators with the goal of rendering it on a specific device and/or transmitting it over a network. These elements (device capabilities, network conditions and natural usage environment) constitute the resource space. Finally, we have the utility value, which is a target function that reflects all QoS related information such as video quality, user satisfaction and cost. The goal of the adaptation process is to optimize the value of this function with respect to constraints expressed in the resource space.

Let us now examine more thoroughly some existing work that has influenced us. In order to describe and analyze this work according to specific criteria, we will consider the usage environment descriptors put forward by digital item adaptation and the proposed conceptual framework as references.

Approaches	Wang and al [11]	Mohan and al [7]	Hsiao and al [9]
Processed media	MPEG-4 video-clips extracted from movies.	Web document image, video, and text items obtained by parsing TV news programs.	Video content MPEG7 test dataset.
Utility function	Learns from training pool and makes predictions by using classification and regression.	Calculates the distortion of each version by mean square error (MSE) and solves resource allocation problem.	Instead of evaluating utility function, bitstream allocation is performed according to the importance of each region of interest (ROI) identified by the attention model.
Content features	DCT coefficients and motion vector are extracted from compressed bitstream.	Content size, display size, color depth, streaming bitrate, and compression format.	Brightness, location, texture complexity, and motion are extracted from compressed video.
Results	Accuracy prediction of UF: 89%.	Not available.	72% satisfaction in subjective tests.
Adaptation operators	Frame dropping (FD) and, DCT coefficient dropping (CD).	Resolution reduction, color depth reduction, compression, transmoding of video to key-frames image.	Spatial size reduction and bitstream allocation.
User characteristics	No	No	No
Device capabilities	No	Screen size, image/video display capabilities.	Screen size and color depth.
Network conditions	Available bandwidth	Available bandwidth	Available bandwidth
Natural usage environment	No	No	No
Processing	Real-time	Static	Real- time

Table 1: Comparison of previous work on image and video adaptation

A quick look at Table 1 reveals that none of the three solutions consider user characteristics, especially preferences and visual impairments, as a criterion in calculating the utility function, either because user evaluation was simulated and thus limited [11] [9], or because an objective quality measure was used instead [7]. Moreover, the natural usage environment which describes the ambient user conditions at the moment of visualizing the adapted content, such as local luminosity and user mobility, was not considered in the three models. In [11], only video coding-related algorithms (frame dropping and coefficient dropping) were used, while the other two approaches used transmoding, which consists in changing the content from one mode to another, e.g., replacing a video sequence by a set of representative key frames. In [11] and [9], only videos were processed, while images adaptation was addressed in [7]. In [7], adaptation was performed in a static manner by preparing several versions of the content, which implies considerable storage space. On the other hand, the space of adaptation operators in [11] and [9] was limited because only online-feasible algorithms, e.g., frame dropping, were selected to satisfy the real-time constraint. In this paper, we attempt to design a general adaptation solution able to address the different challenges involved in prediction and improving user satisfaction, while handling all of the usage environment descriptors specified in MPEG21-DIA, and processing both video and images.

3 Problem Formulation

A variety of essential tools from different fields like image processing, human visual perception measurements, and decision theory frameworks such as machine learning are needed to guide the adaptation process. Content adaptation is a decision-making problem, so the most important segment in the analysis of the adaptation framework is quantitatively defining a target utility function to be optimized. We use this function to represent all quality of service related information and to compare the performances of different adaptation operators. Utility value can be formulated as a multidimensional functional over all parameters to be optimized, such as objective image/video quality assessment using signal level measurement algorithms like the PSNR or the mean square error (MSE). These methods have been extensively exploited because they are simple to calculate and easy to plug into mathematical optimization problems. However, they are not adequate to interpret human perception, because the distortion calculated by an objective quality metric may not match the perception of a human being and doesn't take into consideration human vision system features including foveal vision, light adaptation, and contrast sensitivity function [15]. More recent algorithms have been proposed like the structural similarity index (SSIM) algorithm [14] which can be used in its full-reference version to measure the similarity between the original content (image or video) in the input and the distorted content after applying an adaptation operator. This measure will be denoted by Q. The objective quality depends on the input document (I) and the adaptation operator (O). The second parameter is the subjective evaluation. This measurement is expensive and cannot be performed in realtime but reflects the real opinion of the end viewer (visual preference satisfaction, comprehension) expressed by the indices such as the mean opinion score, and user rating. We denote the user satisfaction by S. This quantity depends on the nature of the input content (I), user characteristics (U), device capabilities (D), and the user's natural usage environment characteristics (C). For instance, colors are not rendered in the same way on cathode-ray tube (CRT) and liquid crystal display (LCD) monitors, so the sensations are not similar; further, satisfaction will be different for a user with a color deficiency and another user with normal color vision. Most visual information is delivered over a network with a specific bandwidth, which imposes a waiting time on the user to upload a video or for initial buffer loading, in addition to the time needed to play the content in the case of streaming. We denote waiting time by T. The time consumed is a function of the input document (I), the adaptation operator (O), and network conditions (N); e.g., consistently available bandwidth will reduce the waiting time needed to upload a video. The final parameter is communication cost, because most mobile devices are connected over paying networks, so the size and the duration of a video is crucial in the adaptation process. We denote the cost by C. This amount varies depending on user location, which belongs to user characteristics (U), the size of the input document (I), and the adaptation operator (O).

Adaptation may be intended to achieve one of two goals. 1) Improving reproduction fidelity: in this situation the signal-level evaluation, which means objective quality will be prioritized, e.g., by performing a gamut mapping in order to match the range of colors available in a destination device. 2) Content customization to meet subjective user needs: user satisfaction will be the heavy criterion for utility value; e.g., hue modulation to satisfy the presentation preferences of a given user.

We can formulate the utility value as a function for each adaptation operation by several combinations of these criteria. We adopt the following formulation:

$$\max_{O} \quad Q(I,O) \tag{1}$$

$$\max_{O} \quad S(I,O,U,D,C) \tag{2}$$

under:

$$T(I,O,N) \leq T_{max} \tag{3}$$

$$C(I,O,U) \leq C_{max} \tag{4}$$

Thus, we can write:

$$\max_{O} \quad Q(I,O)S(I,O,U,D,C) \tag{5}$$

under:

$$T(I,O,N) \leq T_{max} \tag{6}$$

$$C(I,O,U) \leq C_{max} \tag{7}$$

The adapted content is the result of transformations performed by adaptation operators on the input document. So, to make predictions on the utility value, we will need some knowledge on the nature of the input content, different properties of the adaptation operators, and the existence of a correlation between these elements and the utility function components Q, S, T, and C. Adaptation operators modify image/video attributes in order to satisfy user preferences such as hue modulation, while matching resource limits, for instance by resizing the image to fit the display area of a PDA or adjusting video bitstream to a narrow connection bandwidth. These modifications of content are at different levels. The first level is the signal level, e.g., spatial resolution reduction of an image. The second is the structural level, e.g., bidirectional frame dropping in a video sequence. The last is the semantic level which uses the high-level semantic content layer to improve the user's comprehension, e.g., extracting an object in an image or a video. Many operations belong to two categories simultaneously; e.g., video summarization (semantic + structural).

For video, the principal constraint that has been extensively addressed is transmission of videos over networks by reducing the bit rate to fit a specific channel capacity. The set of these techniques is defined as video transcoding [20] which consists in converting an input signal to another in order to realize bitstream reduction and get a new spatial and/or temporal resolution.

Further, we can divide the adaptation space into two categories. First there are the obligatory operations, like converting the format of a video clip from AVI to WMV with the goal of playing this media on a device which supports only the latter format. The second category comprises enhancement operations like modulating the brightness or the hue of an image to satisfy a given user's preference.

In general, adapting media content will include many tasks like color requantization and brightness modulation. For each task there are many candidate algorithms that differ in complexity and performance, and thus, in results. This makes defining an adaptation space without ambiguity a difficult task. Adaptation task scheduling must be defined by the author, using prior knowledge in the image processing field to generate appropriate sequences for specific tasks: e.g., the compression operation must be executed after contrast enhancement and not before.

Like many applications in the imaging domain, such as image retrieval and video indexing, most adaptation solutions are content-based processes. To make decisions, we need input content features which will form the discriminant information between various contents. In our problem, the necessary features to extract will depend strongly on the adaptation space, which contains the set of image and video processing algorithms and their level; e.g., image size is needed to process a spatial resolution modification and edge magnitude is essential to a contrast enhancement algorithm, while high-level semantic features like regions of interest (ROI) or event-based features are required to perform video summarization. Previous work adopted the extraction of low level features from compressed data, such as spatial and texture information embedded in DCT coefficients, in order to meet real-time requirements. However, this was done by making a compromise between calculation simplicity and information accuracy; e.g., the motion vector in MPEG2 gives less texture information than optical flow. More details in compressed domain feature extraction are given in [19]. Further, this process remains dependent on the codec (e.g., MPEG, AVI). In practice, feature extraction from an image with a different format, like JPEG, or from a video, entails decoding the compressed bitstream, extracting the desired features and/or re-encoding the content. This process is infeasible in real-time applications because of the time required and the calculation complexity. To overcome this obstacle, the process can be significantly accelerated by extracting the desired characteristics of the content offline, and preparing them in the form of descriptive metadata which will accompany the visual document. Content features include several types of information, such as color, texture, shape, and regions or points of interest, and differ in their scope: global or local. These features can be represented by various forms including histograms, matrices, probability distribution functions (PDFs), and statistics such as mean and moments. These descriptors can be easily integrated into the MPEG21-DIA content descriptor structure.

The adaptation process must respond to user requests in real time. In addition to content feature extraction, which demands considerable time and calculation power, adaptation operators use algorithms with different performance and complexity. Some algorithms are too highly complex to calculate in real time, e.g., calculation of edge orientation and magnitude in an image or decoding of a video sequence. As a first solution to tackle this problem, we can run these algorithms offline and store the generated version of the input content, to be delivered for later requests [7] [27]. This requires consistent storage space. Another thing which, can help us to overcome the real-time constraint and improve process speed is the increasing programmability and computational power of programmable hardware modules. Some examples are the Field-Programmable Gate Array (FPGA) used in real-time face detection [30], and video feature tracking and matching using the Graphic Processing Unit (GPU) [29].

Among the components of the utility functional, the subjective evaluation reflects user satisfaction, which depends on user presentation preferences. The MPEG21-DIA therefore defines a set of descriptors to identify and quantify these preferences; for example, preferred brightness is defined as the mean of the Y-values in the YCbCr* color space, for all pixels in the image or the frame, uniformly quantified on the range [0,1] [3]. But in the real world, getting such information in an exact and explicit way is a very hard task. For this reason, we try to make inferences and group users who have similar preferences and profiles.

To make an adaptation decision engine capable of choosing the optimal adaptation operator that optimizes the utility functional, the three arguments Q, S, and T must be calculated for each combination of specific content, adaptation operator, user preferences and resource constraints. Most of the existing approaches perform offline calculation of the utility functional on a training pool and define the target functional by an analytical method [7] or by fitting curves on points formed in the adaptation space [22] .

4 Model

Our objective in this work is to perform adaptation for one or more users according to their needs and preferences. Given a set of visual documents, processing operators, and the associated utility value for each of them, the main idea is to select the optimal operator given a user request.

Based on the definition of the utility value and the different factors influencing this functional described in Section 3, we will give more details about data related to adaptation. The input content (I) is a representation of an image or video. So, I is a vector that contains all features related to spatio-temporal and radiometric resolution, color, shape, and texture. Each such feature can be described by integer or real values, e.g. the spatial resolution is $width X height$, and the color is described by a 3D RGB histogram. The user provides a subjective evaluation which depends on the specifities enumerated in MPEG21-DIA such as presentation preferences, and visual impairments as well as facts about the user such as: age, profession, and socio-cultural profile. The vector U includes all of these user attributes. For example, $U = (age,\ \ sex,\ \ profession,\ \ visual\ \ impairment, ...)$, where age is an ordinal variable, sex can be represented by a 0/1-valued variable, visual impairment is an ordinal variable indicating the absence of visual anomaly or the degree of severity of the deficiency otherwise. For instance, for a user who has a severe color deficiency, the value of the associated variable is equal to 0.9, while for another user with normal color vision, the value will be 0. The user displays the adapted content on a device D defined by specific device capabilities such as display area, color depth, and available codecs. The user's device is connected over network N, defined by available bandwidth and latency. Other features can be added, such as connection reliability. The environment surrounding the user and the device is described by a vector C which includes luminosity and location. By organizing all adaptation elements in this way, each user's request will be described by a *situation* vector denoted by X_i and formed by the conjunction of the different factors: $X_i = (U_i, I_i, D_i, N_i, C_i)$. Then, the adaptation engine will apply an adaptation operator that transforms the content I in order to meet the specific needs of user U_i who is in natural environment C_i, matching the constraints imposed by device capabilities D_i and network conditions N_i. For adaptation of a visual document I, one or more image processing or video processing algorithms are used, generating the new document: $I_{new} = O_i(I) = T_n(...T_2(T_1(I)))$. Adaptation operators are formed from different combinations of available algorithms. If we have N algorithms available, the number of different combinations, i.e. the number of possible adaptation operators will be $\sum_{i=1}^{N} P_N^i = \sum_{i=1}^{N} \frac{N!}{(N-i)!}$. For example, if N=3, the number of possible operators will be 15. Fortunately, the actual number of adaptation operators is less than $\sum_{i=1}^{N} P_N^i$ because algorithms must be run on well-defined order. For example, contrast enhancement of a video sequence that will be summarized in key frames is pointless. We will define a set of sequences of algorithms that are scheduled to perform a specific high-level task. For example, reduction of an image resolution sequence for a PDA connected to a paying network with narrow bandwidth is carried out by spatial resolution reduction, color depth reduction, and compression into format available on the destination device. Several algorithms and several parameters can be used. For example, bilinear interpolation, bicubic interpolation, and partial differential equation (PDE) interpolation can be all used. The performance of each of these algorithms is different and depends on visual document properties such as signal to noise Ratio (SNR) and on the parameter values used. We can make variations in the version of each algorithm using a limited subset of the algorithm's parameters, e.g., we can resize a given image by the same algorithm, such as bilinear interpolation, with different scales: 2x2, 4x4. Consequently, a sequence of algorithms for a specific task with specific values of parameters is considered as an adaptation operator. In other words, two adaptation operators can be formed by the same sequence of algorithms, but with different parameter values. The new visual document produced by applying an operator O_i for the situation $X_i = (U_i, I_i, D_i, N_i, C_i)$, is rated by the user to reflect his subjective satisfaction S. Additionally, the objective quality Q of the generated document is measured by using an algorithm such as SSIM. In order to calculate the waiting time T, we use the network bandwidth, the number of frames and the frame rate for a video sequence, and the file size for an image. The utility value $UV_i(X_i, O_i)$, which reflects the "goodness" of the selected adaptation operator O_i, can then be calculated according to equation (5).

In this way, the triplet $(X_i, O_i, UV(X_i, O_i))$ will be considered as an adaptation realization. The basic assumption behind the model is that, for a given user U_i, adaptation realizations that have similar *situation* vectors and are processed by the same operator will generate similar or close utility values; i.e., for two *situation* vectors $X_1 = (U_i, I_1, D_1, N_1, C_1), X_2 = (U_i, I_2, D_2, N_2, C_2)$, if $X_1 \simeq X_2$, then, for a given adaptation operator O_i we have $UV(X_1, O_i) \simeq UV(X_2, O_i)$. So, to select the optimal operator for a new request described by the *situation* vector $X_{request}$, we will retrieve in the history of the past of adaptation realizations, the one that has a similar *situation* vector: $X_{Previous} \simeq X_{request}$. Then, we choose the operator O^* that gives the maximal utility value $O^* = argmax \quad UV(X_{Previous}, O_i)$. However, this process will not be able to handle a new document which has never been adapted or a request from a new user. For that, we adopt collaboration, which involves dividing the history of previous adaptation realizations into different classes, each class containing observations with similar *situation* vectors. The clustering has the advantage that available data about previous adaptation realizations can be used to process future requests. For example, a historic adaptation of user U_1 can be used to processing a request from a new user U_2 who has similar profile, allowing us to avoid calculating the respective utility values for all adaptation operators for the same *situation*.

So, adaptation realizations that have similar *situation* vectors will belong to the same class. For a new request from a user U_i, the vector $X_{Request}$ is formed , and its appropriate class C_k is estimated. We denote by R_{U_i,O_j} the set of realizations processed by a given operator O_j for the user U_i and $X_i \in C_k$. To estimate the utility value associated with the operator O_j, we take the mean of the utility values associated with the elements of R_{U_i,O_j}. Figure 2 illustrates the classification of adaptation realizations based on *situation* vectors and the subsets R_{U_i,O_j} . Further, in order to take into consideration a possible modification in the user's preferences, we will take a mobile average of the utility values of the L last adaptation realizations in the history of U_i. Thus, the estimated utility value is given by:

$$\widetilde{UV}(X_{Request}, O_j) = \frac{\sum\limits_{m=N-L-1}^{N} UV_m(X_i, O_j)}{L} \quad where \quad X_i \in C_k \quad , \quad N = |R_{U_i,O_j}| \qquad (8)$$

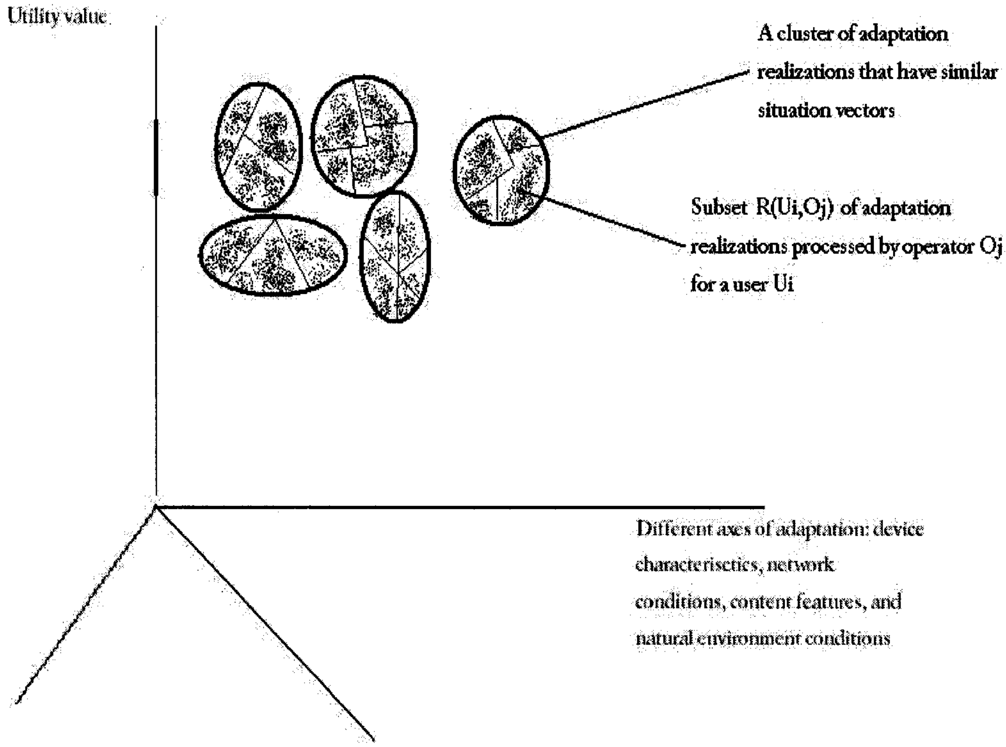

Figure 2: Classification of previous adaptation realizations

As initial approach, we put adaptation realizations in several spaces. Each space contains all adaptation realizations that were processed by the same operator which, means that each space is associated with an adaptation operator. This approach is very useful to study the conditions for application and feasibility of each operator individually. However, the number of operators will grow rapidly with variations in the parameters of available algorithms in the adaptation engine, and this will lead to a considerable number of spaces. In this case, the *situation* vectors must be clustered separately for every space in order to achieve prediction of the utility value, so, when the number of spaces is considerable, the model will be very complex and require more calculation power and time. We therefore choose to adopt one global adaptation space which groups all adaptation realizations regardless of the operator.

We will now deal with the clustering. Let us assume that there are K clusters. A K-dimensional Boolean vector $Y_i = (y_{i1}, y_{i2}, \ldots, y_{iK})$ is associated with the vector X_i. If X_i belongs to a given cluster l, then, $y_{ij} = 1$ if $j = l$ and 0 otherwise, and $\sum_{l=1}^{K} y_{il} = 1$.

The decision to assign a vector X_i to cluster j is given by:

$$y_{ij} = \begin{cases} 1 & if \quad j = argmax \quad P(X_i, \beta | y_j) \\ 0 & otherwise \end{cases} \tag{9}$$

Where β denotes the parameters of interest in the model adopted to perform the clustering. These models are of two types: generative and discriminative models. Their advantages and drawbacks are discussed in [23]. For our problem, the adaptation *situation* vectors are not labeled, so we propose to use a generative model that can estimate the missing data and a discriminative model to carry out the decision. To assign an observation to the appropriate cluster, we estimate the conditional probability that a vector belongs to class $p(y|x)$ using multinomial logistic regression, and benefit from the robustness in accuracy of this discriminative approach [13]. Multinomial logistic regression can be considered as a generalization of logistic regression, which describes the relationship between a set of explanatory variables in the input and a response variable (the outcome). In the case of logistic regression, the outcome is binary, e.g. success and failure, and its conditional distribution follows a Bernoulli distribution, while in the multinomial case, the conditional distribution of the output, which has more than two categories, follows a multinomial distribution. The goal of multinomial logistic regression is to separate m classes on the basis of an input vector X of the observed features of dimension d: $X = [x_1, \cdots, x_d]^T$. This vector includes explanatory variables which can be ordinal or nominal. The set of training data will be $D = \{(X_1, y_1), ..., (X_n, y_n)\}$. The conditional probability that an example X_i will belong to a class j is:

$$P(y_j = 1 | X_i, \beta) = \frac{\exp(\beta_j^T X_i)}{\sum_{k=1}^{m} \exp(\beta_k^T X_i)} \tag{10}$$

Where β^T is a (d+1) dimensional vector of weighting parameters be estimated. In the Bayesian framework, β is viewed as a random variable that has a prior distribution $P(\beta)$. Using Baye's rule, the posterior distribution over β is:

$$P(\beta, X_i|y_i) \propto P(y_i|\beta, X_i)P(\beta, X_i) \tag{11}$$

In general, the unknown parameters β and observations X are assumed to be independent, so:

$$P(X_i, \beta) = P(\beta)P(X_i|\beta) = P(\beta)P(X_i) \tag{12}$$

Assuming that observations are independent and contribute equally in the model, we can consider $P(X_i)$ as a uniform distribution.

When the prior on β is uniform or "flat", estimation of the parameters is achieved by maximizing the log-likelihood function using an optimization algorithm like Newton-Raphson or Fisher scoring [12]. However, with real data, the convergence of these algorithms to an optimal estimate is not guaranteed for the following reasons: (1) entire separability of classes which will give an infinite set of possible values of estimates β^*; (2) the presence of local maxima on the likelihood function doesn't allow us to obtain absolute optimal estimates [24]. To solve this problem, we adopt a univariate Gaussian prior $P(\beta)$ which has a unique mode. This will reshape the likelihood function and save us having to search outside of the parameter space or find local optima. The hyper-parameters are (0,σ), where σ is the variance vector formed by individual variances σ_{kj} of the parameters β_{kj}. Assuming that all components of β are independent, we have:

$$P(\beta_{kj}|\sigma_{kj}) \sim N(0, \sigma_{kj}) = \frac{1}{\sqrt{2\pi}\sigma_{kj}} \exp(\frac{-\beta_{kj}^2}{2\sigma_{kj}}) \tag{13}$$

Now, we estimate the maximum a posteriori (MAP) to find the mode of the posterior distribution:

$$\beta^*_{MAP} = argmax_\beta[\sum_{i=1}^{n} log(P(y_i|x_i, \beta)) + log(P(\beta)] \tag{14}$$

To estimate the number of clusters, many existing information criteria for selecting a model in cluster analysis, such as the Akaike Information Criterion (AIC), the Minimum Description Length (MDL), and Schwartz's Bayesian Inference Criterion (BIC) [31] can be used. These criteria associate the estimated maximum likelihood $L_n(k)$ of a given model with a sample of observations of size n and the number of parameters k, e.g., the AIC [32] is formulated as follows:

$$argmin \quad c_n(K) = -2\frac{ln(L_n(K))}{n} + 2\frac{K}{n} \tag{15}$$

where $ln(L_n(K)) = \sum_{i=1}^{n} log(P(y_i|x_i, \beta))$

As mentioned before, initially the data are incomplete. In order to group similar situation vectors in the same classes according to a specific measure, we perform unsupervised clustering using the fuzzy C means [25].

To conclude, the adaptation algorithm is summarized as follows:

LEARNING:

Run fuzzy C mean to assign data to components.

Estimate the MAP using equation (12).

Calculate AIC to estimate K using equation (13).

Calculate utility value for each realization using equation (5).

DECISION:

For a new request R from a user U_i, we form the input *situation* vector $X_R = (U_i, I_R, D_R, N_R, E_R)$

Calculate conditional probabilities $P(y_{Rj}|X_R) forj = 1, K$ using equation (8).

Assign X_i to the appropriate class C_k.

Find the utility value associated with each operator using equation (6).

Select the optimal operator that gives the maximal predicted utility value.

---

5 Experimental Results and Performance

This section illustrates the validation of our model with real data. In our experiment, we focalized on subjective user satisfaction. We designed a server able to deliver images for four users and videos for two users. During the learning stage, the adapted document is rated by the user using the score S that describes his subjective satisfaction. The score is expressed by the five choices: bad (1), average (2), good (3), very good (4), and excellent (5).

The content is selected from a training pool that contains 1000 images with different attributes in order to represent maximal variety in terms of content (size, color depth, texture, indoor images, natural scenes) and 400 video clips at the original bitrate of 1.5 Mbps, extracted from different movies. For images, the following features were extracted: color histogram in the CIE-L*a*b color space, which is perceptually uniform; image width and height; color depth; and texture, described by correlogram features of the pixel neighborhood with a displacement equal to 1 in the four orientations $\{0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}\}$. The correlogram features were mean, variance, entropy, contrast, energy, and homogeneity [26]. For video, we used the histogram of optical flow, to get texture characteristics [33]; frame width and height; and number of frames. These features were calculated offline from the decompressed video. For the adaptation operators, we defined four operators by organizing a set of operations in sequence to accomplish specific tasks; e.g., reduction of spatial resolution and modification of color depth of an image in order to fit on a PDA screen. For images, these individual operations were size modification, contrast enhancement, brightness modification, color quantification, and conversion to the JPEG format. The operators were as follows: operator 1 consisted of resizing the image to 800X600, and contrast enhancement. Operator 2 was resizing the image to 800X600, contrast enhancement, and brightness modulation. Operator 3 was resizing the image to 300X200, reduction of color depth to 16 bpp, contrast enhancement, and conversion to JPEG format. Finally, operator 4 was resizing the image to 300X200, reduction of color depth to 16 bpp, brightness modulation ,and conversion to JPEG format. These operators were executed using the Magick++ library. For video, the four adaptation operators were formed by combination of spatial resolution modification and bitrate reduction, with variation in the target bit rate: 1 Mbps, 500 Kbps, and 240 Kbps. These tasks were achieved using the Windows Media Encoder Software Development Kit (SDK). All adaptation operations are performed in real time. We used four devices with different characteristics: the Dell precision 380 work station, the HP Compaq nx6110 laptop, the HP iPAQ hw6940 PDA, and the Palm treo 700wx smart phone. Users received images and video clips on their devices over three networks: Université de Sherbrooke Local Area Network, AERIUS wireless network, and Internet. We began performing adaptation by applying operators. For each adaptation realization, we formed the *situation* vector X_i using content features I, user's device characteristics D_i: screen resolution, screen size, color depth, and refresh rate, and for network N_i: the available bandwidth. When about 10000 *situation* vectors had been collected from users over 8 weeks, we collected the scores S_i assigned by users after receiving and displaying the processed content. This allowed us to collect the triplets $(X_i, O_i, S(X_i, O_i))$. The regression algorithm was run to estimate the parameters β of the multinomial regression, and we found the number of clusters for images and videos: $K_{images} = 8$ and $K_{video} = 10$. We checked the accuracy of the classifier using Bayesian multinomial logistic regression on the data labeled by the fuzzy-C means algorithm. The accuracy of classification is shown by the confusion matrices in tables 2 and 3. In these confusion matrices, the cell $(classi, classj)$ represents the number of *situations* from *classi* that were classified as *classj*.

	class1	class2	class3	class4	class5	class6	class7	class8
class1	658	0	0	0	0	6	0	6
class2	0	1044	0	0	0	0	0	0
class3	0	0	624	0	0	0	0	0
class4	0	18	0	562	0	0	0	0
class5	0	0	0	0	888	0	0	0
class6	0	0	0	0	0	746	0	0
class7	0	0	0	6	0	0	600	0
class8	6	0	12	12	12	0	6	1136

Table 2: Confusion matrix for images. Global accuracy rate: 98.68%

	class1	class2	class3	class4	class5	class6	class7	class8	class9	class10
class1	150	0	0	0	0	0	0	0	0	0
class2	0	230	0	0	20	0	0	0	0	0
class3	0	24	904	0	0	0	0	0	0	0
class4	0	0	0	920	0	0	0	0	0	0
class5	0	0	0	0	766	0	0	0	0	0
class6	0	0	0	0	12	304	0	0	0	0
class7	0	0	0	18	0	0	60	0	0	0
class8	0	0	12	0	0	0	0	152	6	0
class9	0	0	0	0	0	0	0	0	304	0
class10	0	0	0	0	0	0	0	0	0	754

Table 3: Confusion matrix for videos. Global accuracy rate: 98.02%

After the training data set was clustered, we examined the distribution of user scores in each subset R_{U_i,O_j} of adaptation realizations inside each cluster. To illustrate these results, histograms of scores relative to each subset R_{U_i,O_j} in a given cluster are presented in figures 3 and 4.

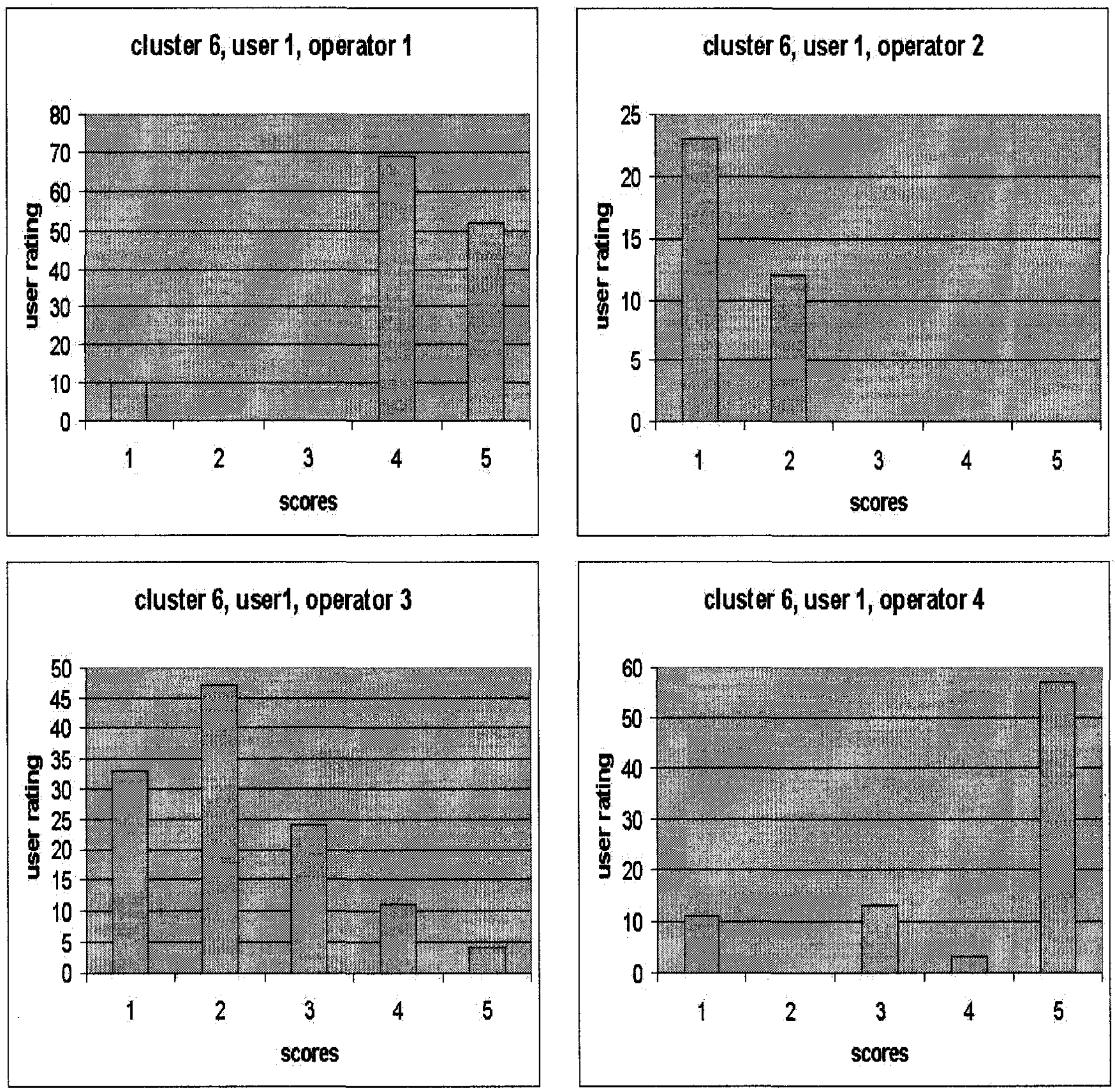

Figure 3: Score histograms for images

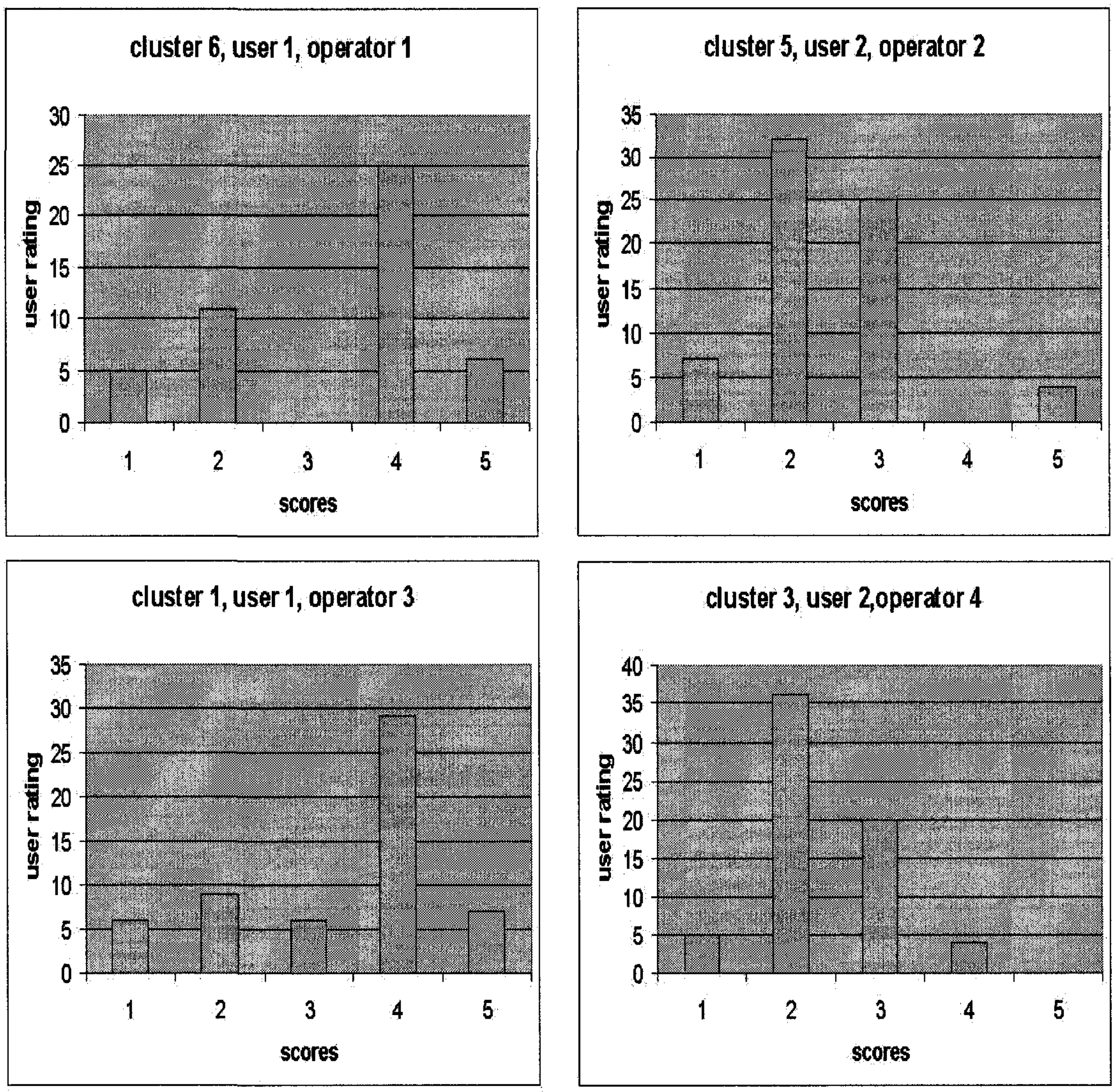

Figure 4: Score histograms for videos

These histograms show that most scores inside each subset of adaptation realizations for each user-operator pair (U_i, O_j) fall into one or two adjacent levels of satisfaction, e.g., "very good (4)", and "excellent (5)", or "bad (1)", and "average (2)". These results confirmed the consistency of the assumption underlying the classification approach behind our model. However, we observed the presence of some irrelevant or "noisy" values; e.g., in the score histogram for cluster 6, user 1, and operator 4 in Figure 4. Further, in some subsets the scores are concentrated in more than two levels; e.g., the histogram of scores in cluster 6, user 1, and operator 3 in Figure 3. This fact can be explained by the inconsistency in the subjective judgement given by the user, and misclassification cases.

The next step in our experiment sought to predict user scores for new requests. From a new test pool containing 200 images and 100 video clips, we processed the user's request by executing adaptation operators. Then, we assigned each *situation* vector to the appropriate class using our classifier. After that, we calculated the predicted score using equation (6) with $L = 10$ (recalling that L is the number of the last adaptation realizations in the subset R_{U_i,O_j}). As this stage of the experiment, the time needed to answer each user request is the total of: 1) the negotiation phase to identify user, device characteristics, and available bandwidth in the network; 2) decision making including prediction of the utility value and selection of the optimal operator; 3) time to execute the selected operator. Comparing the predicted and actual scores, we obtained the following prediction accuracy rates for the individual users: 87.67%, 93.71%, 85.76%, 90.60 % for images; and 87%, 89.32% for videos. We concluded that the prediction model is effective, and can be used to select optimal adaptation operators. At this stage, we performed operator selection by applying the operator that gave the highest predicted score for each request. Tables 4 and 5 show the scores given by users.

	bad	average	good	very good	excellent
user 1	1.42%	6.60%	0%	**23.82%**	**68.16%**
user 2	0%	3.03%	0%	**23.57%**	**73.40%**
user 3	1.30%	0%	3.93%	**37.70%**	**57.07%**
user 4	0%	5.80%	0%	**15.94%**	**78.26%**

Table 4: User scores for images

	bad	average	good	very good	excellent
user 1	0%	4.86%	3.62%	**41.39%**	**50.13%**
user 2	2.15%	0%	4.92	**38.14%**	**54.79%**

Table 5: User scores for videos

From these tables, if we consider that scores "very good" and "excellent" reflects an acceptable threshold of user's satisfaction, the global performance of our model in selection of the optimal adaptation operator is described by the following rates for individual users: 91.98%, 96.97%, 94.77%, and 94.20% for images; and: 91.52 %, 92.93% for videos at reasonable delays. The results confirmed the robustness of our model which uses Bayesian multinomial logistic regression to predict user satisfaction, and selection of the optimal operator online in order to achieve content adaptation automatically.

6 Conclusion

In this work, we devised a new adaptation framework that can handle both images and videos. The framework can be seen as an implementation that fit with the general architecture of MPEG21-DIA. Our formulation proved its capability to take into consideration all of the factors known to influence the adaptation process. Further, the system is open to all of the existing digital image and video processing algorithms. In order to take users's needs and characteristics into consideration, end users were involved at the learning stage. The probabilistic model using classification by Bayesian multinomial logistic regression proved its effectiveness in predicting utility value. We meet the real-time constraint in prediction of the utility value and selection of the optimal adaptation operator. By using the history of previous adaptation operations, the system is able to deal with new situations including device variety, multiple user profiles, and new image and video content. We validated the proposed model by experiments performed with real users in different environments. The results obtained were promising. However, several aspects remain to be explored, such as assignment of new users to existing profiles, features selection to improve the performance of the classifier, and extending the solution to more adaptation operators.

References

[1] A. Perkis, Y. Abdeljaoued, C. Christopoulos, T. Ebrahimi, and J.F. Chicharo. "Universal multimedia access from wired and wireless systems". Birkhauser Boston Trans. on Circuits, Systems and Signal Processing, vol. 10, no. 3, pp. 387-402, 2001.

[2] A. Vetro, T. Christopoulos, and T. Ebrahimi. "Special issue on universal multimedia access". In IEEE Signal Processing Magazine, vol. 20, no. 2, March 2003.

[3] MPEG MDS Group, MPEG-21 Multimedia Framework, Part 7: Digital Item Adaptation (Final Committee Draft), ISO/MPEG N5845. July 2003.

[4] A. Vetro and C. Timmerer. "Digital item adaptation: Overview of standardization and research activities". IEEE Transactions on Multimedia, vol. 7, no. 3, pp. 418-426, June 2005.

[5] G.M Schuster and A.K. Katsaggelos. *Rate-distortion based video compression.* Kluwer Academic Publishers.1997.

[6] A. Vetro, Y. Wangz, and H. Sun. "Rate-distortion optimized video coding considering frameskip". Proceedings in International Conference on Image Processing (ICIP2001), vol. 3, pp. 534-537. 2001

[7] R. Mohan, J. R. Smith, and C. Sheng. "Adapting multimedia internet content for universal access". IEEE Transactions on Multimedia, vol. 1, no. 1, pp. 104-114. March 1999.

[8] D. Mukherjee, E. Delfosse, and Y. Wang. "Optimal adaptation decision-taking for terminal and network quality of service". IEEE Transactions on Multimedia, vol. 7, no. 3, pp. 454-462. June 2005.

[9] M. H. Hsiao, Y. W. Chen, K. H. Chou, and S. Y. Lee. "Content-Aware video adaptation under low-bitrate constraint". EURASIP Journal on Advances in Signal Processing vol. 2007. 2007.

[10] S. F. Chang and A. Vetro. "Video adaptation: concepts, technologies, and open issues". Proceedings of IEEE, vol. 93, no. 1, pp. 148-158, January 2005.

[11] Y. Wang, J. G. Kim, S. F. Chang, and H. M. Kim. "Utility based video adaptation for universal multimedia access (UMA) and content based utility function for real-time transcoding". IEEE Transactions on Multimedia, vol. 9, no. 2, pp. 213-220, February 2007.

[12] A. Schworer and P. Hovey. "Newton-Raphson Versus Fisher Scoring Algorithms in Calculating Maximum Likelihood Estimates". Electronic Proceedings of Undergraduate Mathematics Day, no. 1, pp. 1-11, University of Dayton, OH, USA. 2004.

[13] D. W. Hosmer and S. Lemeshow. *Applied logistic regression.* Wiley Series in Probability and Mathematical Statistics. 1989.

[14] Z. Wang, A. C. Bovik, H. R. Şeikh, and E. P. Simoncelli. "Image Quality Assesment: from error visibility to structural similarity". IEEE Transactions on Image Processing, vol. 13, no.4, pp. 600-612, April 2004

[15] Z. Wang. "Objective image /video quality measurement: A literature survey". For EE381K: Multidimensional Digital Signal Processing. University of Texas at Austin. October 1998.

[16] N. Bjork and C. Christopoulos. "Video transcoding for universal multimedia access". Proceedings of the 2000 ACM Workshops on Multimedia, pp. 75-79, Los Angeles, California, United States, 2000.

[17] Y. Wang, J. G. Kim, and S. F. Chang. "Content-based utility function prediction for real-time MPEG-4 video transcoding". Proceedings of International Conference on Image Processing (ICIP2003), vol. 1, pp. 14-17, September 2003.

[18] B. Shen and I. K. Sethi. "Direct feature extraction from compressed images". SPIE, vol. 2670, Storage & Retreival for Image and Video Databases IV, 1996.

[19] H. Wang, A. Divakaran, A. Vetro, S. F. Chang, and H. Sun. "Survey of Compressed-Domain Features used in Audio- Visual Indexing and Analysis. Journal of Visual Communication and Image Representation, vol. 14, no. 2, pp. 150-183, June 2003

[20] A. Vetro, C. Christopoulos, and H. Sun. "Video transcoding architectures and techniques: An overview". IEEE Signal Processing Magazine, vol. 20, no. 2, pp 18-29, March 2003.

[21] W. Lai, X. D. Gu, R.H. Wang, W. Y. Ma, and H. J. Zhang. "A content based bit allocation model for video streaming". IEEE International Conference on Multimedia, vol. 2, pp. 1315-1318, June 2004.

[22] E. C. Reed and J. S. Lim. "Optimal multidimensional bit-rate control for video communication". IEEE Transactions on Image Processing, vol. 11, no. 8, pp. 873-885, August 2002.

[23] Y. Ng. Andrew and M. I. Jordan. "On discriminative vs. generative classifiers: A comparaison of logistic regression and naive Bayes". Advances in Neural Information Processing Systems, vol. 14, 2002.

[24] K. Riadh, D. Ziou, B. Colin, and F. Dubeau. "Weighted Pseudo-Metric Discriminatory Power Improvement Using a Bayesian Logistic Regression Model Based on a Variational Method". IEEE Transaction on Pattern Analysis and Machine Learning, In press. March 2007.

[25] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern classification.* Second edition. Wiley Interscience. 2001.

[26] M. S. Allili and D. Ziou. "Automatic color-texture image segmentation by using active contours". In Proceedings of Internatianl Workshop on Intelligent Computing in Pattern Analysis and Synthesis, pp. 495-504. 2006.

[27] S. Kopf and W. Effelsberg. "Color adaptation of videos for mobile devices". Proceedings of ACM International Conference on Multimedia. pp 963-964. Santa Barbara, CA, USA. 2006.

[28] H. S. Neoh and A. Hazanchuk. "Adaptive Edge Detection for Real-Time Video Processing using FPGAs". Global Signal Processing Conference, 2004.

[29] S. N. Sinha, J. M. Frahm, M. Pollefeys, and Y. Genc. "GPU-based Video Feature Tracking And Matching". Technical Report (TR) 06-012. Department of Computer Science, UNC Chapel Hill. May 2006.

[30] D. Nguyen, D. Halupka, P. Aarabi, and A. Sheikholeslami. "Real-Time Face Detection and Lip Feature Extraction Using Field-Programmable Gate Arrays". IEEE Transactions on Systems, Man, and Cybernetics - Part B: Cybernetics, vol. 36, no. 4, pp. 902-912. August 2006.

[31] N. Bouguila and D. Ziou. "Unsupervised learning of a finite mixture model based on the Dirichlet distribution and its application. IEEE Transactions on Image Processing, vol. 13, no. 11, pp. 1533-1543. November 2004

[32] H. Akaike: "A New Look at the Statistical Model Identification". IEEE Transactions on Automatic Control, vol. 19, no. 6, pp. 716-723. 1974.

[33] B. K. P. Horn and B. G. Shunk. "Determining optical flow". Artificial Intelligence, vol. 17, pp. 185-203. 1981

Conclusion

Dans ce travail, nous avons proposé une solution pour l'adaptation des images et des vidéos pour différents utilisateurs dans des environnements hétérogènes. La conception de cette solution soulève plusieurs problèmes tels que la formulation nécessaire pour considérer conjointement tous les éléments liés au processus de l'adaptation, la sélection des algorithmes de traitement d'images et des vidéos numériques qu'on doit appliquer sur le document visuel afin de satisfaire l'usager en respectant les contraintes imposées par sa configuration matérielle et sa connexion réseau. L'étude des travaux existants nous a permis de dégager les points primordiaux pour un système d'adaptation efficace. En premier, l'exploitation d'un schéma général pour représenter et décrire tous les facteurs relatifs à l'adaptation et leurs relations mutuelles. Pour cela nous nous sommes inspirés de la partie 7 intitulée Digital Item Adaptation du standard MPEG21. Ensuite, nous avons utilisé le concept d'utilité qui mesure la performance et l'efficacité de chaque opération d'adaptation. Nous avons donné une nouvelle formulation à la fonction d'utilité en combinant tous les critères relatifs à la qualité du document adapté qui sont la satisfaction des besoins de l'usager et ses préférences, la fidélité au document visuel original, le temps et le coût conséquent nécessaires pour consommer le contenu visuel par l'usager. Afin de prendre en considération les besoins et les préférences qui diffèrent d'un usager à un autre, nous avons impliqué les usagers dans l'évaluation des opérations d'adaptation, c'est-à- dire chaque réponse à une requête est notée avec un degré de satisfaction. Cette formulation du problème nous a permis de définir un modèle probabiliste basé sur la classification par la régression bayesienne multinomiale logistique afin que nous puissions sélectionner les traitements adéquats à appliquer pour une nouvelle requête. Le modèle proposé a été validé par la réalisation d'une plate-forme d'adaptation avec des usagers réels. Les résultats de cette expérimentation étaient concluants. Cependant, plusieurs aspects restent à explorer comme la gestion des profils des usagers, par exemple l'affectation des nouveaux usagers à des profils existants, la sélection des caractéristiques pertinentes pour augmenter le taux de précision de la prédiction et enrichir la solution en incluant d'autres opérateurs d'adaptation.

Bibliographie

[1] P. Lyman, and H. R. Varian : How much information. Http :www.sims.berkeley.edu/ research/ projects/ how-much-info-2003/. University of California at Berkeley. 2003

[2] A. Perkis, Y. Abdeljaoued, C. Christopoulos, T. Ebrahimi, and J.F. Chicharo : Universal multimedia access from wired and wireless systems. Birkhauser Boston trans. on Circuits, Systems and Signal Processing, vol. 10, no. 3, pp. 387 to 402, 2001.

[3] A. Vetro, T. Christopoulos, and T. Ebrahimi : Special issue on universal multimedia access. In IEEE Signal Processing Magazine, vol. 20, no. 2, March 2003.

[4] MPEG MDS Group, MPEG-21 Multimedia Framework, Part 7 : Digital Item Adaptation (Final Committee Draft), ISO/MPEG N5845. July 2003.

[5] A. Vetro, and C. Timmerer : Digital item adaptation : Overview of standardization and research activities. IEEE transactions on multimedia, vol. 7, no. 3, pp. 418 to 426, June 2005.

[6] R. Mohan, J. R. Smith, and C. Sheng : Adapting multimedia internet content for universal access. IEEE transaction on multimedia, vol. 1, no. 1, pp. 104 to 114. March 1999.

[7] M. H. Hsiao, Y. W. Chen, K. H. Chou, and S. Y. Lee : Content-Aware video adaptation under low-bitrate constraint. EURASIP Journal on Advances in Signal Processing vol. 2007. 2007.

[8] S. F. Chang, and A. Vetro : Video adaptation : concepts, technologies , and open issues. Proceedings of IEEE, vol. 93, no. 1, pp. 148-158, January 2005.

[9] Y. Wang, J. G. Kim, S. F. Chang, and H. M. Kim : Utility based Video adaptation for universal multimedia access (UMA) and content based utility function for real-time transcoding. IEEE transactions on multimedia, vol. 9, no. 2, pp. 213 to 220, February 2007.