

散在する多視点ビデオ映像を用いたユビキタスビデオ網構想

○市村 哲; 宇田 隆哉; 伊藤 雅仁; 田胡 和哉; 松下 温
東京工科大学

我々は、「実習」を含む教育をリモート環境または独学において可能とすることを目的とした研究を実施しており、この一環として、「MultiVideoPlayer」(略して「MVP」)を開発した。MVPは、複数のビデオカメラを用いて機器操作映像等を多視点撮影し、その多視点映像を同期再生するシステムであるが、ビデオ撮影時の制約が非常に少ないのが特徴である。本稿では、MVPの機能を拡張し、ビデオカメラを持参した人の間で、アドホックにユビキタスビデオ網を構成し、いつでも、どこでも、誰でも、簡単に多面的ビデオ映像の撮影および共有利用ができるようするための方法について述べる。

A Design for an Adhoc Video Network of Multiple Camcorders

○Satoshi Ichimura; Ryuya Uda; Masahito Ito; Kazuya Tago; Yutaka Matsushita
Tokyo University of Technology

Our study has focused on the creation of distance training environment in which people can learn physical operations, such as machine operations or surgical operations, in a remote or asynchronous manner. As a part of the study, we have developed MultiVideoPlayer (MVP for short), a software playing back multiple videos synchronously, each of which is spontaneously videotaped by each different video camcorder. In this paper, we discuss the possible applications of the adhoc video network based on the use of the MVP, allowing multiple users to easily share each user's video.

1. はじめに

我々は、リモートで行う実習教育を「ディスタンストレーニング」と呼び、そのための支援システムの研究を実施してきた。従来は集合教育でしかできないと考えられていた「実習」を含む教育を、リモート環境または独学において可能とすることを目的としている¹。実習を支援するためには、詳細な機器操作映像や、その映像についての解説などを学習者に正確に伝達することが重要である。そのためには、「見たい箇所を多角度から見

られるコンテンツ」、「各部の動きが正しく把握しやすいコンテンツ」などが望まれる。

この研究の一環として、民生用ビデオカメラを複数用いて機器操作映像等を多視点撮影し、その多視点映像を同期再生するシステム「MultiVideoPlayer」(略して「MVP」)の開発を行った。MVPは、複数のデジタルビデオカメラで別々に撮影したビデオ映像中の映像撮影情報を解析してそれら複数のビデオ映像の関連性を自動検出する機能を有しており、別々に撮影した複

数のビデオ映像から関連のある映像箇所を抽出して同期再生できる。

我々は、以上に述べた MVP を、遠隔実習支援以外の用途にも対応できるよう拡張する方法について検討しており、特に本稿では、ユビキタス情報環境の実現に向けたアプローチを議論する。ビデオカメラは、実世界を仮想世界に取り込むための一番身近なツールの一つである。そもそも、ユビキタスコンピューティングというコンセプトは、「実世界と仮想世界とをネットワークによって接合し、これによってコンピュータを意識することなく情報を扱うようにすること」を目指したものであり、ビデオカメラを散在させるというアプローチはユビキタスコンセプトとも親和性が高い。MVP を用いれば、撮影時の制約がほとんどないため、ビデオを持ち歩きさえすれば、いつでも、どこでも、誰でも、簡単に多面的なビデオ映像を作成することができ、ユビキタス情報環境の構築が促進されると考える。

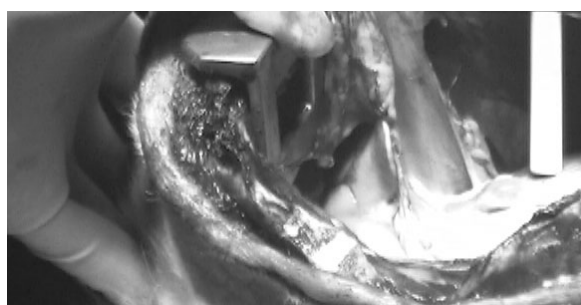
本稿では、MVP 開発の背景となった遠隔実習支援について述べ、本システムが提供する機能について説明する。さらに、ビデオカメラを持参した人の間で、アドホックにユビキタスビデオ網を構成する方法について述べ、ユビキタス情報環境の実現に向けてのシステム拡張可能性について議論する。

2. 遠隔実習支援

2.1. 要求分析

常時接続とブロードバンドの浸透に伴い、職場・学校・公共施設においてインターネットサービスを利用できる環境が整いつつある。このような状況下、E-ラーニングが、ブロードバンドを用いたアプリケーションの一つとして大いに期待を集

めている。しかしながら、現在効果が実証されている E-ラーニングは、マナー研修、英語学習、座学講義など、従来からテキストとビデオテープによる独学環境が存在したような分野に関するものに限られている。これらの E-ラーニングシステムは、講師映像と講義スライド資料とを同期させてインターネットを介して受講者に配信するというものがほとんどである。



臨床工学専攻科（日本工学院専門学校にて撮影）



電気工専科（日本工学院専門学校にて撮影）

図 1 専門学校における実習授業風景

例えば専門学校では、実習を主体とした職業訓練が授業の中で実施されている。このような授業は、講義 30%、実習 70%の割合で構成されるような場合が多く、実習が重視されている。しかしながら、このような実習中心の授業を対象とした場合、講師映像と講義スライド資料とを同期させて配信する従来の E-ラーニングシステムがカバーできるのは 30%の講義部分に関してのみであり、残りの 70%の実習部分に関してはほとんど支援手段とはなり得ないのは明白である。

例えば、日本工学院専門学校（著者らが所属する大学と同一学校法人に属する専門学校）の臨床工学専攻科や理学療法学科においては、日進月歩で進化する手術用医療機器や、実際のリハビリテーションに利用される治療器を日常的に用いて学生に実習させている（図 1 参照）。また、電気工学科や土木建設科では、現場で必要な操作スキルを高めるため、実際に重機を操縦するというような訓練が重視されている（図 1 参照）。しかしながら、講師映像と講義スライド資料とを同期させて配信するという従来の E-ラーニングシステムは、以上のような実習主体の教育においてほとんど支援手段とはなり得ない。

実習を支援するためには、詳細な機器操作映像や、その映像についての解説などを学習者に正確に伝達することが重要である。例えば、「見たい箇所を多角度から見られるコンテンツ」、「各部の動きが正しく把握しやすいコンテンツ」、「全体と細部との対応がわかりやすいコンテンツ」などが望まれる。そのためには我々は、単体のビデオ映像を再生できるようにするだけでなく、多面的にビデオ映像を撮影し、必要なコンテキストでそれら複数ビデオ映像を整理して提示できる機能が必要であると考えた。

2.2. 従来技術の問題点

同じ対象物を同時刻に撮影した映像であっても、複数のビデオカメラで別々に撮影した映像には関連性がなく、再生するにはそれぞれの映像を独立して再生するしか方法がないのが普通である。同時刻に撮影した複数映像を探し出すためには、複数のテープから早送りや巻き戻しをして探し出す必要があり、非常に手間がかかる。このように、複数の人がそれぞれ別のビデオカメラで撮影した場合、映像が別々のビデオテープに記録さ

れるため、複数の映像を同時に再生することが極めて難しいという問題があった。

そこで従来、多面的なビデオ映像を作成するためには、固定カメラを複数台設置し、同時に記録スタートまたは記録終了させるための特別な装置を付加して用いることが必要となっていた。しかしながら、手元の拡大映像が撮影できない、機器や作業者の陰になる部分の撮影が困難、作業者の視点の映像を撮影できないなどの問題や、ビデオ撮影システムが大規模化してしまい可搬性が損なわれるという問題があった。さらに、システムの組み立て、配線、テープの頭だしなどの撮影準備に膨大な手間がかかるため、いつでも簡単に利用できるというシステムではなかった。

複数の手持ちビデオカメラを使って撮影し、後からビデオ編集で複数ビデオ映像間の同期を合わせることによって多視点映像を作成するという方法も考えられるが、実際は、同時刻に撮影した映像を複数ビデオから探すためには、シーンの早送りや巻き戻しを繰り返したり、撮影カットの切れ目を、目で見て判断したりする必要があり、非常に手間がかかる²。

また、タイムコード編集可能なノンリニア編集器を用いた場合においても、撮影を開始するタイミングおよび撮影を終了するタイミングをそのビデオカメラの撮影者が独自に決定するような場合には、ビデオテープのタイムコードと撮影時刻とを一致させることができないという問題がある。また、複数のカットを連続して撮影する場合、いちいちカット毎にビデオテープを最初に巻き戻すようなことはせず、ビデオテープの続きから撮影することが普通であるため、この点からも、ビデオテープのタイムコードと撮影時刻とを一致させることが難しい。このため従来は、ビデオ撮影する度に、撮影内容と、撮影日時と、タイム

コードとの対応関係を紙に書き留めておき、ビデオテープと一緒にこの紙を管理するというのが普通となっている。

3. MVP (MultiVideoPlayer)

MVP は、複数のデジタルビデオカメラで別々に撮影したビデオ映像中の映像撮影情報を解析してそれら複数のビデオ映像の関連性を自動検出する機能を有している。これは、後述する「同一時刻撮影映像の同期再生機能」によって実現されている。この方法を基本として試作システムは、別々に撮影した複数のビデオ映像から関連のある映像箇所を抽出して同期再生する機能を提供している。

MVP を用いれば、多視点映像を作成する場合でも以下のようにビデオ撮影時の制約が極めて少なくなる。

1. 民生用 DV カメラのみを使用し外部制御装置が不要
2. 手持ちカメラを利用でき、撮影位置を固定されない
3. 各人が自由なタイミングで撮影の開始・終了が可能
4. ビデオテープの途中から撮影開始が可能
5. 複数ビデオテープにまたがった撮影が可能

3.1. 同一時刻撮影映像の自動検出

本方式は、複数のビデオ映像から、同一時刻に撮

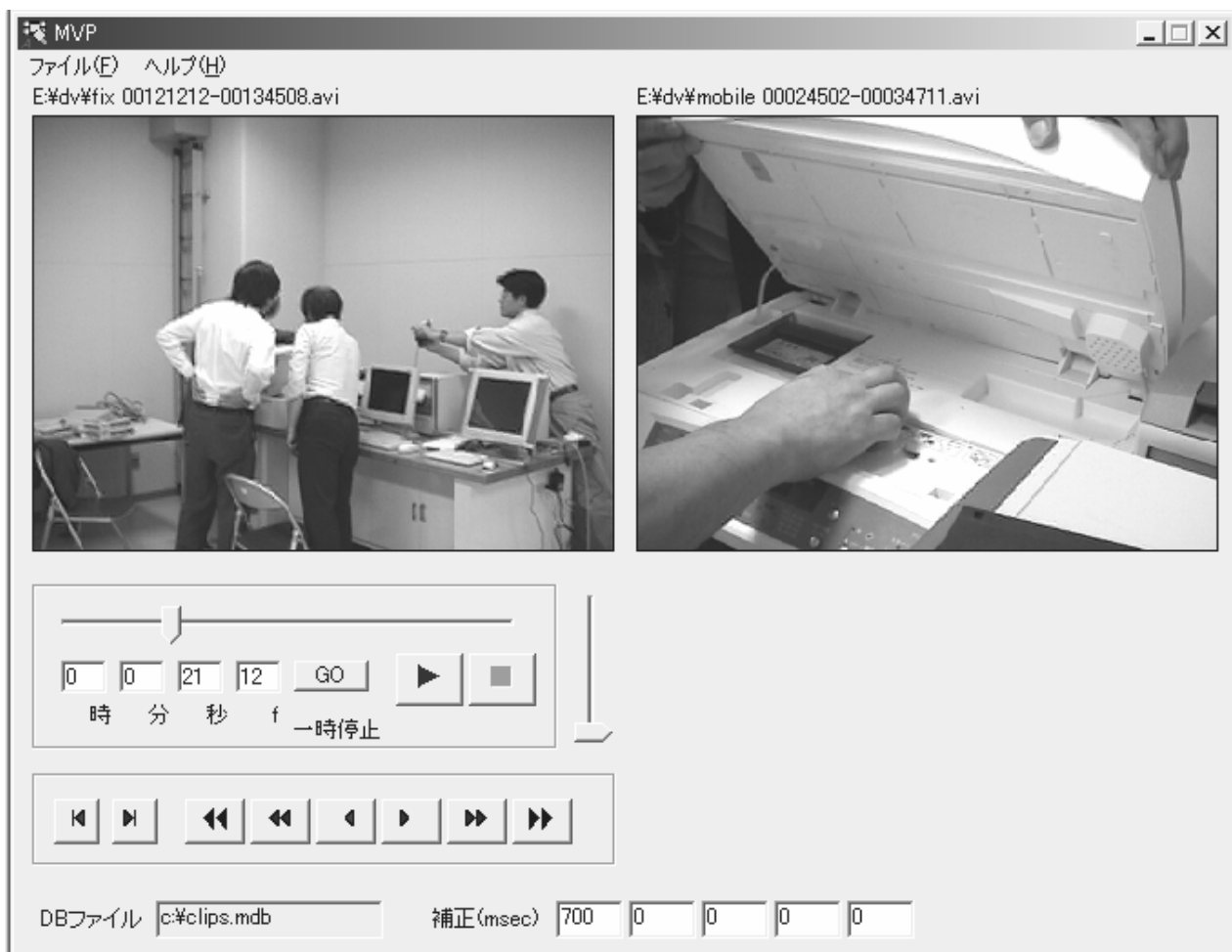


図 2 MVP(MultiVideoPlayer)による複数映像の同期再生

影された映像を自動検出する技術が基本となっている。具体的には、DV カメラで撮影されているライブ映像、DV テープに記憶されたデジタル映像、または、DV キャプチャされた映像ファイルから、その映像が撮影された日時情報を取得し、それに基づいて関連映像の解析を行う。

DV カメラの規格である ISO/IEC 61834 規格³では、撮影日付と撮影時刻とをすべての動画フレームのサブコードエリアに対して埋め込むことが規定されており、本システムはこの情報を利用している。試作システムの実装では、DV テープに撮影された映像は、当該デジタルビデオカメラは IEEE1394 規格のケーブルによって PC と接続され、デジタルビデオキャプチャ（以下、DV キャプチャ）されて PC の動画ファイルとして保存できる。

一般にはあまり知られていないが、DV キャプチャして作成された動画ファイルにも、DV テープに記録されている撮影日時情報がコピーして保存されており、DV キャプチャされたすべての動画フレームの中に、撮影日付情報と撮影時刻情報が記録されている。本試作システムは、Windows オペレーティングシステムの DirectShow の仕組みによって、DV キャプチャによって作成された動画ファイルの中から、動画フレームを 1 フレームずつ抽出し、撮影日付情報と撮影時刻情報を取得するようになっている。すなわち、撮影に用いる DV カメラの内部時計が標準時刻に合わせてありさえすれば、DV テープまたは DV キャプチャファイルに記録された映像からその撮影日時を取得し自動的に利用する。

撮影時刻情報を解析するために要する時間は、CPU Pentium4 1.6GHz の PC を用いて処理した場合、10 分長の動画ファイルに対して約 1 分である。処理時間は CPU 処理能力に左右されるが、最近の市販 PC を利用する限りにおいては、通常再

生の 1/10 以下の時間で解析が終了することを確認している。

3.2. 同一時刻撮影映像の同期再生

図 2は、MVP のユーザインタフェースを示している。左側のパネルにはメイン映像が、右側のパネルにはメイン映像に同期したサブ映像が適宜表示される。映像を再生する場合、ユーザは、メイン映像として再生すべき動画ファイルを、ファイルダイアログやドラックアンドドロップ等で指定する。ここで、撮影者 A が撮影した動画ファイル A をメイン映像として指定した場合には、動画再生部の左側のパネルに動画ファイル A の映像が再生表示されるが、この時、システムが、動画ファイル A と同じ時刻に撮影された映像が他の動画ファイルに記録されていることを検出した場合には、右側のパネルにその検出した動画ファイルをサブ映像として同時再生する。ここで、動画ファイル B および動画ファイル C に、動画ファイル A と同一時刻に撮影された映像が含まれていたとすると、この動画ファイル B および動画ファイル C が検出され、サブ映像として同時再生されることになる（理論上、同時再生するサブ映像数に上限はない）。

この時、動画ファイルの再生箇所の撮影時刻が同一となるように、すべての動画ファイルの再生タイミングをシンクロさせて動画再生する。すなわち、メイン映像である動画ファイル A が、撮影時刻 T1 に撮影した映像を再生している時には、他の動画ファイル B および動画ファイル C も撮影時刻 T1 に撮影した映像を再生する。このとき音声は、動画ファイル A の音声のみ再生し、動画ファイル B および動画ファイル C に記録された音声は再生しないようになっている。なお、撮影に用いた DV カメラの内部時計が多少ずれていた場合に対処するため、その誤差をユーザが補正するため

の GUI も提供している。

一方、撮影者 B が撮影した動画ファイル B をメイン映像として指定することも可能である。この場合には、動画再生部の左側のパネルに動画ファイル B の映像が再生表示され、右側のパネルには、動画ファイル A および動画ファイル C 中の、同時刻に撮影された映像部分が再生表示される。

なお、本システムでは、DV キャプチャされたビデオ映像を解析してインデックス情報を予め作成するインデクシングソフトウェアを提供している。このソフトウェアは、DV キャプチャしたビデオ映像中の撮影日時情報を解析することによって、撮影カット（DV カメラの映像記録開始ボタンが押されてから次に記録停止ボタンが押されるまでの区間の映像）毎の撮影時刻インデックス（撮影カット OP は T_o から T_p までの映像、撮影カット QR は T_q から T_r までの映像、というようなインデックス情報）を作成する。これによって、システムは、映像再生時に、同期場面を他の動画ファイルから高速に検索できる。

また、付加機能として、メイン映像である動画ファイル A の再生が終了した場合には、それに連動して再生している全ての動画ファイルの再生を終了させる機能を有している。加えて、動画ファイル A を一時停止、スロー再生、コマ送り再生、または、早送り再生した際に、動画ファイル A の再生モードと同じモードで全ての他の動画ファイルを再生するようになっている。

4. ユビキタスビデオ網構想

4.1. アドホックビデオ網

近年、デジタルビデオカメラは小型化・軽量化しており、運動会の撮影、旅行先での撮影など、極

めて幅広い分野で用いられている。例えば、小学校の運動会では、数多くの親がビデオカメラで自分の子供の姿を撮影する光景が見られる。運動会では、自分が陣取った場所にビデオカメラを設置して撮影し、運動会の記録ビデオを作成するのが普通である。この場合、運動会の記録ビデオとして自分の手元に残るのは、当然、自分が撮影した映像だけである。

しかしながら、このような、ビデオカメラを 1 台しか用いず、さらに、そのビデオカメラの設置場所を動かすことができないという環境では、撮影したシーンの多くは必ずしも最適な映像でないのが普通である。特に、トラックの徒競走を撮影するような場合には、被写体である子供達の位置や進行方向が大きく変化するため、自分が陣取った場所からの撮影映像だけでは満足な映像を得られない。例えば、スタート付近、または、ゴール付近は、スタート前の緊張した表情を収める、または、ゴールの瞬間を収めることができるということで撮影ポイントとして人気が高い。しかしながら、スタート地点とゴール地点とは位置的に離れていることが普通であり、スタート前の映像とゴールの瞬間の両方を、絶好の位置から撮影するというようなことは事実上不可能である。

そこで、運動会のように多くのビデオカメラが散在している環境において、他の人が撮影したビデオ映像を、自分の記録ビデオとしても活用できるようにする撮影システムを提案する。例えば、運動会撮影ビデオの場合であれば、自分の子供のゴールシーン映像が、他の人が撮影したビデオテープに、より良い映像として撮影されている可能性が多分にあり、その瞬間の映像を自分の記録ビデオとして容易に取り込むことができればビデオコンテンツの共有利用が可能となる。

MVP を用いれば、複数人が個別に撮影したビデオ



図 3 撮影者 3 人によるアドホックビデオネットワーク構成例

映像であっても、撮影時刻が同一の映像を抽出して対応付けることができるため、他人が撮影したビデオ映像を簡単に自分の映像に反映させることができる。この時、上記のような運動会の撮影においても、なんら装置を追加することなく、また、撮影準備として特別な操作をすることなく、他の人が撮影したビデオ映像を自分の記録ビデオとして活用できるという点が MVP の特徴である。我々は、運動会撮影での利用以外にも適用分野は広いと考えており、スポーツの撮影、結婚式の撮影、実験観察のための撮影等においても利用可能と考えている。

4.2. 課題

現状、MVP というアプリケーションがインストールされていない PC では多視点ビデオのビデオ再生できない。そこで、MVP がインストールされていない環境であってもビデオ再生できるようにするため、多視点ビデオを自動生成し、この生成した多視点ビデオを別のビデオファイルとして

保存できるようにすることを考えている。すなわち、現在の実装では、MVP は、関連するビデオ映像を並べて表示するようになっているが、ピクチャインピクチャ等の方法で複数の映像を合成して表示したり、また、それを一つのビデオ映像として新たに保存したりするようにする。このような方法を用いれば、生成された多視点ビデオ映像をビデオストリーミング形式に変換し、インターネットから

ストリーミング配信するというようなことも可能となる。

また、現状のシステムでは、ビデオキャプチャしたビデオファイルの全てから撮影日時情報を検出する必要があり、この処理を効率良く行う必要がある。前記運動会のシナリオでは以下のような方法が考えられる。

1. 各自が撮影したビデオテープを収集し、一括してビデオキャプチャと撮影日時解析とを行う
2. ビデオキャプチャは各自が行い、各自が作成した DV ファイルを収集し、一括して撮影日時解析を行う
3. ビデオキャプチャと撮影日時解析とを各自が行い、各自が作成した撮影日時情報を収集する

素人が撮影者となる運動会撮影のような場合、各自の自宅 PC に必ずしも DV ビデオキャプチャ環境

が整備されているとは限らないため、一番確実なのは方法1であると思われる。ただし、撮影日時解析は撮影時間の10分の1の時間で完了するものの、ビデオキャプチャには撮影時間と同じ時間かかるのが普通であり、ビデオテープの本数が多い場合には多大な時間がかかることを覚悟しなければならない。

方法2については、DVキャプチャして作成されるDVファイルは、10分映像で約2GB強のファイルサイズになるため、そのDVファイルを電子メールに添付したり、FTPでファイル転送したりすることは非現実的である。よって、2の方法がとられるケースは現実的に少ないと考えられる。

DVビデオキャプチャ環境が普及した環境、または、業務として本システムを使用するような場合には、方法3が最も効率の良い方法であると思われる。撮影日時解析によって作成された撮影日時情報は極めて小さなデータであるため、メールに添付して送ることも十分可能である。ただし、多視点映像を作成するためには、撮影日時情報を収集するだけでなく、撮影映像自体も収集しなければならない。この場合、方法2とは異なり、DVファイルから撮影日時情報は抽出済みであるので、DVファイル自体をネットワーク伝送する必要はなく、MPEG、WMV、RMなどの一般的に用いられている動画圧縮形式に変換し、それを伝送すればよい。例えば、10分映像の2GB強のDVファイルを、300KbpsのWMV圧縮形式(ブロードバンド用ストリーミングビデオ程度の画質)に変換した場合、ファイルサイズを約20MBにまで削減できる。これにより、FTPでのファイル転送も現実的に可能となると考える。

5. まとめ

著者らは、あるコンテキストで関連付けられた複

数メディアのことを「グリッドコンテンツ」と名づけている。グリッドコンテンツとは、個別にキャプチャしたメディアを、時間的・位置的・人的関係等の意味的グリッド(格子)にマップすることによって、後から多面的にアクセスできるようにしたコンテンツである。本稿で述べたMVPは、複数人が個別に撮影したビデオ映像を、時間的關係を解析することによって関連箇所を対応付ける機能を提供するため、我々は、グリッドコンテンツ作成ツールの一つと位置づけている。今後、実際の利用者を対象とした利用評価を行い、そのフィードバックに基づいてシステムを改良してゆきたいと考えている。

謝辞

本研究の一部は、富士ゼロックス株式会社ITメディア研究所からの資金援助を得て実施されました。心より感謝申し上げます。

参考文献

1. 市村, 宇田, 伊藤, 田胡, 松下: QuickBoard: 汎用Webブラウザのためのリアルタイムスライド配信サービスの試作, 情報処理学会グループウェアとネットワークサービス研究会報告, GN-46-7, pp. 37-42, (2003).
2. 市村, 福井: 映像ブラウジングのためのビデオファイルブラウザおよびビデオテープブラウザの試作, 情報処理学会インタラクシオン2003, インタラクティブセッション, pp. 171-172 (2003).
3. <http://www.jisc.go.jp/>, ISO/IEC 61834.