

デジタル一眼レフカメラによるマイクロフィルムの デジタル化の予備的研究

酒 井 純

Preliminary research on digitalization of microfilm by digital single lens reflex camera

Jun SAKAI

要 旨

この研究では、マイクロフィルムやマイクロフィッシュなどの文献資料を、高価な専用機器を用いずに、安価なデジタル一眼レフカメラを用いてデジタル文字列データ化するための、予備的な実験を行っている。このために、OCR（光学文字認識）に必要な画像解像度の条件を実験から計算し、普及価格帯のデジタル一眼レフカメラでのデジタル化が数値上可能であることを確認した。その上で、現有の機材を用いて実際にマイクロフィッシュの撮影実験を行った。この結果として撮影条件がそろった場合には、OCRの認識率は紙媒体の資料をスキャナーを用いてデジタル化した場合と遜色なく、いくつかの改善すべき問題点はあるものの、マイクロ資料のデジタル一眼レフカメラによるデジタル化が可能であることが確認できた。

キーワード：デジタル一眼レフカメラ、マイクロフィルム、マイクロフィッシュ、OCR、認識率

1. はじめに

この研究は、マイクロフィルム、マイクロフィッシュなどの各種文献資料（以下マイクロ資料と表記）を、デジタル一眼レフカメラを用いてデジタル画像化し、光学文字認識（以下 OCR と表記）ソフトウェアによりデジタル文字列データ化(以下デジタル化)するための、予備的な実験を行うことを目的としている。

1.1. 研究の背景

これまで筆者は、古ロシア語を含むロシア諸年代記のコンコーダンス作成（いわゆるコーパスの作成）の研究に、主にコンピューター処理の面から協力してきた（「中世ロシア文章語成立に関与するロシア諸年代記の語彙的特徴と表現に関する研究」他、平成14年度～23年度、研究代表者中條直樹）。またこれを補完する研究として、「古ロシア語文献デジタル化の為の、初期キリル文字を含むウェブフォントの開発について」（科研費基盤研究(C)）、課題番号25370390、研究代表者酒井 純）、および酒井2015「拡張キリル文字群を含む Unicode フォントの作成について」なども行ってきた。

これらの研究では主に、製本された紙面の古ロシア語書籍をコンピューターに接続したフラットベッドスキャナーで取り込んで、OCR ソフトで認識し、これを校正して、紙面・またはデジタルのコンコーダンスに変換して発表している。また、2004年には、「古・中世ロシア文字の OCR ソフトによる認識のための実験的予備研究」（科研費萌芽研究、課題番号15652016、研究代表者 MUHTAR Mahsut、研究分担者中條直樹、2004年度研究協力者酒井 純、以下 MUHTAR 2004と表記）にも協力し、この時に手書きマイクロ資料のデジタル化に挑戦している。この時には、透過原稿ユニットを搭載したフラットベッドスキャナーを用いてマイクロ資料を読み取り、デジタル画像化を行い、その画像をOCR ソフトに入力して、デジタル化を試みている。この研究の結果として、マイクロ資料のフラットベッドスキャナーによる読み取りには、困難な点があくつかあることがわかり、この方法でのデジタル化は難しいことがわかっている。

その後、マイクロ資料のデジタル化について、再度挑戦したいと考えていた。

そのような中で、デジタルカメラの高画素・高画質化などの発展により、デジタルカメラの解像度がフィルムを超えてきていることと、写真のフィルムを直接デジタル化するのに用いられるためのフィルムデュプリケーター¹の発売により、マイクロ資料のデジタル化の可能性がでてきたと考えた。つまり、これらデジタル一眼レフカメラおよびフィルムデュプリケーターを用いることで、前回の研究で問題となった点が解決できるのではないかと考え、その条件を検討するとともに予備的な実験を行うこととした。

1.2. 研究の目標

今回の研究の目標の1つ目は、マイクロ資料のデジタル一眼レフカメラによるデジタル化に必要な機材とその性能、および実験環境の必要最低限の条件を明確にすることである。このため、紙媒体の資料をフラットベッドスキャナーとOCRソフトを用いてデジタル化するときの条件をベンチマークとして、デジタル一眼レフカメラによるデジタル化の撮像素子等の理論的な最低条件を検討する。

2つめの目標としては実際の撮影実験から、デジタル一眼レフカメラによるマイクロ資料のデジタル化の実現可能性と問題点、および今後の課題について検討することである。このため、現在筆者の所有する機材を用いて実際に撮影実験を行い、デジタル一眼レフカメラでのデジタル化の可能性を確認するとともに、撮影機材や実験環境についての条件の検討と、問題点の検討を行う。

これらの目的、目標から、今回の研究では次のような手順で実験を進めることとした。

- ① 初期キリル文字を含む、ロシア年代記の実験用元データの作成
- ② ①をもととした、紙原稿データを作成
- ③ ②をもとに、マイクロフィッシュを98コマ、60コマで作成
- ④ ②とスキャナーを用いて、OCRに最低限必要となる解像度を実験から割り出す
- ⑤ ④をもとに、デジタル一眼レフカメラの撮像素子に求められる、最低解像度を割り出す

- ⑥ ③で作製したマイクロフィッシュとデジタル一眼レフカメラを用いて撮影実験を行う

2. マイクロ資料のデジタル化について

マイクロ資料をデジタル化するにはいくつかの方法が考えられるが、ここではそれらに検討を加えるとともに、先行研究での問題点について検討する。

まずマイクロ資料の種類についてであるが、主なものとして35mmと16mmのロール状フィルム、60コマおよび98コマのシート状のマイクロフィッシュ、またカードに貼り付けられたアパーチャカードなどがある（小島2015）。今回の研究では、この中で MUHTAR 2004 にならい、シート状のマイクロフィッシュを実験対象とする。

2.1. マイクロ資料のデジタル画像化

本来マイクロ資料をデジタル画像化するには、マイクロフィルムリーダー²を用いる方法が一番よいが、問題となるのがそのコストである。

まずマイクロフィルムリーダーの直接の購入についてであるが、現在販売されているマイクロフィルムリーダーは、多くの機種が100万円を超えており、大学の研究室単位での導入は難しいと言える。また図書館での導入についても、マイクロ資料の利用が減っている現在では他の研究者の理解が得にくく、難しいと言えるであろう。

2つめの方法としては、コマ単位でのデジタル画像化に対応する業者に外注する方法もある。この方法は、マイクロフィルムリーダーを直接導入するのに比べてコストを低減するこがは可能である。コストについて今回の研究での方法と比較するときには、この価格が比較対象になると考えられる。

そして、今回実験を行うのが、デジタル一眼レフカメラとマクロレンズを用いてのマイクロ資料の直接的な撮影による方法である。この方法であれば、一般的な撮影機材で低コストにデジタル画像化が可能であると考えられる。

またこれ以外の方法として、デジタル画像化の機能のない、画像表示のみのマイクロフィルムリーダーの画面を、直接撮影する方法などが考えられる。

2.2. 先行研究での問題点

MUHTAR 2004では、コンピューターおよびフラットベッドスキャナーによる、手書き資料のマイクロフィッシュの読み込みが試みられている。この時の研究では、手書き文献のマイクロ資料を研究対象としていたが、この研究のデジタル画像の生成、処理の部分の成果については、今回の研究に活かすことができると考えられる。

そこで、MUHTAR 2004、2004年度報告書にある、フラットベッドスキャナーでの読み込みに関する問題点、および成果をまとめたのが次の3つの項目である。

- ① フラットベッドスキャナーに、マイクロフィッシュ専用のフィルム保持枠がないために、フィルムのフォーカスがずれる
- ② フラットベッドスキャナーで取り込んだ画像が解像度不足である
- ③ 画像処理ソフトによる、明るさ・コントラスト・シャープネスの調整は、OCRでの認識率の向上に寄与する

上記①のフィルムのフォーカスについては、保持方法があれば改善した可能性はある。現在であれば、3Dプリンターを用いた方法なども検討に値すると考えられる。また、③については、今回の研究でも作業手順に入れており、調整値についても検討する必要があると思われる。

そしてMUHTAR 2004で問題となった①および②の点については、機材をフラットベッドスキャナーからデジタル一眼レフカメラに置き換えることで、解決されると考えている。

2.3. 実験用のデータについて

今回の研究では、デジタル化の実験用元データとして、『ノヴゴロド第四年代記』（科研費「中世ロシア文章語成立に関与するロシア諸年代記の語彙的特徴と表現に関する研究」研究代表者中條直樹、研究課題番号14310219にてデータを作成）を選んでいる。

これまでの研究協力で、上記年代記を含め『ロシア原初年代記』や『ラジヴィル年代記』、『トロイツァ年代記』など、さまざまな古ロシア語文献のコンコー

ダンスを作成してきているが、実験用元データとしては今回『ノヴゴロド第四年代記』を選んでいる。この選定理由として、次の3つを条件とした。

- ① 初期キリル文字のうち、古ロシア語文献に共通してよく用いられる「Ѣ」「ѣ」「І」「і」の文字が含まれていること
- ② 「Ѣ」「ѣ」「І」「і」以外の初期キリル文字が含まれていないこと
- ③ 初期キリル文字に含まれる、特殊な記号が少ないこと
- ④ 原本のコピーが手元にあること

そして、『ノヴゴロド第四年代記』のデータのうち、冒頭10ページ分を実験用元データとしている。

次の表1は、今回作成した10ページ分のデータの数値である。

表1 ノヴゴロド第四年代記冒頭10ページ分のデータ一覧

名称	ノヴゴロド第四年代記
部分	冒頭10ページ分 脚注番号および脚注除くタイトルおよび本文
総文数	249文
総単語数	3201単語
異なり単語数	1693単語（語形変化は別として計上）
総文字数	空白を含む19066文字 空白を除く15859文字
異なり文字数	76文字
使用されている文字	і І а А б Б в В г Г д Д е Е ж Ж з З и И к К л Л м М н Н о О п П р Р с С т Т ф Ф х Х ц Ц ч Ч ш Ш щ Щ ъ Ъ ы Ы ю Я ъ
使用されている記号	0 1 2 3 4 5 6 7 8 - ! , . : ; ? [] « »

この10ページ分のデータを元に、レーザープリンターで印刷した紙面を、OCR解像度の実験用元データとしている。また、この紙面をもとにして、98コマと60コマの2種類のマイクロフィッシュを外注にて作成している。この時には、フォントの違いによるOCRの認識率を比較する実験も今後予定しているため、原本のコピーを含め、4種類の原稿をもとにした。図1が、実際に作成したマイクロフィッシュ（60コマ）である。

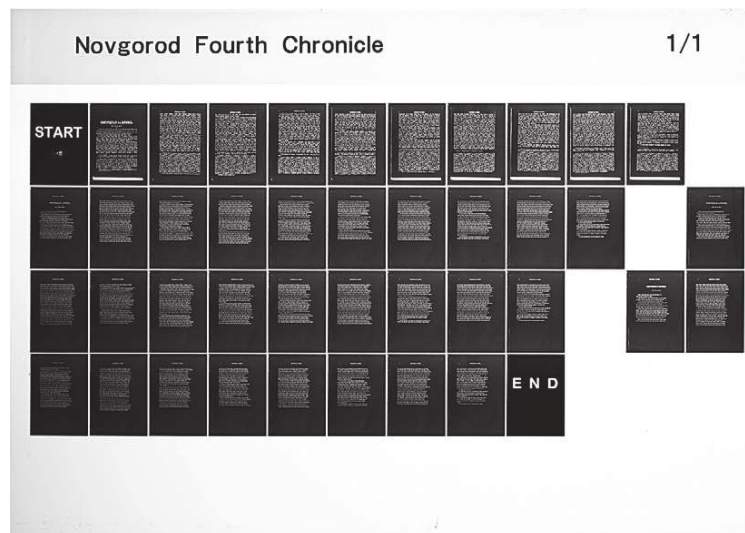


図1 作成したマイクロフィッシュ

また、今回の実験では、上記10ページのうち、7ページ目のみを用いて実験を行っている。このページを選んだ理由としては、まず1ページ目のようなタイトルがなく、脚注も少ないことにより本文の行数が多いこと、また使用されている数字・記号の種類が少ないことの2点である。そしてこの7ページ目のデータをまとめたのが次の表2である。(実際の7ページ目の文面全文については6.3を参照)

表2 ノヴゴロド第四年代記 p.7データ

名称	ノヴゴロド第四年代記
部分	p.7、脚注番号および脚注除く本文
総文数	26文
総単語数	329単語
異なり単語数	231単語 (語形変化は別として計上)
総文字数	空白を含む2078文字 空白を除く1768文字
異なり文字数	空白記号含む39文字
使用されている文字	і І а А б Б в В г Г д е ж з З и И к К л м н о О п П р с С т Т у х Х ц ч ш щ ъ ы ь ю я ъ
使用されている記号	-, ., ;

3. デジタル化に必要なとなる機材と理論値

ここではまず、マイクロ資料の撮影に必要なとなる機材と、最低限必要と考えられるデジタル一眼レフカメラの撮像素子の解像度について述べる。

3.1. 撮影機材

マイクロ資料をデジタル化するのには、デジタル一眼レフカメラと専用マクロレンズを含めて、以下の機材が必要と考えられる。

3.1.1. デジタル一眼レフカメラ

デジタル一眼レフカメラについてであるが、この研究の趣旨から普及価格帯の APS-C サイズの撮像素子を持つデジタル一眼レフカメラが、その目的に一番合致していると考えられるであろう。

現在販売されているデジタル一眼レフカメラは、カメラ内のミラーの有無によって、一眼レフタイプとミラーレスタイプの2種類に大きく分けることができる。この2つについて、コスト面で大きな差は無いと考えられるが、機種や交換レンズの選択肢が多い点では一眼レフタイプに、またミラーの動作がないことによる撮影時の振動の少なさ³という点ではミラーレスタイプに利点がある。

次にミラーレスカメラを含むデジタル一眼レフカメラの撮像素子の大きさとしては、35mmフルサイズ、APS-C サイズ、4/3型⁴の3種類が現在主流である。今回の研究では、1：1の等倍まで拡大して撮影を行うため、撮像素子の大きさは1度に撮影可能なページ数に直結する。

次の図2は、60コマタイプのマイクロフィッシュの1コマの大きさを、それぞれの撮像素子の大きさにあてはめたものである。この図からもわかるように、60コマのマイクロフィッシュで換算して、フルサイズの撮像素子でA4を3枚分、APS-C サイズで2枚、4/3型で1枚をデジタル画像化できる。一方でカメラ本体の価格については、フルサイズの撮像素子を持つデジタル一眼レフカメラは高価である一方⁵、

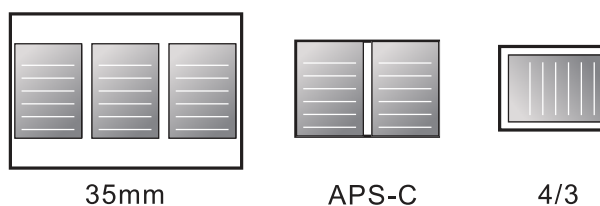


図2 撮像素子とマイクロフィッシュのコマの対比

APS-C サイズと4/3型のカメラはともに安価なものも多くあり、コストの差があまり無いことから、一度に撮影できる枚数が2枚になる APS-C サイズが目的に合致すると考えられる。

また、撮像素子の解像度については、「3.2 OCR に必要な解像度とデジタル一眼レフカメラでの理論値」で詳しく述べる。

3.1.2. マクロレンズ

デジタル一眼レフカメラでのマイクロ資料のデジタル化を考える場合、必要となるのがマクロレンズで、60mmの焦点距離⁶を持つ等倍マクロレンズが目的に合致する。

撮影倍率については、現在市販されているマクロレンズは、ほとんどのものが等倍マクロ、または0.5倍マクロレンズであり⁷、解像度から等倍マクロレンズが必要である。

次にマクロレンズの焦点距離については、60mm程度と、100mm程度のレンズに分かれる。この2つを比較すると、被写界深度を深めにとれるという点で60mmのレンズに、また撮影距離を取りやすいという点で100mmのレンズに利点があると言える。ただし撮影距離はそこまで問題にならないこと、また100mmマクロの方がブレやすいことと高価である点から、60mmのマクロレンズを選択すべきであると考えられる。

またレンズの性能の問題として、レンズの中心部分の解像度と周辺部分での解像度には差があるため、これが最終的な OCR の認識率の違いにつながるかどうか、今後検討が必要になると考えられる。

3.1.3. 三脚

デジタル一眼レフカメラでマクロレンズを用いて撮影を行う場合には、カメラ本体を固定する三脚が必ず必要となる。特にマクロレンズを用いて等倍付近で撮影する場合には、ほんの数ミリの違いがピントのずれに反映されるため、カメラをしっかりと固定する必要がある。

また、今回の撮影では一般的な三脚のみを用いているが、三脚の上に取り付ける前後左右の微調整をする台があり、今後こういった製品の導入も検討すべ

きと考えている。

3.1.4. 照明

マイクロ資料は、そのベースフィルムが透明であるため（図3）、フィルムを挟んで反対側からの照明が必要となる（図4）。また、この照明も単に照らすだけでなく、フィルム面全体を均一の明るさに照らす必要がある。このためには大きく3つの方法が考えられる。

1つめの方法は、フィルムデュプリケーター（図5）を使う方法である。フィルムデュプリケーターは、写真用フィルムをデジタル一眼レフカメラを用いて撮影するための専用機材で、現在は RICHO 社から、35mmサイズと4×5サイズの2種類が発売されている。これと外部ストロボを使うことによって、マイクロフィッシュを撮影することができると考えており、今後導入の上で実験を行いたい機材である。

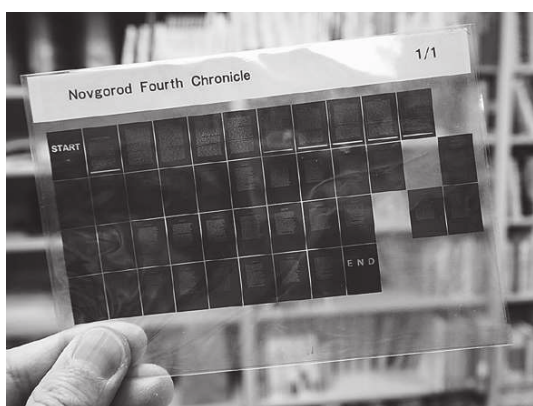
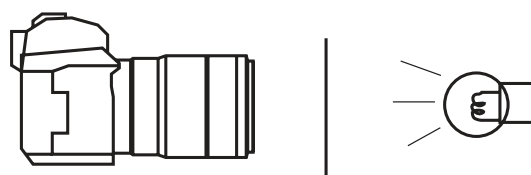


図3 透明なベースフィルム



カメラ フィルム 光源

図4 照明の位置関係



図5 フィルムデュプリケーター
RICHO サイトより引用

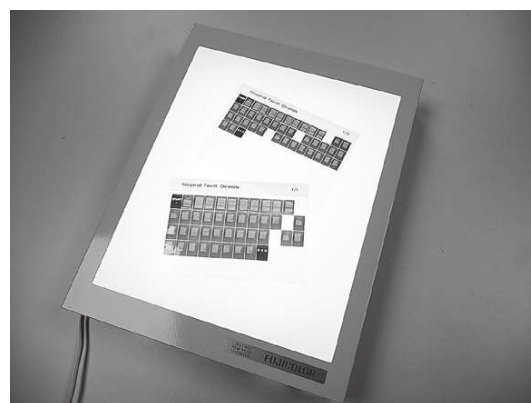


図6 ライトボックスの例

次に2つ目の方法としては、写真用フィルムを確認するためのライトボックス（図6）を照明とするものである。ライトボックスは、光源と光拡散用半透明板が組み合わさっており、もともと写真用フィルムを確認するために使われる機器である。光源が均一であるという点、フィルムを載せる面の平面性が高いという点では今回の研究の目的に合致しているが、目視のために必要な光量しかないために撮影の照明としては光量が不足すると考えられる。

3つめの方法は、白色の半透明な板と光源として何らかの照明器具、または外部ストロボを用いる方法である。

これは、デュプリケーター、またはライトボックスを自作するのに等しい方法であり、光源として様々なものを利用可能である点が有利である。光源としては、蛍光灯や電灯、デスクライトなど日常的な照明器具を用いる方法と、カメラの外部ストロボを用いる方法の2つが考えられる。日常的な照明を用いる方法では、新たな機材の購入が必要ないなどメリットもあるが、上記ライトボックス以上の光量を得るには、相応の明るさを持つ照明器具が必要となる。これに対して外部ストロボを用いる方法では、ストロボ本体と接続用の専用コードの導入が必要となるが、光源としての光量は十分であり、費用対効果は高いと考えられる。

また、この自作する方法では、半透明な板を撮像素子に対して平行に、また平面状に保持する器具を作成する必要がある、この点が不利と言える。

3.1.5. 画像処理ソフト

2.2の③でも述べているように、MUHTAR 2004によれば、撮影したデジタル画像データは、画像の明るさおよびコントラストを調整すること、シャープネスを上げることで、OCRソフトでの認識率が上がることがわかっている。このため、画像の明るさとコントラスト等を編集できる画像処理ソフトが必要となる。また、マイクロ資料には、白黒反転したネガ資料と、現行そのままの白黒のポジ資料があり（小島2015）、今回のようなネガ資料の場合には、画像処理ソフトにネガポジ反転が必要となる。

一方で、明るさやコントラストを同一条件で一括して画像処理できるバッチ

機能やドロップレット機能もあることが、利便性の面からも求められるであろう。

3.1.6. OCR ソフト

光学文字認識ソフトウェア、いわゆる OCR ソフトについては、今回の実験対象がロシア年代記であることから、古ロシア語への対応が条件となる。特に、現代ロシア語では用いられない「Ѣ」「ѣ」などを含む初期キリル文字を認識できるかどうかにより、文字認識率が大きく変わると考えられる。

現在日本国内で入手可能な OCR ソフトで、古ロシア語に対応したものとしては、ABBYY 社の Fine Reader Pro と、OMNI 社の Omni Page の 2 つがあげられ、また両ソフトともに欧文の認識率には定評がある

3.1.7. その他

その他実験に必要なになる機材としては、画像処理ソフトや OCR ソフトを用いるためのコンピューター、デジタル一眼レフカメラの記録用メモリーカード、またコンピューターにメモリーカードからデータを読み込むためのカードリーダーなどがあげられる。またパソコンとデジタル一眼レフカメラを直結して撮影を行う場合には、USB 接続ケーブルなども必要となるだろう。

3.2. OCR に必要な解像度とデジタル一眼レフカメラでの理論値

デジタル一眼レフカメラでマイクロ資料をデジタル化する際に必要となる撮像素子の解像度は、フラットベッドスキャナーを用いて紙媒体をデジタル化する際の解像度を参考にすることで、擬似的に計算できると考えた。このため、まず実験用元データを印刷した紙媒体の古ロシア語の文章を、フラットベッドスキャナーでいくつかの解像度でスキャンしてデジタル画像データを作成し、これを OCR ソフトで文字認識させたときの認識率を比較することで、最低限必要となる解像度を検討した。またこれを元に、撮影に用いるデジタル一眼レフカメラに必要なとされる解像度を計算した。

3.2.1. OCR に必要な解像度

紙媒体をスキャンして OCR ソフトを用いてデジタル化する場合に、通常どれくらいの解像度が必要となるのか、実験を行ってその数値を求めた。このた

め、「2.4 実験用のデータおよびマイクロフィッシュについて」で作成した実験用元データを印刷した紙原稿を、フラットベッドスキャナーを用いて各解像度でスキャンし、これを OCR ソフトで認識させて、元の原稿データと比較することで OCR での単語認識率を計算した。この結果をまとめたのが、次の表 3 である。

表 3 紙原稿の dpi 別 OCR 認識率

	100dpi	110dpi	120dpi	130dpi	140dpi	150dpi	160dpi	170dpi
1 回目	84.2%	92.0%	95.5%	95.8%	97.9%	97.3%	94.3%	96.1%
2 回目	89.0%	90.5%	96.1%	95.2%	95.8%	96.4%	97.9%	95.2%
3 回目	80.7%	91.1%	92.6%	93.5%	95.2%	97.6%	96.1%	97.3%
平均	84.6%	91.2%	94.7%	94.8%	96.3%	97.1%	96.1%	96.2%

	180dpi	190dpi	200dpi	210dpi	220dpi	230dpi	240dpi
1 回目	95.8%	95.8%	96.4%	95.8%	97.6%	96.4%	95.8%
2 回目	94.6%	96.1%	96.1%	94.0%	94.3%	95.5%	96.1%
3 回目	95.8%	96.1%	97.6%	95.2%	96.1%	96.1%	94.3%
平均	95.4%	96.0%	96.7%	95.0%	96.0%	96.0%	95.4%

また、このデータを元に作成したのが、次の図 7 のグラフである。

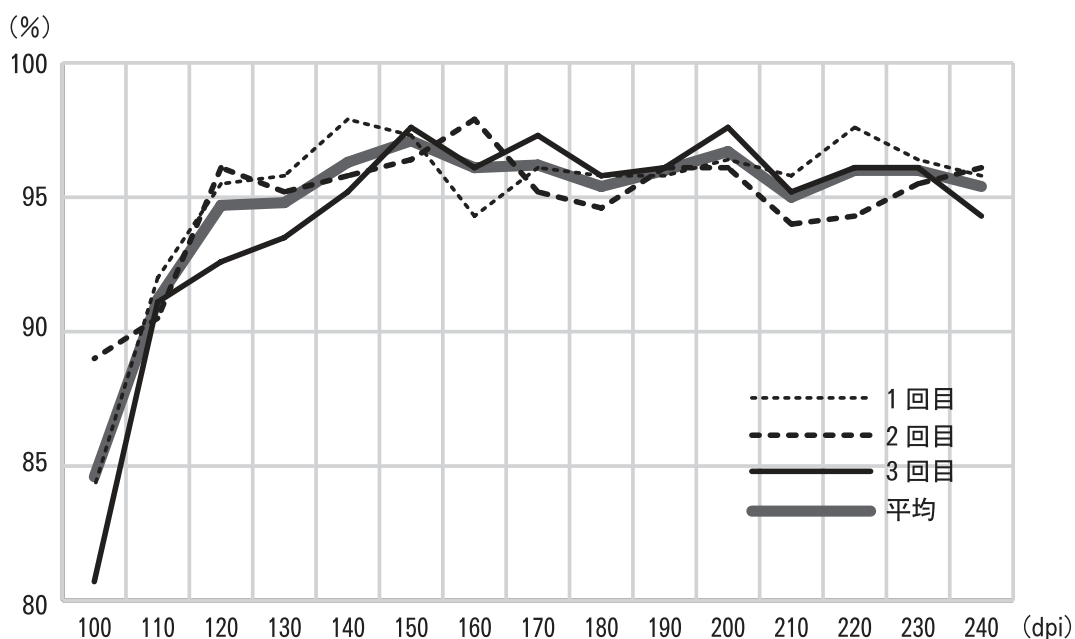


図 7 紙原稿の dpi 別 OCR 認識率

このグラフからもわかるように、紙原稿の OCR 認識率は120dpi ほどで実用域に入り、150dpi あたりをピークにそれ以上は横ばいである。このことから、OCR ソフトに入力するデジタル画像の解像度としては、最低で120dpi 必要であり、また200dpi 以上は必要ないことも分かった。

3.2.2. 必要とされる撮像素子の解像度

今回の実験では、マクロレンズを用いて1：1の等倍で撮影するため、撮像素子の性能として重要となってくるのが解像度 (dpi、1インチ当たりの画素数) である。また、マイクロ資料はもととなる原稿を数十分の一に縮小しているため、この縮小分を計算に入れて実験をする必要がある。

結果としては、撮像素子の必要解像度 (dpi) = OCR に必要な解像度 (dpi) ÷ マイクロ資料の縮小倍率となる。OCR に必要な解像度を上記3.2.1の結果から120dpi と設定して計算した結果をまとめたのが、次の表4である。

表4 必要解像度

コマ数	縮小倍率	必要解像度
60コマ	1/20	$120 \div 1/20 = 2400\text{dpi}$
98コマ	1/25	$120 \div 1/25 = 3000\text{dpi}$

ただし、デジタル一眼レフカメラの撮像素子は、ベイヤー配列といわれる、赤緑青のフィルターが1画素ごとに配置されており、単純にフィルムと同等の解像度が得られるわけではない。それでも現在のデジタル一眼レフカメラであれば、普及機でさえ6000dpi 以上とこれらの数値を大きく超えており、理論的にはデジタル一眼レフカメラによる撮影は十分に可能であると考えられる。

4. デジタル化の実験

この実験の目的は、現有のデジタル一眼レフカメラを用いてマイクロ資料のデジタル化が可能であることを確認するとともに、撮影等にかかわる問題点について検討することである。また、これらの問題点について、今度どのような方法によって解決していくのかについても検討する。

4.1. 実験手法

今回の実験は、以下にあげる機材を用いて、まず実験用マイクロフィッシュ(60コマタイプ)の7ページ目を、3つのカメラで撮影しデジタル画像化する。そしてこれを画像処理ソフトにて、ネガポジ反転、明るさ・コントラスト・シャープネスの調整をして、OCRソフトで文字認識、デジタル化を行う。この結果を、実験用元データと比較して単語単位での認識率の計算を行う。

4.2. 実験に用いる機材

今回の実験は、筆者が現在所有または利用可能なデジタル一眼レフカメラ(ミラーレスカメラを含む)で行っており、必ずしも前述した条件すべてに合致しているわけではない。

4.2.1. カメラ本体

今回の実験では、撮像素子の大きさ、解像度の違いを調べる目的もあり、次の3つのカメラを用いて撮影を行った。

表5 実験で用いたデジタル一眼レフカメラの仕様一覧

	カメラ1	カメラ2	カメラ3
メーカー	Canon	Canon	Olympus
機種名	EOS 5D mark III	EOS 40D	Pen Lite E-PL6
発売年	2012年	2007年	2013年
タイプ	一眼レフ	一眼レフ	ミラーレス一眼
撮像素子	35mmフルサイズ	APS-C サイズ	4/3型
解像度	約2210万画素 5760×3840	約1010万画素 3888×2592	1605万画素 4608×3456
撮像素子	36×24mm	22.2×14.8mm	17.3mm×13.0mm

解像度 (dpi)	4064.0dpi	4448.4dpi	6765.5dpi
60コママイクロフィッシュでの疑似解像度	201.7dpi	220.8dpi	335.9dpi
98コママイクロフィッシュでの疑似解像度	161.0dpi	176.2dpi	268.0dpi

また、60コマおよび98コマのマイクロフィッシュを1：1の等倍撮影した場合に、これをA4の大きさ拡大した時の画像の解像度についても計算している。

ただし前述の通り、デジタル一眼レフカメラを含めるデジカメの撮像素子は、赤緑青緑の4つのピクセルでひとまとまりとなる、ベイヤー配列をしており、単純にフラットベッドスキャナーの場合とは比較できない。このため、上記表の解像度がそのまま利用できるとは考えにくい。しかしながら、どのカメラもOCRに必要とされる120dpiは大きく超えており、十分に条件を満たしていると考えられる。

4.2.2. マクロレンズ

マクロレンズについては次の表6にあるように、各カメラに対応したレンズを用いた。

表6 実験で用いたマクロレンズの仕様一覧

使用したカメラ	カメラ1	カメラ2	カメラ3
メーカー	Canon	Canon	Olympus
使用レンズ	EF100mm F2.8L マクロ IS USM	EF-S60mm F2.8 マクロ USM	M.ZUIKO DIGITAL ED 60mm F2.8 Macro
焦点距離 (35mm換算)	100mm	96mm	120mm

マクロレンズについては3.1.2でも述べているように、35mm換算で100mm付近の望遠マクロよりも、60mm付近の標準マクロのレンズが向いていると考えられるが、手持ちの機材の都合により、3つのカメラともに100mm付近のマクロレンズを使用している。

4.2.3. 三脚、照明、ソフトウェア等

三脚については、SLIK カーボンスプリント613 PRO という水準器付きの、一般的な形状のものを用いている。また、照明およびマイクロフィッシュの保持板として、フジカラー社の「ライトボックス NEW5000インバーター」を用いている。蛍光灯を用いた、インバーター電源のものであるが、光量が少ない点が問題であった。

画像処理ソフトについては、ネガポジ反転とコントラスト、明るさの調整、アンシャープマスクフィルターの機能を持つ、Adobe 社 Photoshop CC 2017を用いている。これは3.1.5で述べた機能すべてを含むソフトウェアである。また OCR ソフトとして、ABBYY 社の Fine Reader Pro 12を用いている。これは、古ロシア語へ対応している OCR ソフトの中で、これまでのロシア諸年代記コンコードンス作成で用いてきたことから採用している。そして、OCR ソフトによる認識率の計算には、MS-Excel を用いた。

4.3. 撮影実験の実施

撮影は、すべて ISO 感度100、絞り f8.0 の条件にそろえて各カメラ、3回ずつ行い、画像をコンピューターに取り込んで明るさ、コントラスト、シャープネスをそれぞれ調整して画像を作成した（図8）。そしてこれを OCR ソフトに読み込んで文字認識し、出力された文字列を元データと比較して、単語単位での認識率を計算した。

всякъ звѣрь, ядоуще все нечисто, срамословные в нихъ предъ родители, и племенїи не стыдятся, браци не бѣваху въ нихъ, но играша между селы ихъ; и сковахоуся на играша и на вся бесовская пласания, и ту оумьахоу собѣ жены, с нежеже кто сѣчався, и имажуть же и 2 и по 3 жены. Аще кто оумрише оу нихъ, и творяхоу тризну надъ нимъ, и по семь събраше кладу велику, и възложатъ на кладу мертвца и съжгоуть и, и по семь събраше кости и вложатъ в соудиноу малу, и поставяхоуть на поутехъ на столпѣхъ; еже творить Вятци и до сего дни. Сии творить обычи и Кривци и прочи погани, не ведоуще закона Божїа, но творяще сами собѣ законъ.

Глаголетъ бо Георгїи въ лѣтописани: ибо коемоуждо языку, овѣвъ исписанъ законъ есть, другимъ же обычая; зане безаконнымъ отечество мнитса. Отъ нихъже Переди и Сїри, живоуша на конецъ земли, законъ имоуть и отечество ихъ обычаи: не любодѣти, ни красти, ни оклеветати, ни сбити, ни злодѣяти веси. Законъ же Оуктирякъ, глаголеми Врахмане и Островьници, иже отъ прадѣвъ наказаніемъ и благочестіемъ мясь не ядоуще, ни вина пьюще, ни блуда творяще, сътраха ради многа, ни злобы творяще никоеже. Ибо друзини близъ тѣхъ соуть инымъ закономъ, съхверно творяще все, гѣлвалии паче естества; вноутреніише странѣ, и бѣахоу челоуѣкъ ядоуще и странствующихъ оубивахоу, паче ядоху яко пси. Ииъ же законъ Халдѣемъ и Вавилоняномъ: матери помати, съ братними чады блудъ творити и оубивати, и всякое бестодное дѣяние творять. Ииъ же законъ Тилномъ: женимъ оу нихъ оротъ, и храмы създадоуть и всяко мужская орудья творять, и прелюби творитни еи, еликоже хотять, и не въздержани отъ мужъ своихъ выину: и соуть жены ты храбри; ловить звѣрь крѣпко, и мужи своими владють жены тыи. І въ Вергани же мнози мужи съ единою женою спятъ, и жены съ единымъ мужемъ похотьсъвоуютъ: безаконныи и законъ творять незавестно, ни въздержано. Амазоняни же мужа не имажуть, но акы скотъ бесловесныи, но одиноко днѣмъ къ весниимъ днѣмъ оземствении боудуть, съчтаются съ окрестными ибо мужи, яко иѣаоторое имъ торжество великое празднество

図8 処理の終わった画像の例

4.4. 結果

この実験の結果をまとめたのが次の表7である。

表7 デジタル一眼レフカメラによる OCR 認識率

	カメラ 1	カメラ 2	カメラ 3
1回目	93.8%	91.7%	92.6%
2回目	82.7%	90.2%	92.0%
3回目	89.0%	95.8%	86.6%
平均	88.5%	92.6%	90.4%

実験の結果から、条件が整えばデジタル一眼レフカメラによるマイクロ資料のデジタル化は可能であることが、まず証明できたと考えられる。カメラ2の

3回目のように、条件が揃えば95.8%の認識率となり、これは紙媒体からの文字認識と比較しても差はなく、十分に実用に足るといえる。また、カメラ1でも最大93.8%、カメラ3でも92.6%と、撮像素子の大きさにかかわらず、デジタル化が可能であることも確認できた。

その一方で、撮影回数ごとの認識率のばらつきが大きいことも明らかである。今回の実験でもカメラ1の2回目には82.7%と、実用域から大きく外れる認識率となっており、次に述べるいくつかの問題点を解決して、安定的に認識率を上げられるようにする必要がある。

4.4.1. 今回の実験で確認できた問題点

① 撮影対象と撮像素子の平行の調整

今回の撮影実験で難しかったのが、撮像素子とマイクロフィッシュのフィルム面を平行に調整、維持することであった。縦方向の平行性（図9）については、照明とフィルムの維持に用いたライトボックスの形状、および三脚の水準器により、ほぼ平行に保つことができた。しかしながら、横方向の平行性（図10）については、ライブビューで拡大してピントを見ながら、カメラとライトボックスを移動させて調整する必要があり、この調整に非常に時間と手間がかかり、また左右でピントのずれ＝平行性が悪い画像も多かった。

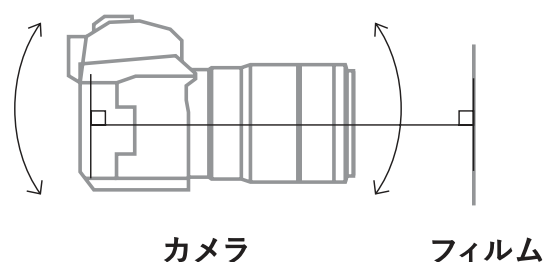


図9 縦方向の平行性

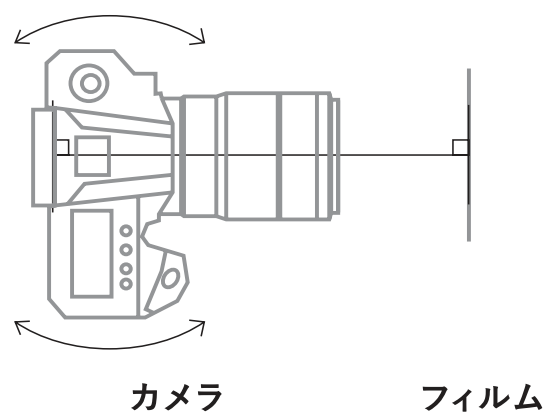


図10 横方向の平行性

このため、横方向の平行性の調整と維持が、今後の課題になると考えられる。

② ライトボックスの光量不足

これは実験開始前からわかっていたことであるが、やはりライトボックスの光量ではマイクロ資料の撮影には光量不足であった。実際に ISO 感度100、絞

り f 8.0の時に、シャッタースピードが1/15～1/50秒となり、後述する三脚の保持力不足もあって画像のブレが何度も生じてしまった。このため、今後は照明にストロボを用いる方法を検討すべきと考えられる。

③ 三脚の保持力不足

今回用いた三脚は、カーボン製の軽量なものであったこともあり、マクロ撮影には保持力不足であった。上記②の理由によりシャッタースピードが遅いこともあり、一眼レフタイプ（5D3、40D）では、ミラーアップをした状態で撮影しないと、シャッターを切ったときのミラーアップ・ミラーダウンの振動で、画像にブレが生じてしまった。また、操作のためにカメラ背面のボタンを押すだけで、ピントがずれることも多くあった。

このため、次回以降の実験時にはもっと重量のある、剛性の高い三脚を使う必要があるだろう。

④ 画像のブレについて

③、④にも書いたとおり、いくつかの理由から画像にブレが生じることが多かった。根本的には、ライトボックスのかわりにストロボを証明として使用すること、またしっかりとした三脚を使うことで解決可能である。しかしながら、カメラ本体の撮影技法として、ミラーアップした状態での撮影や、リモートリリースなどを用いた撮影方法でも改善は可能であり、これらも組み合わせて問題の解決につなげられればと考えている。

⑤ 一コマごとの撮影時間

①で述べている横方法の平行性の調整に特に手間がかかることもあり、1コマごとの撮影に時間がかかってしまった。これは、今後撮影枚数が増えれば増えるほど、問題になることは確実である。このため、ストロボと半透明版を用いた照明・マイクロ資料保持装置の自作時に、この問題を解決するための機構を組み込むことができればと考えている。

5. おわりに

今回の研究の目的である、デジタル一眼レフカメラによるマイクロ資料のデ

ジタル化が可能なが、実際の撮影実験から証明できたと考えている。先行研究である MUHTAR 2004では、マイクロフィッシュのデジタル画像への取り込みにフラットベッドスキャナーを用いていたが、これをデジタル一眼レフカメラによるものに変更することで、状況が改善できたと考えている。

また、目標としていた撮影条件の明確化についてであるが、まずデジタル一眼レフカメラの撮像素子の解像度については、現在の普及機の性能で十分であることがわかった。また、撮像素子の性能よりも、フィルムと撮像素子の平行性や、三脚によるブレの発生など、撮影時に発生する問題の方が深刻であることもわかった。これらの問題点については、今後実験を重ねながら改善していく必要があると考えている。

6. 参考資料

6.1. 参考文献

- 小島浩之2015『図書館資料としてのマイクロフィルム入門』日本図書館協会
- 酒井 純2000「古ロシア語文献のコンコードダンス化における技術的問題について」『岡崎女子短期大学研究紀要』第38号 p.95-103、
- 酒井 純2005「古ロシア語文献に用いられる初期キリル文字の Unicode フォントの作成について」神戸親和女子大学『言語文化研究』創刊号 p.103-128
- 酒井 純2015「拡張キリル文字群を含む Unicode フォントの作成について」神戸親和女子大学『言語文化研究』第9号 p.1-20
- 中條直樹、酒井 純2005『ノヴゴロド第4年代記コンコードダンス』日本古代ロシア研究会

6.2. 関係する科学研究費助成事業

「中世ロシア文章語成立に関与するロシア諸年代記の語彙的特徴と表現に関する研究」科学研究費基盤研究B、研究課題番号14310219、研究代表者中條直樹
<https://kaken.nii.ac.jp/ja/grant/KAKENHI-PROJECT-14310219/>

「古・中世ロシア文字の OCR ソフトによる認識のための実験的予備研究」科

研費萌芽研究、課題番号15652016、研究代表者 MUHTAR Mahsut、研究分担者
中條直樹、2004年度研究協力者酒井 純

<https://kaken.nii.ac.jp/ja/grant/KAKENHI-PROJECT-15652016/>

「キリル文字とその変形文字による旧ソ連圏の英雄叙事詩 -PC 利用の資料研究」
科研費基盤研究 B、課題番号18652033研究代表者 中條 直樹、研究分担者
塚原 信行

<https://kaken.nii.ac.jp/grant/KAKENHI-PROJECT-18652033/>

「古ロシア語文献デジタル化の為の、初期キリル文字を含むウェブフォントの
開発について」科研費基盤研究 C、課題番号25370390、研究代表者酒井 純

<https://kaken.nii.ac.jp/ja/grant/KAKENHI-PROJECT-25370390/>

6.3. 参考サイト

リコーイメージング「PENTAX FILM DUPLICATOR」

http://www.ricoh-imaging.co.jp/japan/products/accessory/index35_others.html

(2016年11月30日確認)

6.4. 参考資料

実験用元データ『ノブゴロド第四年代記』p.7全文

всякы звѣрь, ядоуще все нечисто, срамосьловіе в нихъ предъ родители, и племеніи не стыдятся, браци не бывахоу въ нихъ, но игрища межоу селы ихъ; и схожахоуся на игрища и на вся бесовская плясания, и ту оумыкахоу собѣ жены, с неюже кто свѣчався, и имяхоуть же и 2 и по 3 жены. Аще кто оумряше оу нихъ, и творяхоу тризну надъ нимъ, и по семь творяхоу кладу велику, и възложить на кладу мертвеца и съжгоуть и, и по семь събравше кости и вложить в соудиноу малоу, и поставляхоуть на поутехъ на столпѣхъ; еже творять Вятици и до сего дни. Сии творять обычаи и Кривици и прочии погании, не ведоуще закона Божия, но творяще сами собѣ законъ. Глаголетъ бо Георгъгии въ лѣтописании: ибо коемоуждо языкоу, овѣмъ исписанъ законъ есть, другимъ же обычая; зане незаконнымъ отечествие мнися. Отъ нихъже Персиди и Сірии, живоуща на конецъ земля, законъ имоуть и отечество ихъ обычаи: не любодѣяти, ни красти, ни оклеветати, ни оубити, ни злодѣяти весма. Законъ же Оуктирьякъ, глаголеми Врахмане и Островъници, иже отъ прадѣдъ наказаниемъ и благочестиемъ мясь не ядуще, ни вина пьюще, ни блоуда творяще, сътраха ради многа, ни

злобы творяще никоеяже. Ибо друзии близъ тѣхъ соуть инымъ закономъ, скъверно творяще все, гнѣвливи паче естества; вноутренѣишее странѣ, и бѣяхоу челоуѣкы ядоуще и странѣствующихъ оубивахоу, паче ядохоу яко пси. Инѣ же законъ Халдѣемъ и Вавилоняномъ: матери поимати, съ братними чады блоудъ творити и оубивати, и всякое бестоудное дѣяние творять. Инѣ же законъ Тилиомъ: женыи оу нихъ орють, и храмы съзиждають и всяко мужеская орудья творять, и прелюбы творити и еи, еликоже хотять, и не въздержими отъ моужь своихъ выиноу: и соуть жены ты храбрыи; ловять звѣрь крѣпко, и моужи своими владѣють жены тыи. I въ Вертании же мнозѣ моужи съ единою женою спять, и жены съ единымъ моужемъ похотѣсьтвоуютъ: незаконныи и законъ творять независтно, ни въздержанно. Амазоняни же моужа не имають, но акы скотъ бесловесныи, но одиноу днемъ къ вешнимъ днемъ оземьствении боудуть, съчтаются съ окрестными ибо моужи, яко нѣкоторое имѣ торжество великое празднество

注

- 1 リコーイメージング「PENTAX FILM DUPLICATOR」。
- 2 現在はデジタル画像化を可能にする、デジタルマイクロフィルムリーダーが主流となっている。
- 3 一眼レフタイプでも、ミラーアップしてライブビューモードにすることで、この問題は回避可能である。
- 4 (マイクロ) フォーサーズマウントのカメラの撮像素子のサイズ。
- 5 2016年12月現在、最低で12万円。
- 6 フィルム一眼レフカメラ35mm換算の焦点距離。以降すべて同様。
- 7 例外としては、1～5倍の撮影が可能な「Canon MP-E65mm F2.8 1-5 ×マクロフォト」もあるが、非常に高価である。