

SAS-Makro-Sammlung zum Geocoding von Photos – graphische Benutzeroberfläche inklusive

Thomas G. Grobe
ISEG
Lavesstr. 80
30519 Hannover
grobe@iseg.org

Zusammenfassung

Der Beitrag verfolgt Ziele auf zwei Ebenen: Auf einer allgemeinen Ebene werden Möglichkeiten von SAS zum Lesen und Schreiben von Daten in primär SAS-fremden Formaten erläutert sowie Elemente zur Programmierung graphischer Benutzeroberflächen mit SAS/AF in der Praxis vorgestellt. Der Beitrag soll dabei Hinweise zu Zugriffsmöglichkeiten von SAS auf sehr unterschiedlich strukturierte Daten liefern sowie Anregungen für einen Einstieg in die SAS/AF-Programmierung geben, die insbesondere bei oft benötigten und thematisch geschlossenen Programmroutinen mit einer größeren Zahl von Optionen Vorteile bietet.

Auf einer spezifischen, inhaltlichen Ebene wird eine Makro-Sammlung und deren Benutzeroberfläche zum Geocoding bzw. Geotagging von Photos vorgestellt, die eine automatisierte Zuordnung von geographischen Koordinaten aus sequenziellen Aufzeichnungen eines GPS-Empfängers zu digital aufgenommenen Bilddateien erlaubt. Dabei werden SAS-Datenbanken sowohl zu GPS-Informationen als auch zu den Bilddateien erstellt und ergänzend spezifische Dateien geschrieben, welche eine Visualisierung der geographisch zugeordneten Photos im Kontext mit zum Teil hochauflösenden Satellitenaufnahmen in Google Earth® erlauben. Die Makro-Sammlung ist auf der Begleit-CD enthalten und kann zur freien eigenen Verwendung bei Bedarf auch vom Autor per Mail angefordert werden.

Schlüsselwörter: SAS/AF-Beispiele, SAS-Makro-Programmierung, XML, binäre Daten, JPEG-Format, Exif-Header, Geocoding, Geotagging, GPX-Format, KML-Format

1 Geocoding von Photos – Hintergrund

Jedem Photo ließen sich, neben Informationen zum Aufnahmezeitpunkt, auch Informationen zur Aufnahmelokalisation zuordnen, wobei dieser Vorgang bei Zuordnung der geographische Koordinaten als Geocoding, oder, in bestimmten Fällen, alternativ auch als Geotagging bezeichnet werden kann. Beide Merkmale, Zeitpunkt und Lokalisation einer Aufnahme, können als universelle Ordnungs- bzw. Abfragekriterien für Photodatenbanken gelten. Die geographische Position dürfte dabei insbesondere bei Photos aus der Umwelt mit dokumentativem Charakter, z.B. bei wissenschaftlichem Hintergrund, aber auch bei Pressephotos, eine ausgesprochen relevante und erwünschte Information darstellen.

Von nahezu allen aktuellen Digitalkameras wird bereits automatisch der Aufnahmezeitpunkt erfasst und zumindest bei Verwendung des JPEG-Bilddateiformates relativ einheitlich in so genannten Exif-Datei-Headern abgelegt.

Nur wenige Kameras bieten allerdings die Möglichkeit, auch Informationen zur geographischen Position bereits während der Aufnahme zu erfassen und gleichfalls im Exif-Header abzulegen. Eine praktische Alternative zur Kamera-internen Speicherung der geographischen Position in der Bilddatei bereits während der Aufnahme besteht jedoch in der Verwendung eines separaten GPS-Gerätes mit kontinuierlicher Aufzeichnung der geographischen Position in so genannten Track-Dateien (z.B. Aufzeichnung einer Wegstrecke mit Armbandgerät über einen gesamten Tag). Da bei jeder GPS-Positionsbestimmung die aktuelle Zeit erfasst wird, ist auch eine spätere Zuordnung von Positionen und Bildaufnahmen anhand des Aufnahmezeitpunkts im Rahmen der computergestützten Weiterverarbeitung der Daten möglich. Vorteile bietet dieses Vorgehen dadurch, dass neben der geographischen Lokalisation der Aufnahme auch ergänzende Informationen zu zurückgelegten Wegstrecken zwischen einzelnen Aufnahmen verfügbar sind, die sich zumindest für Überprüfungen der Positionsangaben nutzen lassen, sofern sie an sich nicht von primärem Interesse sind.

Die hier vorgestellte SAS-Makro-Sammlung erlaubt den Aufbau einer Bilddatenbank mit Einlesen von Informationen aus EXIF-Headern von JPEG-Bilddateien sowie die automatisierte Zuordnung von Geo-Koordinaten (geographische Länge und Breite sowie Höhe) aus Track-Aufzeichnungen handelsüblicher GPS-Geräte. Alle wesentlichen Programmfunktionen und -optionen können über eine graphische Oberfläche gesteuert werden, die in SAS/AF realisiert wurde. Zudem ist ein Nutzer-Management implementiert, welches die Präferenzen und Einstellungen für unterschiedliche Nutzer automatisch speichert und bei einem späteren Programmaufruf erneut bereitstellt.

2 Programmkomponenten

2.1 Übersicht

Die Anwendung erlaubt in einem typischerweise ersten Arbeitsschritt die systematische Aufarbeitung und Überprüfung von GPS-Track-Dateien, die im GPX-Format, einem spezifischen XML-Format, vorliegen müssen. Mit einem Aufruf können automatisiert Track-Dateien aus zuvor ausgewählten Verzeichnissen in großer Zahl aufgearbeitet werden. Die Auflistung von Dateien zur nachfolgenden Abarbeitung erfolgt dabei durch ein allgemein formuliertes Makro, welches an unterschiedlichen Stellen im Programmablauf eingesetzt wird. Aufgearbeitete Track-Dateien stehen anschließend als SAS-Datendateien in einem Nutzer-abhängig festlegbaren Verzeichnis zur Verfügung.

Aus Bilddateien werden, gleichfalls optional nach Auswahl ganzer Verzeichnisse, in weiteren SAS-Programmroutinen neben diversen Aufnahmedaten auch Thumbnail-Bilder extrahiert und separat gespeichert. Je Auswahl, z.B. eines Verzeichnisses, resultiert eine SAS-Datendatei mit Informationen zu allen ausgewählten Bildern.

Für die Zusammenführung von Bildinformationen mit GPS-Daten anhand der Zeitangaben in einem weiteren Programmteil können Korrekturwerte für die kamerainterne Zeit-

erfassung sowie weitere Optionen gewählt werden. Zudem werden erfasste sowie abgeleitete Bild- und Track-Informationen in ein proprietäres KML-Format exportiert, welches bei einer Internetanbindung die Visualisierung von Tracks, Aufnahmelokalisationen und Bildern mit zugehörigen Daten im aktuell frei verfügbaren Programm Google Earth[®] und damit im Kontext hochauflösender Satellitenaufnahmen der jeweiligen Umgebung erlaubt.

Ausgewählte Aspekte zu den einzelnen Programmkomponenten sowie die Benutzeroberfläche werden nachfolgend näher vorgestellt. Weitere Informationen sind auch der Kommentierung des Programmcodes zu entnehmen.

2.2 Listing von Dateien in Verzeichnissen (%DIR_READ)

Für das Listing von beliebig eingrenzbaren Dateien aus ausgewählten Verzeichnissen in einer SAS-Datendatei ist nach Kenntnis des Autors stets ein Rückgriff auf Betriebssystem-abhängige Befehle erforderlich, welche entsprechende Informationen bereitstellen. Im Falle der Verwendung von Microsoft Windows[®] ist dies der schon aus DOS-Zeiten bekannte Befehl DIR. Dieser erlaubt optional die Ausgabe der Informationen in eine ASCII-Datei, die anschließend in SAS eingelesen werden kann. Externe Programme, wie hier der Befehl "DIR" mit einigen Optionen, lassen sich allgemein wie folgt innerhalb von SAS aufrufen (im Beispiel liest DIR Angaben zum Stammverzeichnis der Festplatte C: und schreibt das Ergebnis in die angegebene ASCII-Datei):

```
systask command "dir C:\ > C:\TEMP\datdir.txt";
```

Sofern Programme (wie DIR) Standard Input und Standard Output nutzen, besteht eine elegantere Alternative in der Verwendung von "Unnamed Pipes", die den Output ohne Erstellung einer intermediären Datei in SAS bereitstellen:

```
%LET SSTRING=C:\;           * Makro-Variable mit Verzeichnisangabe;

filename d pipe "dir ""&SSTRING."" /-C ";
data datdir;

    format text $char500.;
    infile d trunccover ;
    input @1 text $char500.;
    text=tranwrd(text,'84'x,'E4'x);      *Anpassung von Umlauten etc.;
    *u.s.w.....;
run;
```

Das Makro %DIR_READ verwendet "Unnamed Pipes" wie im Beispielcode und stellt nach Aufarbeitung der Informationen diverse Angaben zu fast beliebig selektierbaren Dateien in einer SAS-Datentabelle WORK.Datdir bereit, wobei je Tabellenzeile alle relevanten Informationen zu jeweils einer identifizierten Datei inklusive der vollständi-

gen Pfadangaben verfügbar sind. Eine allgemeine Nutzung des Makros unter Windows 2000[®] oder Windows XP[®] in deutscher Sprachversion sollte auch bei Umlauten und Leerzeichen in Datei- und Verzeichnisnamen keine Probleme bereiten, bei anderen Betriebssystemen und Sprachen ist jedoch eine Anpassung grundsätzlich erforderlich (vgl. auch Kommentierung im Programmcode).

2.3 GPS-Aufarbeitung (%gpx_parameter)

Eine wesentliche Teilaufgabe bildet die Aufarbeitung von Track-Dateien, die zuvor vom GPS-Gerät auf den Computer übertragen werden müssen. Zur Übertragung existieren in der Regel Programme der Gerätehersteller, empfehlenswert erscheinen jedoch auch Geräte-unabhängige Programme, die teilweise frei verfügbar sind [1]. Ein relevantes Kriterium für die Auswahl entsprechender Software sollte sein, dass primär im Gerät erfasste Informationen zu geographischen Koordinaten nicht unnötig durch die Rundung oder Beschneidung von Nachkommastellen beim Abspeichern von Track-Dateien reduziert werden. Als flexibles Format zum Speichern von Track-Dateien ist, vermutlich nicht nur im vorliegenden Kontext, das GPX-Format zu empfehlen. Dabei handelt es sich um ein spezifisches XML-Format, das unter Verwendung einer entsprechend angepassten MAP-Datei problemlos von SAS gelesen und in eine SAS-Datendatei geschrieben werden kann:

```
filename gpx "&GPX_DIR.\&GPX_FILE..gpx"; *Angabe der Track-Datei;
filename map "&GPX_MAP."; *Angabe der MAP-Datei inkl. Pfad;
libname gpx xml xmlmap=map; *XML-Libname;

data x;
set gpx.TRACK;
 *u.s.w.....;
run;
```

Bei anfänglicher Nutzung des XML-Formates ungewohnt erscheint, dass die über das Filename-Statement im Beispiel als "gpx" bezeichnete XML-Datei im nachfolgenden Data step im Sinne eine Library verwendet wird. Dies ergibt sich jedoch logisch aus der Tatsache, dass über eine MAP-Datei für eine XML-Datei mehrere unterschiedliche Datentabellen definiert werden können. Zum Lesen relevanter Informationen des GPX-Formates ist allerdings die Definition einer einzelnen Tabelle ausreichend. Die zugehörige Syntax, über die eine Tabelle mit dem Namen "TRACK" in der MAP-Datei definiert wird, ist auszugsweise mit Festlegung von einigen Variablen bzw. Tabellenspalten (vgl. jeweils <COLUMN> - Tags) nachfolgend dargestellt. Entscheidend für die Interpretation der Daten ist die Angabe in den <TABLE-PATH> - Tags, die festlegt, auf welcher Gliederungsebene der strukturierten Datei einzelne Beobachtungen der Datentabelle zu finden sind. Ein wesentlicher Vorteil des XML-Formates liegt darin, dass unter Beibehaltung dieser Gliederung ansonsten mehr oder weniger viele Merkmale in den Daten erfasst sein können, ohne dass deshalb Ausleseroutinen zu Basismerkmalen an-

gepasst werden müssten, wie dies beispielsweise beim Einlesen von ASCII-Daten mit fester Spaltenbreite regelmäßig der Fall wäre. Weitere Hinweise zum XML-Format in SAS finden sich in der entsprechenden Dokumentation auch im Internet [2].

```
<?xml version="1.0" encoding="UTF-8" ?>
<SXLEMAP version="1.2" name="SXLEMAP">

  <TABLE name="TRACK">
    <TABLE-PATH syntax="XPath">/gpx/trk/trkseg/trkpt</TABLE-PATH>

    <COLUMN name="SEG_NO" retain="YES" ordinal="YES">
      <INCREMENT-PATH syntax="XPath">/gpx/trk/trkseg</INCREMENT-PATH>
      <TYPE>numeric</TYPE>
      <DATATYPE>integer</DATATYPE>
      <FORMAT width="8">Z</FORMAT>
    </COLUMN>

    <COLUMN name="LATITUDE">
      <PATH syntax="XPath">/gpx/trk/trkseg/trkpt/@lat</PATH>
      <TYPE>numeric</TYPE>
      <DATATYPE>double</DATATYPE>
      <FORMAT width="15" ndec="10">Z</FORMAT>
    </COLUMN>

    <COLUMN name="time">
      <PATH syntax="XPath">/gpx/trk/trkseg/trkpt/time</PATH>
      <TYPE>numeric</TYPE>
      <DATATYPE>dateTime</DATATYPE>
      <FORMAT width="20"> IS8601DZ</FORMAT>
      <INFORMAT width="20"> IS8601DZ</INFORMAT>
      <LENGTH>20</LENGTH>
    </COLUMN>
  </TABLE>

</SXLEMAP>
```

Zu den einzelnen Track-Punkten, die nach dem Einlesen jeweils eine Beobachtung bzw. Zeile in einer SAS-Datentabelle bilden, sind neben der Zeitangabe (grundsätzlich als UTC-Angabe, also ohne Anpassung an lokale Zeitzonen) insbesondere Angaben zur geographischen Länge und Breite sowie ggf. auch zur Höhe über dem Meeresspiegel verfügbar. Einzelne Track-Dateien umfassen zur Zeit gerätebedingt üblicherweise nicht mehr als 10.000 Track-Punkte, was jedoch für eine angemessen genaue Wegaufzeichnung über einen Tag in nahezu jedem Fall ausreichen sollte. Unter Berücksichtigung von Informationen aus zeitlich benachbarten Trackpunkten lassen sich eine Vielzahl

von Parametern ableiten. Im Ablauf des Makros %gpx-parameter werden unter Verwendung von Lag-Funktionen entsprechend Distanzen bzw. Wegstrecken, Geschwindigkeiten, Bewegungsrichtungen und Beschleunigungswerte ermittelt, die nach Festlegung geeigneter Grenzwerte in einer Konfigurationsdatei auch für eine Plausibilitätsprüfung der Positionsbestimmung genutzt werden (Konfigurationsparameter finden sich in der Datei gpx_parameter_config.sas). Um Zugriffe bei größeren Track-Sammlungen zu vereinfachen, werden Track-Dateien im SAS-Format nach Von- und Bis-Zeitpunkt der jeweiligen Aufzeichnung benannt. Schließlich werden zur Visualisierung Tracks als KML-Dateien exportiert (vgl. spätere Abschnitte zum KML-Format), wobei potentiell als fehlerhaft erkannte Wegaufzeichnungen farblich hervorgehoben sind. Track-Dateien besitzen bei intelligenter Aufzeichnung (z.B. mit reduzierter Speicherung bei unveränderter Position) und vor dem Hintergrund der aktuell maximal erreichbaren Genauigkeit mit Positionsbestimmungstoleranzen im Bereich von einigen Metern ein recht übersichtliches Datenvolumen. Für einen Nutzer dürften bei weitgehend vollständiger Registrierung aller zurückgelegten Wegstrecken außerhalb geschlossener Räume in einem Jahr nur selten mehr als 1 Mio. Track-Punkte erfasst werden. Diese ließen sich in der beschriebenen Form innerhalb weniger Minuten aufarbeiten, der Bearbeitungsschritt erfordert also nur verhältnismäßig geringe Ressourcen.

2.4 Bilddaten einlesen (%picture_parameter)

Das Makro %picture_parameter erlaubt das Einlesen von Informationen aus Exif-Headern, die von nahezu allen Digitalkameras beim Speichern der Aufnahmen im häufig verwendeten JPEG-Format in die Bilddateien geschrieben werden [3]. Neben Angaben zum Aufnahmezeitpunkt finden sich im Header einer Bilddatei üblicherweise technische Angaben zu Kamera, Objektiv, Bildgröße, Blende, Belichtungszeit, verwendeter Brennweite, Fokuseinstellung u.s.w., von denen die Wesentlichsten über das Makro nach Auswahl einer zumeist größeren Zahl von Bilddateien ausgelesen werden und die anschließend aufgearbeitet in einer SAS-Datendatei verfügbar sind. Mit einem typischen Aufruf ließen sich so beispielsweise Informationen zu allen Aufnahmen eines Tages, die zuvor in einem Verzeichnis abgelegt wurden, einlesen. Im Zuge des Einlesevorgangs ist optional die Extraktion von Thumbnail-Bildern, also kleinen Vorschaubildern, möglich. Eine Vorläuferversion dieses Makros wurde bereits auf der KSFE 2003 vorgestellt [4], weshalb auf Details der relativ umfangreichen Programmierung an dieser Stelle nur mit zwei Beispielen eingegangen werden soll.

```
data d&R;    *Einlesen langer Zeichenstrings in nummerierte Dateien;
  infile "&&DATNAME&R " lrecl=65519 recfm=f truncover obs=1;
  input @1 header $ASCII32765. @32754 header_ext $ASCII32765.;
run;
```

Wesentliche Voraussetzung für die Aufarbeitung der Bilddaten in der hier realisierten Form ist die Fähigkeit von SAS, auch Zeichenstrings mit einer extremen Länge von bis zu 32.767 Byte in eine Variable zu lesen, welche anschließend ohne erneuten Zugriff

auf die Originaldatei aufgearbeitet werden können (vgl. zuvor aufgeführtes Syntaxbeispiel zum Einlesen zweier überlappender Strings in die Variablen "header" und "header_ext" mit jeweils 32.765 Byte Länge in eine über die Makro-Variablen &R nummerierte SAS-Datei d&R aus Bilddateien, deren Bezeichnung sich in der nummerierten Makro-Variablen &&DATNAME&R befindet). Das Herausschreiben eines Thumbnail-Bildes ist nach der Identifikation und Auswahl des entsprechenden Abschnitts des Strings über eine Substring-Funktion im selben Data step relativ simpel (Syntax hier ohne die im Original eingesetzten Variablen- und Makrovariablenbezeichnungen):

```
file "THUMBNAILDATEI_NAME.JPG"  recfm=f  lrecl=1;
thumbnail=substr(SOURCE_STRING,ERSTES_BYTE,ANZAHL_BYTES);
put thumbnail;
```

Die Extraktion der Bildinformationen erfordert einige Ressourcen, zumal zwischenzeitlich Header-Daten in Form der langen Zeichenstrings zu allen ausgewählten Bildern innerhalb einer Datei verarbeitet werden. Die Verarbeitung mehrerer hundert Bilder in einem Aufruf ist nach eigenen Erfahrungen unproblematisch, der Zeitbedarf liegt im Minutenbereich. Mit sehr hohen, 5stelligen Bildzahlen liegen keine Erfahrungen vor.

2.5 Zusammenführung (%picture_gps_merge)

Die Zusammenführung von Bild- und GPS-Daten erfolgt im Rahmen des Makros %picture_gps_merge mehrstufig, wobei Zuordnungen für alle in einer SAS-Datendatei erfassten Anzahl von Bildern in einem Aufruf vorgenommen werden (sofern GPS-Angaben aus dem zeitlichen Umfeld der einzelnen Aufnahmen vorliegen). Zunächst wird der Gesamtzeitraum bestimmt, in welchem die Bilder entstanden sind. Anschließend werden alle aufgearbeiteten Track-Dateien mit Relevanz für diesen Zeitraum in einer Datei zusammengeführt. Über Lag-Funktionen werden Trackangaben zu vier weiteren Zeitpunkten in jeder Tabellenzeile (redundant) bereitgestellt. Über eine Makro-Schleife wird eine Datei generiert, die für jedes einzelne Bild Angaben zu allen zeitlich potenziell relevanten Trackpunkten enthält (wobei Angaben, die für mehrere Bilder relevant sind, entsprechend mehrfach als Zeilen, also nochmals redundant, in der Datei vorkommen). Nachfolgend werden über Proc means mit einer Bildidentifikationsnummer als class-Variablen die zeitlich nächstliegenden Trackpunkt-Angaben (inklusive der Angaben zu vier weiteren Punkten im zeitlichen Umfeld) in einer Tabelle bereitgestellt, die entsprechend nur noch eine Zeile je Bild enthält. Aus den Trackpunkt-Angaben im zeitlichen Umfeld der Bildaufnahme werden ergänzende ableitbare Maße berechnet. Bei Bildaufnahmen zwischen zwei Trackaufzeichnungszeitpunkten werden bei Einhaltung bestimmter Kriterien die geographischen Angaben zur Bildlokalisierung interpoliert.

Ein zweiter Teil des Makros dient der Erstellung einer KML-Datei. Dabei handelt es sich um ein spezifisches XML-Format, welches die Darstellung von geographisch zugeordneten Informationen in Google Earth[®] erlaubt [5]. Die vom Makro erstellten KML-Dateien zeigen Bildlokalisierungen und zugehörige Track-Verläufe. Über eine Auswahl der Bildmarkierungen im Kontext der von Google[®] bereitgestellten Satellitenaufnahmen

werden Thumbnails einzelner Bilder mit ausgewählten Informationen gezeigt, die Aufnahmen in voller Auflösung können über einen Link aufgerufen werden (vgl. Abbildung 1).

Die Erstellung der KML-Datei erfolgt im Makro, unabhängig von spezifischen XML-Funktionen in SAS, durch die Ausgabe der gewünschten Textstrings über entsprechende Put-Anweisungen in eine ASCII-Datei. Dabei lassen sich konstante Textelemente wie KML-Tags recht flexibel mit Werteangaben aus SAS-Datentabellen in unterschiedlichen Abfolgen und Umfängen kombinieren und Tags in Abhängigkeit von Variablenwerten bei Bedarf modifizieren (vgl. Programmcode und erstellte KML-Dateien).

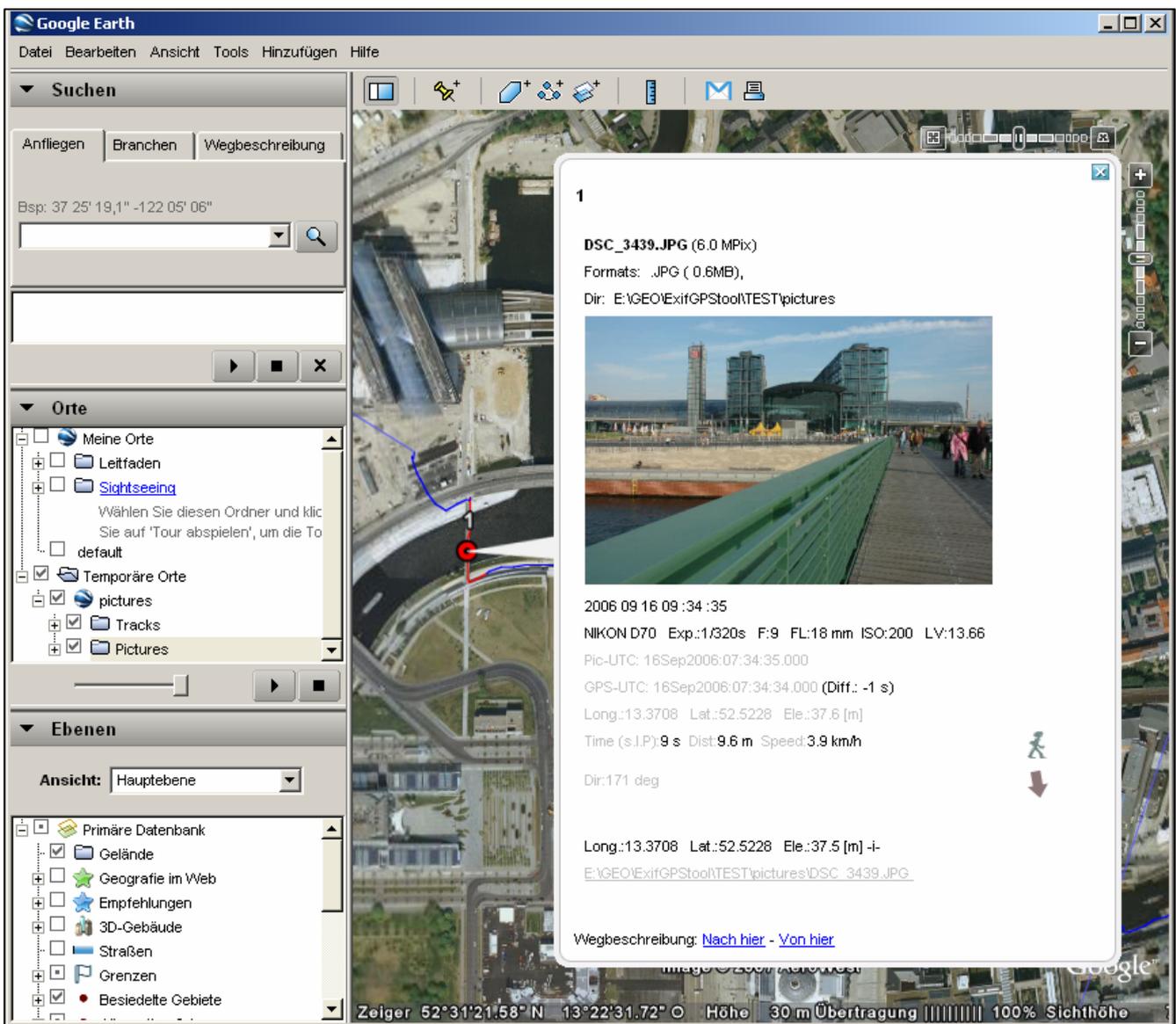


Abbildung 1: Darstellung von geographisch zugeordneten Bildern in Google Earth®

3 Benutzeroberfläche

Der Aufruf der zuvor dargestellten Programmkomponenten bzw. Makros könnte grundsätzlich, wie üblicherweise praktiziert, über entsprechende Programmzeilen im Editor erfolgen. Allerdings wäre die Formulierung des Aufrufs in Anbetracht der relativ umfangreichen Optionen verhältnismäßig fehlerträchtig und wäre kaum ohne die intensive Auseinandersetzung des Anwenders mit einzelnen Funktionen der Makros möglich. In einer entsprechenden Situation kann eine graphische Benutzeroberfläche für Anwender, die lediglich vorgegebene Funktionen von Makros nutzen wollen, einen erheblich vereinfachten Zugang bieten.

Graphische Benutzeroberflächen lassen sich in SAS relativ komfortabel und mit einer umfangreichen Zahl von unterschiedlichen Elementen in SAS/AF gestalten (während zur Programmierung der Oberflächen eine Lizenz für SAS/AF vorhanden sein muss, ist diese Lizenz auf dem Rechner des Anwenders nicht erforderlich). Im nachfolgenden Text vorgestellt wird die Benutzeroberfläche zur Steuerung der bereits besprochenen Programmkomponenten zum Geocoding. Ziel ist dabei auch, durch ein unter SAS in einer Windows-Umgebung bereits lauffähiges Programmbeispiel Anregungen für eine eigene SAS/AF-Programmierung zu liefern. Einige Besonderheiten bei der Verwendung von SAS/AF und der zur Einbindung erforderlichen SAS Component Language (SCL) werden im Text aufgegriffen. Weitere Hinweise auf konkrete Umsetzungsmöglichkeiten sind dem Quellcode der Anwendung zu entnehmen.

3.1 Hinweise zum Einstieg in SAS/AF

Einen guten ersten Einstieg in die Programmierung von Benutzeroberflächen unter SAS/AF bietet ein auf den SAS-Seiten im Internet verfügbares Dokument [6]. Neben diesem Dokument sind eine Reihe weiterer Informationen online verfügbar [7, 8]. An dieser Stelle soll lediglich eine Kurzanleitung für einen ersten Nutzeroberflächen-Programmierungsversuch unter SAS/AF mit Hinweisen auf einige Besonderheiten geliefert werden.

Als erster Schritt empfiehlt sich die Festlegung eines Verzeichnisses für die Versuche über ein entsprechendes Libname-Statement, z.B. für Versuche im Verzeichnis C:\test mit der SAS-internen Bezeichnung TESTLIB über die Programmzeile:

```
libname TESTLIB "C:\test";
```

Anschließend kann bereits die erste Oberfläche bzw. der erste Frame durch Eingabe eines BUILD-Befehls in SAS erstellt werden:

```
BUILD TESTLIB.SCLTEST.FRAME_1.FRAME
```

Im Befehlsbeispiel wird im Katalog SCLTEST, der in TESTLIB gespeichert wird, ein Frame mit der Bezeichnung Frame_1 erstellt, oder, falls bereits vorhanden, zur Bearbeitung geöffnet. Neben dem anfänglich leeren Frame-Fenster sollte nach Ausführung des Befehls gleichfalls ein Fenster mit potenziellen Frame-Komponenten geöffnet werden, die nach Belieben mit der Maus in den Frame gezogen und dort hinsichtlich ihrer Platzierung und Größe modifiziert werden können. Die verwendeten Elemente werden zunächst automatisch nach ihrer Typenbezeichnung mit fortlaufender Nummerierung innerhalb des Frames benannt, bei Verwendung vieler Elemente empfiehlt sich jedoch

eine Umbenennung unter Verwendung sprechender Namen. Eigenschaften bzw. Attribute des Frames und der platzierten Elemente inklusive deren Namen können in einem Fenster modifiziert werden, welches sich nach Klick mit der rechten Maustaste und Auswahl von "Properties" bzw. Eigenschaften öffnet. Bei der alternativ dabei gleichfalls möglichen Auswahl von "Frame SCL" kann der dem Frame zugeordnete SCL-Code in einem separaten Fenster zunächst erstellt oder nachfolgend weiter bearbeitet werden.

Der SCL-Code legt fest, welche Aktionen beim Öffnen oder Schließen des Frames oder bei der Bedienung von bestimmten Elementen des Frames, ggf. auch eingabeabhängig, ausgeführt werden, wobei die entsprechenden Code-Abschnitte durch die Angaben `INIT:`, `Term:` bzw. `Elementname:` eingeleitet und mit `return;` abgeschlossen werden. Code in einem mit `Main:` eingeleiteten Abschnitt wird bei jeglicher Interaktion des Anwenders mit der Oberfläche ausgeführt.

Gewöhnlicher SAS-Programmcode, wie auch der Aufruf oder die Definition eines Makros, muss bei der Angabe innerhalb des SCL-Codes mit `submit continue;` eingeleitet und mit der Angabe von `endsubmit;` abgeschlossen werden. Hingewiesen sei an dieser Stelle abschließend auf die erforderliche SCL-Syntax, um Werte von Makro-Variablen an Frame-Variablen (z.B. für eine Anzeige im Frame) weiterzugeben:

```
framevariable = symget("MAKROVARIABLE");
```

Weitere Besonderheiten sollten sich beim Durcharbeiten von Beispielprogrammen erschließen.

3.2 Benutzeroberfläche zum Geocoding

Die Benutzeroberfläche besteht aus insgesamt drei Fenstern bzw. Frames, die sequenziell nacheinander aufgerufen und in umgekehrter Reihenfolge wieder geschlossen werden, dem Startfenster, dem Nutzerauswahlfenster sowie dem Hauptanwendungsfenster. Vor der ersten Nutzung (oder Veränderung) der Oberfläche muss die im Programmverzeichnis befindliche Datei `CONFIG_lib_gps_scl.sas` durch die Angabe des Programmverzeichnisses einmalig angepasst und anschließend ausgeführt werden. Bei regelmäßiger Nutzung empfiehlt sich die Erstellung einer Desktopverknüpfung, über die ein direkter Aufruf der Benutzeroberfläche möglich ist. Details hierzu sind in der zuvor genannten Konfigurationsdatei beschrieben. Der zur Oberfläche gehörige SCL-Code inklusive der Dateien bzw. Informationen zu Nutzern (s.u.) ist im Unterverzeichnis "SCL" des Programmverzeichnisses abgelegt, welches durch die Konfigurationsdatei als Library "Gps_scl" definiert wird.

Ein Doppelklick auf den in "Gps_scl" enthaltenen Katalog "Exif_gps" ermöglicht den Zugang zum Start der Oberfläche ("Exif_gps_start" markieren, rechte Maustaste, "Run" wählen) oder zur Modifikation der Oberflächenelemente (durch Doppelklick auf das entsprechende Symbol für Frame oder SCL-Code).

3.2.1 Startfenster

Das Startfenster beinhaltet lediglich zwei Buttons, die entweder den Aufruf des nachfolgenden Fensters oder ein Verlassen des Programms erlauben. Die im Hintergrund ab-

laufende Funktion dieses Fensters besteht darin, eine Kopie der Nutzerdatei zu erstellen, die im nachfolgenden aufrufbaren Fenster bearbeitet werden kann.

3.2.2 Nutzerauswahlfenster

Über dieses Fenster ist die Auswahl eines bereits bestehenden oder die Festlegung eines neuen Nutzers möglich, der anschließend in der Nutzerdatei erfasst wird. Zu jedem neuen Nutzer wird automatisch eine separate Datendatei mit Nutzerpräferenzen angelegt, die zunächst Default-Werte beinhaltet und die ansonsten die jeweils beim letzten Programmaufruf durch den Nutzer gewählten Einstellungen im Hauptanwendungsfenster erneut bereitstellt. Das Listing zur Auswahl eines Nutzers erfolgt allgemein in alphabetischer Reihenfolge, wobei, abweichend von dieser Regel, der Nutzer beim letzten zeitlich vorausgehenden Aufruf des Hauptanwendungsfensters an erster Stelle genannt wird. Die zugehörigen Programmroutinen sind im SCL-Code des Nutzerauswahlfensters implementiert.

3.2.3 Hauptanwendungsfenster

Im Hauptanwendungsfenster sind alle nutzerspezifisch relevanten Funktionen zugänglich. Die Anordnung der einzelnen Elemente von oben nach unten entspricht weitgehend dem typischen Arbeitsablauf (vgl. Abbildung 2). Bei erstmaligem Aufruf durch einen neu spezifizierten Nutzer sollten zur Sicherstellung des regulären Programmablaufes zunächst zwei separate Verzeichnisse zur Ablage von aufgearbeiteten GPS- bzw. Bilddaten ausgewählt werden. Eine Auswahl – und bei Bedarf auch Neuerstellung – von Verzeichnissen ist nach Klick auf die Buttons "GPS SAS Dir" bzw. "PIC SAS Dir" möglich. Diese beiden Verzeichnisse enthalten später alle Resultate von Programmabläufen und sollten aus Gründen der Übersichtlichkeit und Datensicherheit keine anderweitig erstellten Dateien enthalten.

Anschließend können typischerweise zunächst GPS-Daten im GPX-Format zur Aufarbeitung ausgewählt werden. Über entsprechende Buttons ist entweder die Auswahl einzelner Dateien oder die Auswahl ganzer Verzeichnisse, optional mit Berücksichtigung von Dateien aus Unterverzeichnissen, möglich. Nach einer Auswahl wird der Einlesevorgang über den Button "Import sel. GPX Data" gestartet. Für die Plausibilisierung während der Einleseroutine können unterschiedliche Grenzwerte ausgewählt werden, die in einer Konfigurationsdatei abgelegt sind, welche nach Klick auf den Button "Edit GPX imp. config" im Windows-Programm Notepad editiert werden kann. Nach dem erfolgreichen Einlesen sollten aktualisierte Basisangaben zu den insgesamt verfügbaren aufgearbeiteten GPS-Daten im rechten Teil des Fensters erscheinen.

Für das Einlesen von Bilddaten lassen sich analog Verzeichnisse mit entsprechenden Dateien auswählen. Je Auswahl wird eine Datendatei mit Bildinformationen erstellt, die automatisch den Verzeichnisnamen erhält (weshalb ausgewählte Verzeichnisse immer eindeutige Benennungen tragen sollten, da andernfalls bereits erstellte Dateien überschrieben werden). Der Einlesevorgang wird durch Klick auf den Button "Read PIC Info" gestartet. Auch nach diesen Einlesevorgängen werden aktualisierte Statistiken in der rechten Fensterhälfte bereitgestellt.

Die Bezeichnung der zuletzt erfolgreich eingelesenen Datendatei mit Bildinformationen erscheint als Voreinstellung im Auswahlfeld im unteren Drittel des Hauptanwendungsfensters. Nach Klick auf den Button "Merge GPS+PIC Info" wird die Datendatei um ggf. verfügbare GPS-Informationen aus dem entsprechenden SAS-GPS-Datenverzeichnis ergänzt, wobei vorausgesetzt wird, dass verfügbare GPS-Daten bereits zuvor aufgearbeitet wurden.

Für die Zusammenführung lassen sich unterschiedliche Parameter spezifizieren. "Use track time... before / after" ermöglicht die Spezifikation des maximalen zeitlichen Ranges um den Aufnahmezeitpunkt (in Sekunden), in den GPS-Positionsbestimmungen fallen müssen, um beim Geotagging und der Trackdarstellung ggf. noch berücksichtigt zu werden. Die Eingabe von "Interpol. limits" ermöglicht die Angabe des maximalen zeitlichen Abstandes zweier Positionsbestimmungen vom Aufnahmezeitpunkt, bei dem die geographische Position der Aufnahme aus beiden Punkten interpoliert wird. Die Option "Validated GPS only" führt zur ausschließlichen Verwendung von als valide eingestuft Trackpunkten bei der Positionszuordnung.

Substanziell für eine korrekte Zuordnung ist bei den Angabemöglichkeiten in der Regel jedoch ausschließlich eine möglichst genaue Angabe im Feld "PIC – UTC deviation", in welchem die Differenz der Kamera-intern erfassten Zeitangabe (üblicherweise Ortszeit) zur UTC-Zeit, die von einem GPS-Gerät gespeichert wird, angegeben werden muss. An dieser Stelle lassen sich ggf. auch kleinere, nachträglich festgestellte, Ungenauigkeiten der Kamera-internen Zeitangaben ausgleichen. Zu beachten ist, dass in einem Zuspielvorgang immer nur eine einheitliche Korrekturangabe verwendet werden kann. Insofern müssen Bilder nach jeder Veränderung der Kamerazeiteinstellung, z.B. auch bei Wechsel von Sommer und Winterzeit, separat eingelesen werden. Optional werden bereits beim Einlesen der Bilder Angaben zur Zeitabweichung berücksichtigt und im Label der SAS-Datei mit Bilddaten gespeichert, sofern diese in einer Textdatei mit dem Namen PIC.UTC.TIME im Originalbildverzeichnis abgespeichert sind. Sollen diese ggf. vorhandenen Angaben später überschrieben werden, muss dies explizit durch die Auswahl der entsprechenden Option angegeben werden.

Die Statistik zu Bilddateien gibt nach der Zusammenführung Auskunft, bei welchem Anteil der Dateien im verwendeten Verzeichnis aktuell bereits eine Zusammenführung von Informationen vorgenommen wurde.

Über den Button "View in Google Earth" wird schließlich bei Bedarf die im Rahmen der Zusammenführung im Sinne eines Geotaggings gleichfalls erstellte KML-Datei aufgerufen und in Google Earth[®] – sofern dies Programm auf dem genutzten Computer installiert ist – angezeigt.

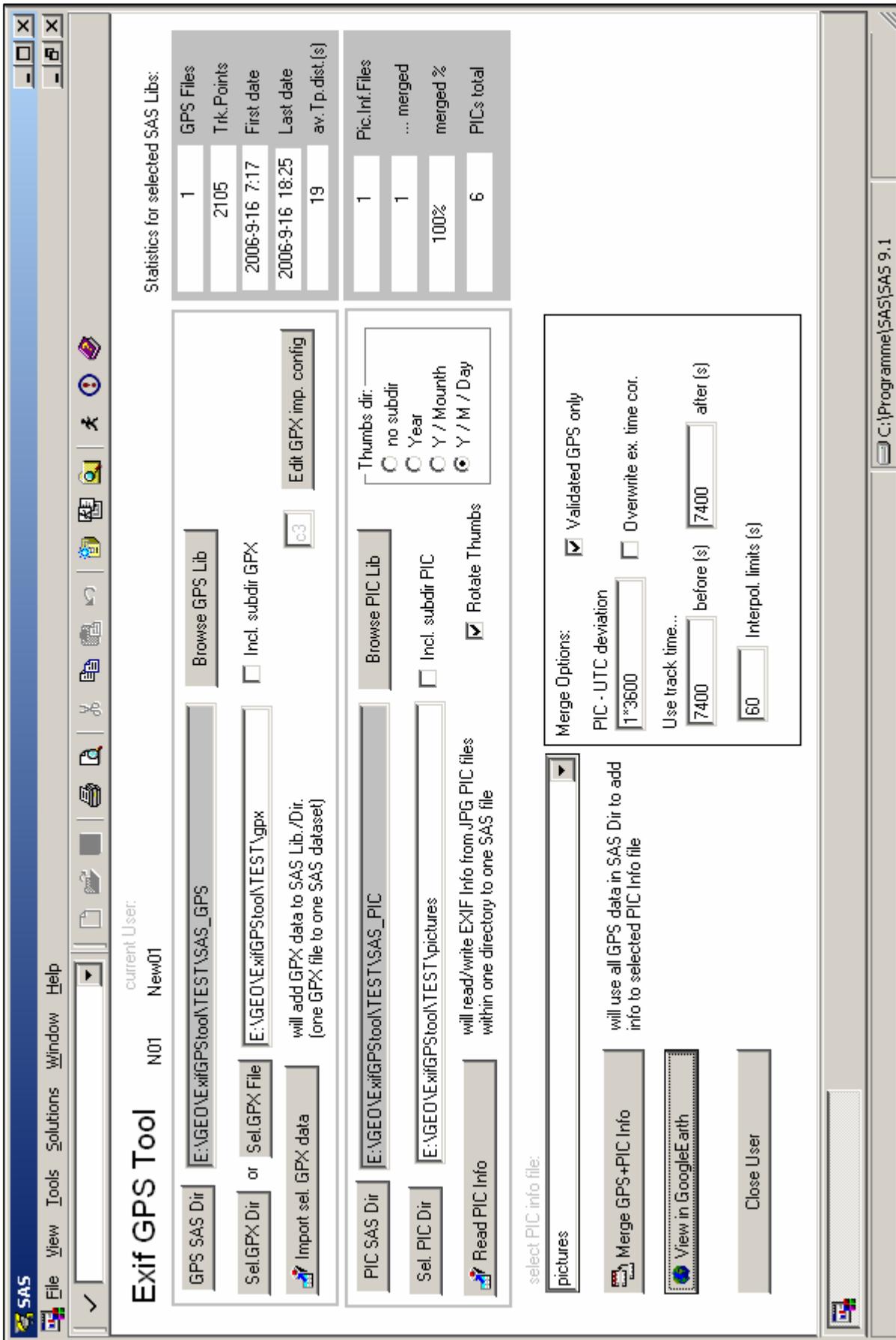


Abbildung 2: Hauptanwendungsfenster

4 Resümee

Mit dem vorliegenden Beitrag sollten Anregungen für den Umgang mit Daten in unterschiedlichen Formaten aus unterschiedlichen Quellen sowie für Einsatzmöglichkeiten einer graphischen Benutzeroberfläche in SAS gegeben werden. Der Autor hofft, dass dies zumindest aus der Sicht einiger Leser gelungen ist. Zweifellos konnten im Rahmen des Textes nur kleine Bruchstücke der verwendeten Programmsyntax vorgestellt werden. Da die Syntax jedoch in vollständigem Umfang verfügbar ist, sollte es möglich sein, eine Reihe von Fragen nach der technischen Umsetzungen von Funktionen in der Makrosammlung als Nutzer selbst auf der Basis des Programmcodes zu beantworten.

Die durch die Makrosammlung nebst Oberfläche bereitgestellten Funktionen zum Geotagging dürften in dem jetzt realisierten Umfang für eine Reihe von konkreten Anwendungen bei Verwendung von geeigneten Betriebssystemen bereits ausreichen und insbesondere relativ einfach nutzbar sein. Bei Wünschen nach erweiterten Funktionen oder der Anwendung entsprechender Programmabläufe unter anderen Betriebssystemen bliebe die Möglichkeit, Programmcode partiell oder modifiziert in eigenen Anwendungen zu nutzen.

Das Thema Geotagging bzw. Geocoding allgemein dürfte innerhalb der nächsten Jahre sicherlich insgesamt noch an Bedeutung gewinnen.

Literatur

- [1] Programmempfehlung: GPS TrackMaker[®] <http://www.gpstm.com/>
- [2] SAS 9.1.3 XML LIBNAME Engine: User's Guide
http://support.sas.com/documentation/onlinedoc/91pdf/index_913.html
- [3] Informationen zum Exif-Format: <http://www.exif.org/>
- [4] T. G. Grobe: Extraktion von Informationen aus binären Dateien: Ein SAS-Makro zur Erfassung von strukturierten Meta-Informationen aus Bilddateien aktueller Digitalkameras.
In C. Becker, H. Redlich (Hrsg.): Proceedings der 7. KSFE. Shaker Verlag, 2003
- [5] Informationen zum KML-Format (Google Earth[®]):
<http://code.google.com/apis/kml/documentation/>
- [6] Getting Started with SAS/AF and Frames
http://support.sas.com/documentation/onlinedoc/91pdf/index_913.html
- [7] SAS Component Language 9.1: Reference
http://support.sas.com/documentation/onlinedoc/91pdf/index_913.html
- [8] SAS/AF 9.1 Procedure Guide
http://support.sas.com/documentation/onlinedoc/91pdf/index_913.html

Anhang A: Installationshinweise – Kurzanleitung

Wie bei jeder Ausführung von Programmen sollten vor der Anwendung der Makrosammlung Maßnahmen zum Schutz vor Datenverlusten ergriffen werden. Die Nutzung der bereitgestellten Programme erfolgt ausschließlich in eigener Verantwortung des jeweiligen Nutzers! Insbesondere sollten sich keine relevanten Dateien aus anderen Programmabläufen im SAS-Work-Verzeichnis befinden!

Die Makrosammlung sollte grundsätzlich in SAS 9.1 bei Betrieb unter Windows 2000® oder Windows XP® (Betriebssystem in deutscher Sprachversion) lauffähig sein. Für andere Windows- oder Sprachversionen sind Anpassungen im Makro %DIR_READ (insbesondere zum Lesen der Datei-Listings) sowie in %picture_parameter erforderlich (relevante Stellen finden sich im Makro-Code bei der Suche nach "systask" – an entsprechenden Stellen werden Betriebssystemabhängige Befehle vorrangig zum Erstellen von Verzeichnissen aufgerufen. Zwischen Verzeichnis- und Dateinamen werden im Makroablauf an unterschiedlichen Stellen "\"-Zeichen gesetzt – sofern dies nicht Betriebssystem-konform ist, sind umfangreiche Anpassungen erforderlich. Leerzeichen in Datei- oder Verzeichnisnamen können allgemein zu Fehlfunktionen führen und sollten unbedingt bei der Benennung des Programmverzeichnisses sowie der Bildquellverzeichnisse vermieden werden, letzteres, da die Verzeichnisnamen zur Benennung von SAS-Datentabellen verwendet werden!

Installationsanleitung:

- Entpacken der Datei ExifGPStool.zip unter Beibehaltung der Unterverzeichnisse in ein beliebiges Verzeichnis auf einer Festplatte.
- Obligat erforderliche Anpassung der Pfadangabe zur Programmlokalisation in der Datei CONFIG_lib_gps_scl.sas, die sich im Stammverzeichnis des Programms befindet. Die Datei beinhaltet weitere Hinweise zum Start des Programms sowie zur optionalen Erstellung einer Desktop-Verknüpfung zum direkten Programmstart.
- Zur (verlustfreien) Rotation von extrahierten Thumbnail-Bildern wird im Makro-Ablauf das Programm jpegtran.exe der "Independent JPEG Group" (IJG) genutzt, welches frei verwendet werden kann. Weitere Informationen finden sich im Internet unter <http://www.iijg.org> sowie in der Datei jpegtran.txt im Programmverzeichnis.
- Für Probeläufe findet sich im Programmverzeichnis ein Unterverzeichnis TEST mit vier weiteren Unterverzeichnissen: Das Verzeichnis "gpx" beinhaltet eine Track-Datei im GPX-Format. Im Verzeichnis "Pictures" finden sich exemplarische Bilddateien sowie u.a. eine Datei PIC.UTC.TIME.txt mit einer Angabe zur Korrektur der Bildzeit versus Koordinierte Weltzeit UTC in Sekunden, die für alle Bilder aus dem Verzeichnis einheitlich verwendet wird (Bildzeit – UTC). Die leeren Verzeichnisse SAS_GPS und SAS_PIC können beim Probelauf nach einer entsprechender Auswahl im Haupteingabefenster als Speicherorte für aufgearbeitete SAS-Datendateien mit GPS-Track-Daten bzw. mit Bildinformationen verwendet werden.