

## ДИГИТАЛИЗАЦИЯ НА ПОВРЕДЕНИ РЪКОПИСИ

Дигитализацията на писменото наследство може да преследва много различни цели: от простото опазване на ръкописа до неговия анализ и издаване. Благодарение на компютъра и съответния софтуер дигиталните „изображения“ могат да се обработват, т.е. да се подобрява качеството им и резултатите да се използват за по-нататъшен анализ на документите, например за оптично разпознаване на буквите (Optical Character Recognition) или за изследване на буквите. Това е основното предимство спрямо аналоговата ера, която също произведе – понякога дори по-хубави – снимки.

В зависимост от състоянието на обекта днес се използват различни начини за обработка на изображенията. За добре запазени ръкописи като старата част на Зографското евангелие в РНБ е достатъчен и „нормален“ цифров фотоапарат – и е за предпочитане, защото „цветните“ снимки, направени с него, съответстват на спектрите на абсорбция на човешкото око – така нареченото трихромно зрение, червено, зелено, синьо – на кратко RGB<sup>1</sup>.

От друга страна, ако ръкописът е повреден и текстът му не се разчита, RGB апаратът и цветните му снимки няма да свършат работа. В такъв случай първият избор ще бъде фотоапарат с по-широк спектрален обхват или с обхват различен от този на способността на очите ни. Както е известно, избледнял текст обикновено се вижда най-добре на ултравиолетови флуоресцентни лампи, защото пергаментът става флуоресцентен на ултравиолетова светлина, а желязно-галовото мастило отслабва отразената УВ светлина. Обратното е в сила за потъмнели (най-вече от въглен) текстове, написани с карбоново мастило. Те се виждат най-добре под инфрачервено осветление. Въпреки това е установено, че за различните обекти и цели най-добре е да се избират различни тесни спектрални ленти, тъй като те могат да се комбинират, за да се извлече възможно най-много информация. Наричаме това мултиспектрална обработка на изображенията или MSI.

Да поясним, когато разделяме електромагнитния спектър на малки ленти, както се прави първоначално за заснемане на отдалечени и разположени в пространството изображения, ние получаваме в зависимост от броя на лентите или мулти- или хиперспектрални изображения, в про-

---

<sup>1</sup> За разлика от много животни, хората и някои други бозайници са развили трихромно зрение, т.е. тяхната ретина съдържа три типа цветови рецептори.

тивовес на панхроматичното заснемане, което запечатва цялостния интензитет на лъчението върху всеки пиксел. Всъщност най-новият тип е ултраспектралното заснемане, но само за отдалечени обекти.

Така от множество различни камери – видео или CCD (Charge-coupled Device), инфрачервени камери и инструменти, способни на мулти- или хиперспектрална обработка на образи – най-добрият вариант е мултиспектрална камера с обхват от ултравиолетово до почти инфрачервено или, в цифри, от около 300 до около 1000 нанометра. Въпреки това е трудно да се каже точно колко и какви ленти са необходими, за да се получи максимум информация. Естествено, качеството и количеството на тези ленти зависи отчасти от обекта, отчасти от техниките за обработка. Но са нужни много опит и преценка, за да се направи правилен избор. Същото се отнася и за въпроса как да се получат най-добре спектралните ленти, т.е. с помощта на специални филтри или чрез осветяване.

Основното предимство на дигиталните изображения е възможността да бъдат увеличавани на екрана на компютъра. В много случаи това помага на филолозите да разчетат някой пасаж или да открият нещо неоткриваемо с просто око в ръкописа или снимката с нормални размери. Но и нерядко остават случаи, които налагат специална употреба на микроскоп, като редовете на текста на пренаписан ръкопис. В тези случаи събирането на данни може да се извърши чрез CCD или мултиспектрална камера, монтирана на оптичния път на микроскопа.

Накрая, ако никой от описаните начини не ви позволява да прочетете измит или изтрит текст, все още има (макар и малка) надежда. Нещата стоят така, когато текстът е написан с вещество като споменатото преди малко желязно-галово мастило върху метална основа. Металните остатъци от мастилото могат да се открият с рентгенови лъчи или синхротронно лъчение, като в момента текат първите опити за работа с тези нови средства. Химиците в нашата лаборатория във Виена вече са конструирали рентгенов флуоресцентен скенер за тази цел. В момента той може да се използва само във вертикално положение за разлика от удобния XRF-спектрометър, който може да се обръща на обратно.

Нека накратко ви представя нашата лаборатория. Преди година основахме така наречения Център за анализ на изображенията и материалите в културното наследство (СИМА). Той е последният от поредица проекти, които започнаха малко след началото на новото хилядолетие с основаването на така наречения виенски археографски форум. Подобно на двата предходни проекта, посветени на глаголическото наследство на планината Синай, той е интердисциплинарен и обединява три научни дисциплини: филология, компютърни науки и химия (на материалите). Както се вижда от заглавието на въпросния проект, ние не се занимаваме само с компютърни изображения, а и с химически анализ на писменото култур-

но наследство. За тези цели използваме всички модерни инструменти за изследване на ръкописи и когато се наложи, разработваме и нови. Инструментите ни трябва да отговарят на четири изисквания: приложението им трябва да бъде недеструктивно и неинвазивно; трябва да бъдат здрави и леснопреносими, за да могат да бъдат транспортирани до всяко хранилище и да бъдат възможно най-здравеоинсталирани и прилагани.

Работата по компютърните изображения на СИМА е съсредоточена във Виенския технологически университет и може да бъде отнесена към две изследователски полета: обработка на мултиспектрални изображения и дигитална обработка на изображения. Сега само ще очертаем тези части на програмата, които засягат мултиспектралните изображения и подобряването на качеството на изображенията.

Мултиспектралното заснемане (MSI) може да бъде описано като техника, която произвежда изображения на една сцена в многобройни теснолентови спектрални обхвати. Чрез употребата на удължен светлинен спектър в сравнение с човешкото око (видимият спектър е приблизително 380–700 nm) и възможността за различни комбинации MSI доказва, че може да увеличи контраста на разрушени или избледнели текстове в сравнение с осветяването с обикновена бяла светлина.

За да можем да изследваме ръкописите на различните места, на които се съхраняват, ние създадохме преносима MSI система. Тя се състои от две камери, два осветителни панела с LED лампи и други аксесоари. Камерите са традиционна трихромна (RGB) камера и мултиспектрална камера. Първата е Nikon с резолюция 4928 x 3280 пиксела. Тя се използва, за да се снимат видими на светлината и УВ флуоресцентни изображения. Първите, цветните изображения, служат за многобройни цели като кодикологически и полеографски изследвания или факсимилни издания. За мултиспектралните изображения използваме черно-бяла CCD камера Hamamatsu с резолюция 4000 x 2672 пиксела. Спектралната чувствителност на тази втора камера се простира от УВ почти до ИЧ обхват, а именно от 330 до 1000 nm.

За особени случаи, например за ръкописи, написани с карбоново мастило, в Института за наука и технология на Академията за изящни изкуства има допълнителна камера. Тази камера с красивото име Озирис има чувствителност между 900 и 1700 nm и резолюция 4096 x 4096 пиксела.

От двете страни на камерата са монтирани първо мултиспектралните LED панели, на които са прикачени четирите бели LED панела. Още два дифузера са монтирани между осветителните тела и обекта, така че осветяването да бъде равномерно. Изследваните обекти се поставят на дъска, която е прикрепена за специална стойка. Стойката позволява автоматично преместване на обекта между двете камери. Целият монтаж

(който обикновено включва и черна тента около системата за снемане на изображения) отнема около два часа.

Мултиспектралните LED панели произвеждат 11 различни дължини на вълните от 365 до 940 nm.

В първия ни проект се използваше „нормално“ осветление, а мултиспектралният обсег се постигаше чрез филтри, които се задвижваха автоматично на едно колело, монтирано през камерата Namamatsu. Сега употребата на LED панелите с различна дължина на вълните обикновено прави филтрите и колелото с филтрите ненужни. Въпреки това в два случая такива оптични филтри продължават да се използват: за УВ флуоресцентни изображения (400 nm нискочестотен филтър) и за УВ рефлектография (365 nm високочестотен филтър). Това се дължи на факта, че и в двете техники се използва едно и също УВ лъчение, но трябва да се заснемат различни свойства на светлината. Когато УВ лъчението стигне обекта, лъчите се отразяват, предават или поглъщат. Докато при УВ рефлектографията важни са отразените лъчи (те се регистрират от филтъра), при флуоресцентната фотография от камерата (филтъра) се регистрират само тези лъчи, които се поглъщат от обекта, след което се отразяват като видима светлина.

Затова в момента за една страница се заснема поредица от най-малко 12 изображения с мултиспектралната камера и още две изображения с трихромната (RGB) камера, или общо 14 образа.

Нашите американски колеги, които работят по така наречения проект EMEL, използват друго решение. Те използват двойни LED лампи и филтри, но само една мултиспектрална камера, с която обикновено правят по над 30 снимки на страница, включително снимки с препредадена и непряка светлина. Въпреки че последните със сигурност са необходими за специални цели, съмнително е доколко има значима полза от голямото количество изображения в приблизително еднакъв спектрален диаметър (приблизително от 340 до 940 nm). Спестяването на RGB камерата има това предимство, че не се налага обектът да бъде местен между двете камери, а образите да се регистрират след това. (След малко ще се върна на това.) От друга страна, те не могат да направят „нормални“ цветни снимки, а трябва да направят съставни образи от трите канала за червено, зелено и синьо.

Сега да се върнем на обработката на изображенията.

След заснемането на мултиспектралните изображения се прилагат няколко стъпки за последваща обработка: най-напред образите се наслагват един върху друг. След това се прилагат техники за подобряване на изображенията, за да се увеличи видимостта на буквите.

Ключовата дума е наслагане: изображенията, заснети с различни камери и филтри, не са подравнени едно към друго поради геометрични

изкривявания. Така че, за да се продължи със стъпките по последващата обработка, е необходимо правилно подравняване на различните източници на изображения, което се нарича регистрация. За тази цел екипът ни разработи специален алгоритъм за регистрация, който се основава на характеристиките на изображенията.

Подобряване на изображенията: Често мултиспектралните изображения не са достатъчни за разчитането на повредени ръкописи или палимпсести. Някои части на ръкописите остават нечетими, често поради „шум“ (фонов шум). Тогава на помощ за реконструиране на тези части идват методи на обработка като редуция на измеренията. Тези техники се използват, за да се редуцира третото измерение на мултиспектралния скенер, за да се извлече необходимата информация – в нашия случай ръкописния текст. Например за ръкописи, които съдържат само един почерк, MSI сканирането може да се сведе до едно изображение, което подчертава древния текст. За палимпсести, от друга страна, третото измерение на MS скенер може да се сведе до две изображения, които подчертават двата различни слоя на текста. За да се маркират тези слоеве, може да се използва фалшиво оцветяване. Това може да се направи индивидуално от филолога с помощта на нашия „Nomacs“-Viewer или с добре познатия софтуер Photoshop.

Популярна техника за редуция на измеренията е Principal Component Analysis (PCA). Ние го използвахме за първоначалното подобряване на изображенията на Синайския требник. Вторият опит беше направен с Multivariate Spatial Correlation (MSC), която се оказва твърде времеемка. Друга техника за редуциране на измеренията се нарича Linear Discriminant Analysis (LDA). Основната разлика между PCA и LDA е, че PCA е автоматичен подход, а LDA не е, тоест LDA изисква предварителна информация за подгрупа мултиспектрални образци, докато PCA не изисква такава, тъй като необходимата информация се извлича автоматично.

Преди две години Фабиан Холаус и колегите му предложиха нова техника за подобряване на изображенията, базирана на LDA. Той отбелязваше образците като принадлежащи или на текста, или на фона. След това предложи алгоритъм за автоматично отбелязване, който използва основно алгоритъм, който открива редовете. Това се дължи на обстоятелството, че в повредените текстове редовете обикновено се виждат по-ясно от буквите.

Както виждате, съществуват много начини да се подобри качеството на изображенията на ръкописите, а в бъдеще те ще станат още повече. За филолога това има двоен ефект – от една страна, радост, защото той или тя може да прочете повече текст, а от друга – гняв и страх, защото всеки път, когато изображенията се подобряват, има нужда от нов прочит, както в случая със Синайския требник, където вече достигнахме до трети прочит!

*Превод: М. Панева*