GEOMETRIC AND INFORMATION CONSTRAINTS FOR AUTOMATIC LANDMARK SELECTION IN COLPOSCOPY SEQUENCES

Juan D. García-Arteaga, Jan Kybic
Center for Machine Perception, Czech Technical University, Karlovo Náměstí 13, Prague, Czech Republic
(garcia, kybic)@fel.cvut.cz

Jia Gu , Wenjing Li
STI Medical Systems, 733 Bishop Street, Suite 3100, Hawaii, USA
(jgu, wli)@sti-hawaii.com

Keywords: Medical Image Analysis, Registration,Computer-Aided Diagnosis, Colposcopy, Validation, Mutual Information,

Abstract: Colposcopy is a diagnostic method to visually detect cancerous and pre-cancerous tissue regions in the uterine cervix. A typical result is a sequence of cervical images captured at different times after the application of a contrast agent that must be spatially registered to compensate for patient, camera and tissue movement and on which progressive color and texture changes may be seen. We present a method to automatically select correct landmarks for non-consecutive sequence frames captured at long time intervals from a group of candidate matches. Candidate matches are extracted by detecting and matching feature points in consecutive images. Selection is based on geometrical constraints and a local rigid registration using Mutual Information. The results show that these landmarks may be subsequently used to either guide or evaluate the registration of these type of images.

1 INTRODUCTION

Colposcopy is a well established diagnostic method for early cervical cancer detection (Ferris, 2002). It is performed during a gynecological examination usually, but not always, following an abnormal pap smear test (Wright et al., 2002). During the exam a low concentration acetic acid (3 to 5%) is applied to the cervix to induce temporal color and texture changes in cancerous and pre cancerous tissue areas. An expert (colposcopist) observes the cervix through a low magnification microscope (colposcope) during the appearance change process (three to five minutes). He then reports his findings and, if necessary, recommends a biopsy to confirm the diagnosis. Envisioned CAD diagnostic systems aim to aid the physician in his diagnosis by quantitatively measuring and combining features with high prognostic values by analyzing high resolution images captured during the colposcopy exam (Lange and Ferris, 2005). This will result in more reliable diagnosis, minimization of the variability among colposcopists and the proportion of unnecessary false-positive biopsies.

Due to camera, patient and tissue movement, images taken during the exam must be spatially registered before the features extracted from each of them can be fused with the rest of the set. Several retrospective image-registration schemes have been proposed to rectify colposcopy image sequences, either considering each image individually (Lange et al., 2004) or using several frames from the sequence (Barreto-Flores et al., 2005). However there is no "gold standard" against which to compare the registration results other than using landmarks chosen by a human operator. Landmark selection, even when done by an expert, is subjective and difficult to reproduce.

In a recent article we showed how consistent landmarks can be semiautomatically extracted from a colposcopic image sequence by using a general feature detector and putative matches between consecutive images (García-Arteaga and Kybic, 2007). At the end of the process an expert chooses among the detected feature points the final landmarks.This approach proved robust in consistently detecting and matching correct features. The landmarks can then be used for image registration validation or as an additional cue for elastic registration. It relies however on a subjective judgement to validate the match.

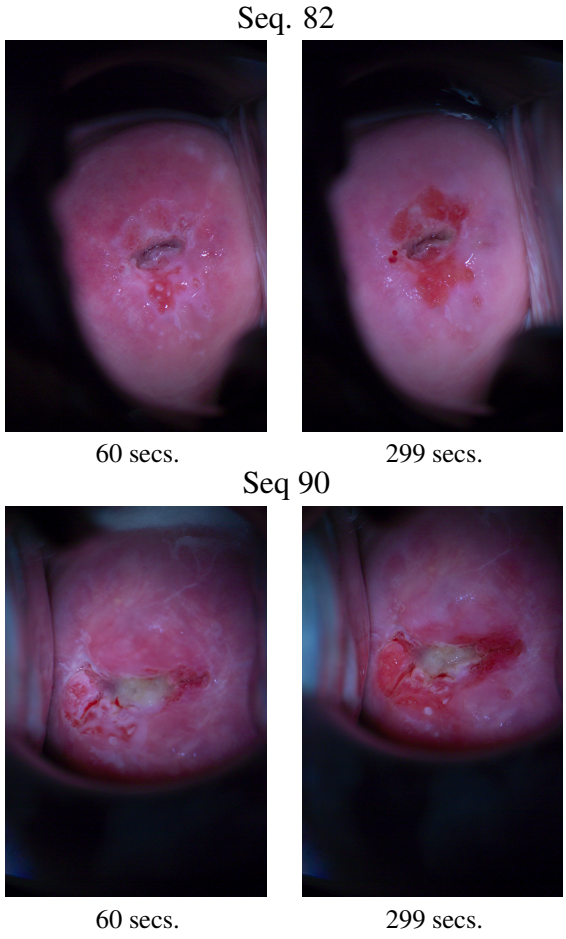

Figure 1: The first and last frames of the colposcopy sequences used for the experiments were captured 60 seconds (left column) and 299 seconds (right column) after the application of acetic acid. They correspond to the maximum color reaction and to the return to normal color, respectively.

In the present work we expand the previous methodology and automate the landmark selection process by adding additional information and geometrical constraints. We assume that two detected and matched feature points are corresponding landmarks if an image similarity metric in the neighborhood of the matched points is maximized when the landmarks overlap. In other words, we assume that registering the images in the neighborhood of the landmarks will make the landmarks overlap only if they are correctly matched. This assumption allows us to test and fully automate the process of landmark selection. Additionally, the similarity criterion gives us a quantitative measure of the quality of our match.

2 METHODS

The landmark selection process can be divided in three main steps. First pairwise candidate matches are calculated between every two consecutive frames of a colposcopy image sequence as detailed in (García-Arteaga and Kybic, 2007). A global rigid transformation that roughly aligns the images is calculated. Then, for each of the point pairs, an initial translation that makes the candidate points overlap is calculated. Using this translation as starting point, the small neighborhood of each of the candidate points is registered individually using different similarity criteria.

2.1 Candidate Point Extraction

On every image, feature locations are extracted using the Harris corner detector (Harris and Stephens, 1988), a well studied method that detects sharp changes in the first derivative of an image. On sharp polyhedral constructions (such as buildings) these features will correspond to corners. For other imaged objects the Harris detector will be activated by other features such as tight curves.

Because color and textural changes are gradual, good putative matches can be made between the points of interest found by the Harris detector in images captured at short time intervals. These matches are made between all consecutive frames.

To select the candidate landmarks between images m and n of the sequence only the points of interest that are matched in all frames between m and n are taken into account. This means that for the candidate feature point j in the initial image m ($\mathbf{X}_m^j$) and the candidate feature point j in the last image n ($\mathbf{X}_n^j$) to be accepted as landmarks there must exist an equivalent candidate feature point ($\mathbf{X}_k^j$) in image k for every k between m and n. In other words, $(\mathbf{X}_m^j, \mathbf{X}_n^j)$ is a candidate pair if and only if:

$$\forall k, k \in \mathbb{N}, m \leq k < n \quad \exists \quad \mathbf{X}_k^j, \mathbf{X}_{k+1}^j | \mathbf{X}_k^j \leftrightarrow \mathbf{X}_{k+1}^j$$

Interest points based on weak features will not survive for many frames. If all the matches throughout the sequence are correct then the locations of each of the candidate pair points corresponds to the same tissue region.

2.2 Similarity Measures

We have used two common image similarity metrics: Sum of Square Differences (SSD)(Zitova and Flusser, 2003) and Mutual Information (MI) (Viola and Wells, 1997).

The SSD criterion is one of the simplest and most commonly used similarity measures, having a low computational cost and smooth gradients. It assumes equal intensity values for corresponding pixels in the images. The SSD criterion of two images $\mathbf{g}$ and $\mathbf{f}$ within an interest region Ω is:

$$J_{SSD}(\mathbf{f}(i,j),\mathbf{g}(i,j)) = \sum_{(i,j)\subset\Omega} (\mathbf{f}(i,j)-\mathbf{g}(i,j))^2 \quad (1)$$

The MI similarity criterion measures the statistic dependency between two data sets. Because it does not assume any *a priori* known dependency between the images it has been extensively used for multi-modality image registration (Maes et al., 2003) (Pluim et al., 2003). The MI criterion is usually written in terms of entropy H:

$$J_{MI}(\mathbf{f},\mathbf{g}) = I(F,G) = H(F)+H(G)-H(F,G) \quad (2)$$

$$H(F) = -\int p_F(\xi)\log p_F(\xi)\mathrm{d}\xi \quad (3)$$

where F and G are random variables and pixel values $f_i = \mathbf{f}(\mathbf{x}_i)$, $g_i = \mathbf{g}(\mathbf{x}_i)$ are assumed to be realizations of F, G:$f_i \sim F$, $g_i \sim G$, $(f_i,g_i) \sim (F,G)$. The unknown probability distribution functions of F, G and (F,G) may be estimated by a normalized histogram of the intensity values.

For colposcopy images different tissue types subject to the contrast agent will have different reactions (e.g. normal epithelia will not change whereas metaplastic epithelia will turn opaque white). This suggests that the use of a multi-modality registration framework is necessary.

2.3 Deformations

The cervix is an elastic tissue and, as such, it will undergo elastic deformations during the colposcopy exam. However, we assume that the local deformation in the neighborhood of a point of interest (e.g. a landmark) can be approximated by a rigid transformations consisting of a translation and a rotation. Following the notation of (Ibáñez et al., 2005) we define this transformation as:

$$\mathbf{g}^*_{(\theta,\mathbf{T},\mathbf{C})}(x',y') = \mathbf{g}(x,y)$$

$$x' = (\cos\theta-\sin\theta)(x-C_x)+(T_x+Cx)$$

$$y' = (\sin\theta+\cos\theta)(y-C_y)+(T_y+Cy) \quad (4)$$

where θ is the rotation angle, (C_x,C_y) are the coordinates of the rotation center and (T_x,T_y) are the components of the translation.

2.4 Optimization

The registration is defined as the minimization of the similarity criterion J:

$$\arg\min_{\theta,T_x,T_y} J(\mathbf{f}(i,j),\mathbf{g}^*_{(\theta,\mathbf{T},\mathbf{C})}(i,j)) \quad (5)$$

where J will be either J_{SSD} or J_{MI}. We search for the θ, T_x and T_y that solve this equation using a gradient descent algorithm for a window around every feature point pair $\mathbf{X}^j$. The center of rotation corresponds to the test image feature coordinates (for simplicity we will drop the superscript to refer to each candidate pair in the rest of the article):

$$\mathbf{C} = \mathbf{X}_{test}$$

The initial translation $\mathbf{T}_0$ is chosen so the feature point in the test image $\mathbf{X}_{test}$ is mapped to the location of the feature point in the reference image $\mathbf{X}_{ref}$, making them overlap:

$$\mathbf{T}_0 = \mathbf{X}_{ref} - \mathbf{X}_{test}$$

All operations are carried over a square 65x65 Region of Interest (ROI) centered on the reference image feature point location, in this case:

$$\forall i,j\in\mathbb{N} \quad | \quad \begin{array}{l} X_{refx}-32 \le i \le X_{refx}+32 \\ X_{refy}-32 \le j \le X_{refy}+32 \end{array}$$

This window size keeps the local deformation assumption valid while still providing enough statistical information to evaluate the MI criterion.

As a preprocessing step, all images are roughly aligned by a sequential rigid registration of the frames (García-Arteaga and Kybic, 2007) making initial angle $\theta_0 = 0$ for all feature point pairs of the image.

We consider the optimization process as an iterative refinement of our initial estimation of the test image feature point location. The new test image feature point location is chosen so that when transformed it will overlap the reference image feature point. Setting $(x',y') = \mathbf{X}_{ref}$ in Equation 4 we find the new test image feature location (x,y):

$$x = (x'-T_x-C_x)(\cos(\theta)-\sin(\theta))^{-1}+C_x$$

$$y = (y'-T_y-C_y)(\sin(\theta)+\cos(\theta))^{-1}+C_y$$

$$\mathbf{X}_{testNew} = (x,y)$$

The ITK implementations of the Mattes MI criterion and of the gradient descent optimization were used (Ibáñez et al., 2005).

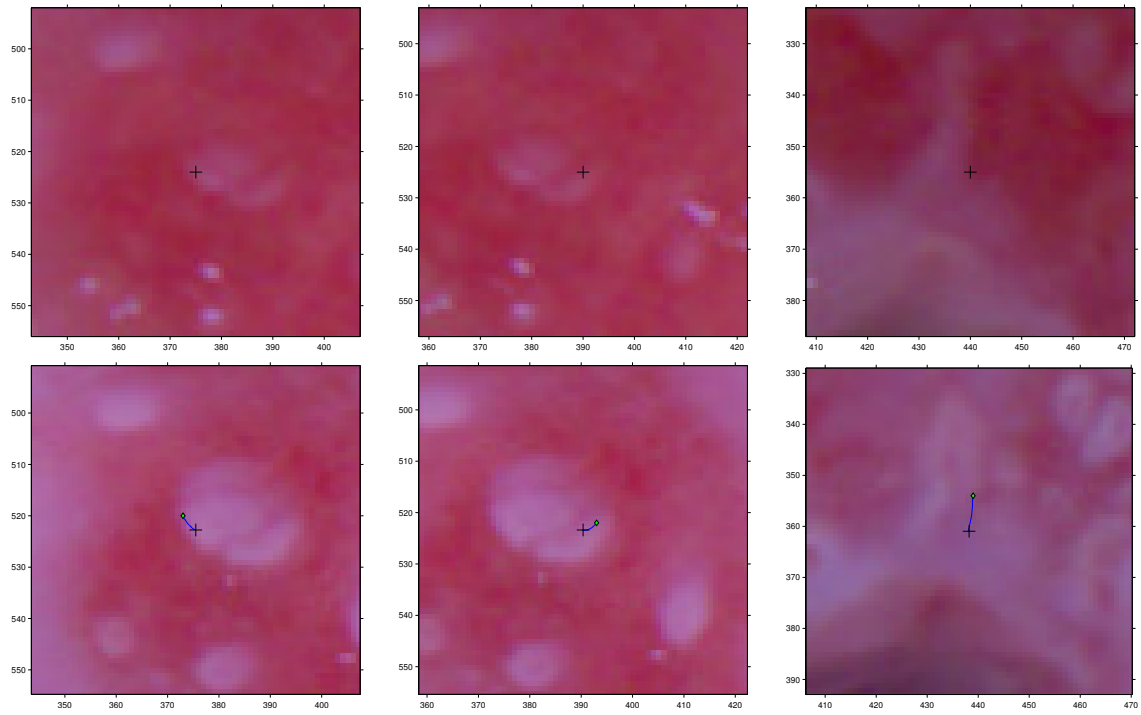

Figure 2: Final candidate pair location after MI registration is shown with a cross in the the template (top) and the test image (bottom) for initially misaligned pairs. The line from the final to the initial location of the test image feature point (marked with a diamond) illustrates the path followed by the optimization. The images shown correspond to the ROIs used for the registration.

3 EXPERIMENTS

We have tested our method on the initial and final frames of two sequences of images taken from different patients between 60 and 299 seconds after application of the acetic solution (sequences 82 and 90, seen in Figure 1). These images correspond to the moment of maximal acetic whitening and to its return to normal color. The sequences are composed of nineteen 750x1125 pixel color images taken at intervals of between 5 and 12 seconds with a cross-polarized filter to reduce glint and a black opaque speculum that partially masks the background. All operations are done with grayscale versions of the images although final results are presented in color. The landmark extraction algorithm produced 20 and 23 tentative matches, respectively.

3.1 Results MI

After registration the number of correct matches (evaluated manually) goes from 6 to 15 out of 23 candidate pairs for sequence 82 and from 10 to 20 out of 23 for sequence 90. It should be noted that most of the mismatches in sequence 82 occur in smooth featureless zones where, even for a human, it is difficult to distinguish salient features. In this areas glint is mistakenly detected as a valid candidate point.

The results of using MI for registration are clearly visible in strongly misaligned windows as those seen in Figure 2. The misalignment is usually due to one or more mismatches during the candidate extraction process, i.e. there is at least one k such that:

$$\mathbf{X}_k \nleftrightarrow \mathbf{X}_{k+1}$$

When this happens the whole landmark selection process is corrupted. By re-placing the test image feature point location as described in Section 2.4, we get good final matches from an originally mismatched pair. This increases the number and quality of the final landmarks.

3.2 Results SSD

It was observed that in many cases the registration using the SSD criterion converged to minima far away from the correct transformations, further misaligning the landmarks as seen in Figure 3. It is difficult to asses if this behavior is inherent to the appearance changes induced by the contrast agent or to the severe

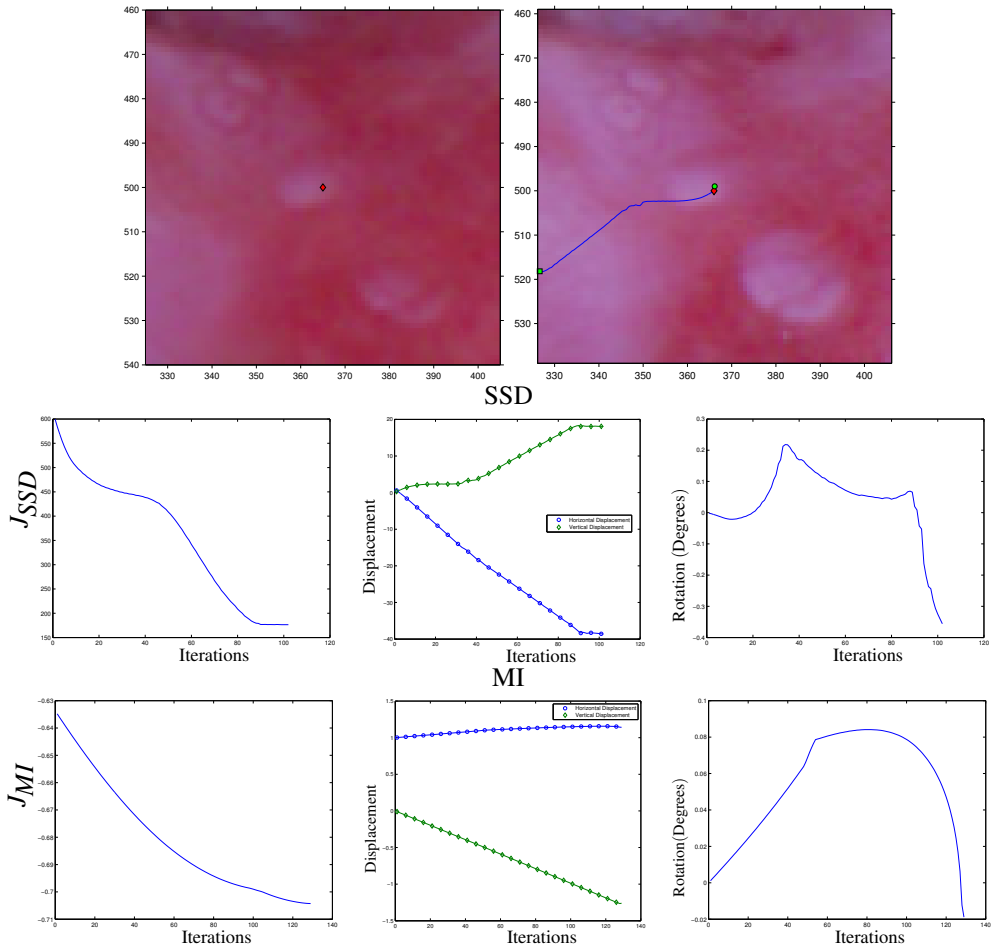

Figure 3: A correctly matched candidate point pair is shown in the reference (top left) and test image (top right) marked as diamonds at the center of each image. Both the SSD (second row) and the MI (third row) criteria are minimized (leftmost column). The final transformation (composed of the translations shown in the center column and the rotations on the right column) re-locates the test image candidate point near its initial position when using MI (shown as a circle overlapping the initial feature, top right) and drifts away from it when using SSD (final location shown as a square, top right).

illumination inhomogeneities known to be present in cervical imagery (Dvir et al., 2006). This, together with the low number of correct matches after registration (2 out of 20 for sequence 82 and 5 out of 23 for sequence 90), leads us to believe that the SSD criterion is not appropriate for cervical images at the level of detail required for this application.

3.3 Geometric Validation

To select the final set of landmarks we check that their position is consistent with the other landmarks, that is, we add a geometric constrain. We used the RANSAC algorithm (Fischler and Bolles, 1987) to robustly fit the candidate points to a projective transformation both before and after registration using MI. Projective transformations have been used previously for colposcopic sequence alignment (Barreto-Flores et al., 2005) and their robust estimation from observed data is a well studied problem (Hartley and Zisserman, 2004). The results of using the original (unregistered) and the registered set of correspondences may be seen in Table 1. The results shown are the average of running RANSAC 120 times with each of the sets.

Table 1: Results of RANSAC fitting. N is the number of candidate points and Iterations the average number of iterations over all runs.

	Iterations	Inliers	N
82	402	0	23
82_{MI}	80	6	23
90	267	0	20
90_{MI}	46	5	20

Points selected in 90% or more of the runs are considered inliers. Visual inspection of the selected candidate points confirms that all but one of the inliers selected corresponded to correct matches. The remaining one is located in a smooth area where it is not possible to determine if it is a correct match or not. The very restrictive distance threshold between the data points and the model used for outlier rejection (equal or less than a pixel) explains the relatively low number of inliers when compared to manual evaluation.

It should be noted that it takes a much larger number of iterations to arrive to the final answer in the original set of points than when fitting the MI registered set. This suggests a higher proportion of outliers in the original set than in the registered set.

4 CONCLUSIONS

We have presented an improvement over our previous landmark extraction method for registration validation. The previous method made no direct comparison between the registered images to guarantee the similarity of the candidate points. We have shown that this can be done by using a local estimate of the MI criterion as a similarity measure and registering locally with rigid transformations.

We have also observed that although low complexity similarity criteria like the SSD are good for matching consecutive frames with small appearance changes, they are not appropriate to compare images captured at long time intervals. The failure of the SSD similarity criterion to align even correctly matched candidate points confirms that multi modality registration framework is better suited for this type of images.

An additional geometrical constraint was added to select the final landmarks. Such a constraint is necessary because the MI similarity criterion, as implemented, is not sufficiently discriminative to separate landmarks from mismatches. Further research in the use of similarity criteria that takes into account important visual information such as color and edges together with a better geometric model of the cervix deformation should improve the robustness of the system and reduce the number of false negatives (landmarks classified as mismatches).

REFERENCES

Barreto-Flores, A., Altamirano-Robles, L., Morales-Tepalt, R. M., and Cisneros-Aragon, J. D. (2005). Identifying precursory cancer lesions using temporal texture analysis. In *Second Canadian Conference on Computer and Robot Vision*, pages 34–39. IEEE Computer Society.

Dvir, H., Gordon, S., and Greenspan, H. (2006). Illumination correction for content analysis in uterine cervix images. In *CVPRW '06: Proceedings of the 2006 Conference on Computer Vision and Pattern Recognition Workshop*, page 95, Washington, DC, USA. IEEE Computer Society.

Ferris, D. (2002). *Modern Colposcopy: Textbook and Atlas*. Kendall Hunt Publishing Company, Dubuque, Iowa, 2nd edition.

Fischler, M. A. and Bolles, R. C. (1987). *Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.

García-Arteaga, J. D. and Kybic, J. (2007). Automatic Landmark Detection for Cervical Image Registration Validation. *SPIE Medical Imaging 2007, in press*.

Harris, C. and Stephens, M. (1988). A combined corner and edge detection. In *Proceedings of The Fourth Alvey Vision Conference*, pages 147–151.

Hartley, R. I. and Zisserman, A. (2004). *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN: 0521540518, second edition.

Ibáñez, L., Schroeder, W., Ng, L., and Cates, J. (2005). *The ITK Software Guide*. Kitware, Inc., second edition.

Lange, H., Baker, R., Håkansson, J., and Gustafsson, U. P. (2004). Reflectance and fluorescence hyperspectral elastic image registration. In Fitzpatrick, J. M. and Sonka, M., editors, *Medical Imaging 2004: Physiology, Function, and Structure from Medical Images. Edited by Amini, Amir A.; Manduca, Armando. Proceedings of the SPIE, Volume 5370, pp. 335-345 (2004).*, pages 335–345.

Lange, H. and Ferris, D. G. (2005). Computer-aided-diagnosis (CAD) for colposcopy. *Medical Imaging 2005: Image Processing*, 5747(1):71–84.

Maes, F., Vandermeulen, D., and Suetens, P. (2003). Medical image registration using mutual information. *Proceedings of the IEEE*, 91(10):1699–1722.

Pluim, J. P. W., Maintz, J. B. A., and Viergever, M. A. (2003). Mutual-information-based registration of medical images: a survey. *Medical Imaging, IEEE Transactions on*, 22(8):986–1004.

Viola, P. and Wells, W. M. I. (1997). Alignment by maximization of mutual information. *Int. J. Comput. Vision*, 24(2):137–154.

Wright, T. C. J., Cox, T. J., Massad, L. S., Twiggs, L. B., Wilkinson, E. J., and for the 2001 ASCCP-Sponsored Consensus Conference (2002). 2001 Consensus Guidelines for the Management of Women With Cervical Cytological Abnormalities. *JAMA*, 287(16):2120–2129.

Zitova, B. and Flusser, J. (2003). Image registration methods: a survey. *Image and Vision Computing*, 21(11):977–1000.