Class-based feature matching across unrestricted transformations

Evgeniy Bart and Shimon Ullman

The authors are with the Department of Computer Science and Applied Mathematics, Weizmann Institute of Science, Rehovot 76100, Israel
E-mail: bart@caltech.edu, shimon.ullman@weizmann.ac.il

APPENDIX I
DERIVATION OF THE DECOMPOSITION EQUATIONS

We derive below the forms of eqs. (4), (6).

By definition of conditional probability, $P(X,Y) = P(X)P(Y|X)$. A reasonable simplifying assumption is that the fragments depend only on their parents in the pyramid, and are independent given the parent's correct match. In this case, $P(X)$ can be factored as

$$P(X) = P(X_R) \prod_{f \in \mathcal{F} \setminus \{R\}} P(X_f | X_{\pi(f)}), \qquad (10)$$

where $\pi(f)$ is the parent of f in the pyramid and R is the root fragment. The product is over all fragments except the root.

Next, we assume that the observations $C(f,g')$ involving some source fragment f are determined by the value of X_f (its true match) and g'. Therefore, given X_f, the observation $C(f,g')$ is independent of all other observations of the same fragment, $C(f,t')$ $(g' \neq t')$, of matches of other fragments X_h $(h \neq f)$ and of their observations $C(h,t')$ $(h \neq f)$. This assumption is standard for similar models [24].

Note that high consistency $C(f,g')$ indicates that f is likely to match g'. This increases the likelihood that $X_f = g'$ and therefore decreases the likelihood that $X_f = h' \neq g'$. This, in turn, suggests that the consistency value $C(f,h')$ is unlikely to be high. The observations $C(f,g')$ and $C(f,h')$ are therefore dependent. The reason for this dependence is that both observations affect the distribution of X_f. The assumption made above is that X_f is the only source of the dependence and therefore the observations become independent when the value of X_f is fixed.

Under this assumption, $P(Y|X)$ can be rewritten as

$$P(Y|X) = \prod_{f \in \mathcal{F}} P(Y_f | X_f), \qquad (11)$$

where Y_f is the set of all observed consistency values that involve the fragment f:

$$Y_f = \{C(f,g') | g' \text{ at the same level as } f\}. \qquad (12)$$

Due to conditional independence,

$$P(Y_f|X_f) = \prod_{g' \in M'(f)} P(C(f,g')|X_f), \qquad (13)$$

where $M'(f)$ is the set of the possible matches of f, i.e. $M'(f)$ is the set of candidate target fragments g' at the same level as f.

Combining the assumptions above results in the desired factorization

$$P(X,Y) = P(X_R) \prod_{f \in \mathcal{F} \setminus \{R\}} P(X_f|X_{\pi(f)}) \prod_{f \in \mathcal{F}} P(Y_f|X_f). \qquad (14)$$

APPENDIX II
ESTIMATING THE MODEL'S PROBABILITIES

Here, we justify in more detail eqs. (7), (8), (9).

By Bayes formula,

$$P(C(f,f')|X_f = f') = P(X_f = f'|C(f,f')) \frac{P(C(f,f'))}{P(X_f = f')}. \qquad (15)$$

Since high consistency is an evidence of a match, a natural assumption is that $P(X_f = f'|C(f,f'))$ is proportional to $C(f,f')$, or, more generally, to $[C(f,f')]^\gamma$ for some power γ. Values of $\gamma > 1$ make the maxima of $C(f,f')$ sharper, while values of $\gamma < 1$ make the maxima of $C(f,f')$ wider. Assuming also that

$$\frac{P(C(f,f'))}{P(X_f = f')} = const, \qquad (16)$$

we arrive at the following estimate:

$$P(C(f,f')|X_f = f') = \frac{1}{Z_1}[C(f,f')]^\gamma, \qquad (17)$$

where Z_1 is an appropriate normalization factor. Note that the actual distribution $P(C(f,f'))$ could be estimated during training since this does not require examples of correct matches. However, good performance was obtained simply by using eq. (17).

Experiments with two different approximations of $P(C(f,g')|X_f = f')$ were performed. The first approximation was to use

$$P(C(f,g')|X_f = f') = 1 - P(C(f,g')|X_f = g') \qquad (18)$$

(where $P(C(f,g')|X_f = g')$ is estimated from eq. (17)). This assumes that high consistency values are unlikely for a non-matching pair, while lower consistency values are more likely. The second approximation assumed a uniform distribution for $P(C(f,g')|X_f = f')$. The intuition here is that two non-matching fragments f, g' are independent; therefore, all consistency values are equally likely. In particular, high consistency for a non-matching pair can be obtained by chance. In the experiments, the performance of the two approximations was equivalent. However, the second approximation allows a more efficient implementation: since the terms $P(C(f,g')|X_f = f')$ are constant, $P(Y_f|X_f = f')$ can be written simply as

$$P(Y_f|X_f = f') = \frac{1}{Z_1}[C(f,f')]^\gamma, \qquad (19)$$

where Z_1 is adjusted so that $P(Y_f|X_f = f')$ is a valid probability distribution. This approximation was used in the final scheme.

The distribution $P(X_f = f'|X_{\pi(f)} = f'_\pi)$ implements the proximity constraint, using the hierarchical structure. Assume that $\pi(f)$ matches f'_π. By definition, f is a child of $\pi(f)$. Similarity of the hierarchical structure requires f' to be a child of f'_π. Therefore, $P(X_f = f'|X_{\pi(f)} = f'_\pi)$ should be high if the smaller fragment f' falls inside the larger fragment f'_π, and should decrease when f' becomes more distant from f'_π. Note that the probability $P(X_f = f'|X_{\pi(f)} = f'_\pi)$ should be defined for all fragments f'_π, even if f'_π is actually not a parent of f'. Setting $P(X_f = f'|X_{\pi(f)} = f'_\pi)$ to a small but non-zero value if $\pi(f') \neq f'_\pi$ allows the scheme to cope with occasional deviations from the proximity constraint.

The actual distribution $P(X_f = f'|X_{\pi(f)} = f'_\pi)$ could be learned from examples. However, since no observable examples of correct matches are available in our setting, the following estimate conforming to the qualitative requirements listed above was used:

$$P(X_f = f'|X_{\pi(f)} = f'_\pi) = \frac{1}{Z_2}\frac{1}{d(f', f'_\pi)^\beta}. \qquad (20)$$

Here $d(f', f'_\pi)$ is the distance from the center of f' to the center of f'_π, and Z_2 is the appropriate normalization factor, calculated so that the right-hand side in eq. (20) sums to 1. In our experiments, this estimate was sufficient to obtain good performance. Experiments with different parameters α, β in the family

$$P(X_f = f'|X_{\pi(f)} = f'_\pi) = \frac{1}{Z_2}\exp(-\alpha d^\beta(f', f'_\pi))$$

were also done, but resulted in poorer performance.

REFERENCES

[1] C. Tomasi and T. Kanade, "Detection and tracking of point features," Carnegie Mellon University, Tech. Rep. CMU-CS-91-132, Apr. 1991.

[2] T. Tuytelaars and L. V. Gool, "Wide baseline stereo matching based on local, affinely invariant regions," in *British Machine Vision Conference*, 2000, pp. 412–425.

[3] K. Mikolajczyk and C. Schmid, "An affine invariant interest point detector," in *Proceedings of the European Conference on Computer Vision*, 2002, pp. 128–142.

[4] M. Brown and D. Lowe, "Invariant features from interest point groups," in *British Machine Vision Conference*, 2002.

[5] R. Basri and D. Jacobs, "Recognition using region correspondences," *International Journal of Computer Vision*, vol. 25, no. 2, pp. 141–162, 1997.

[6] S. Agarwal and D. Roth, "Learning a sparse representation for object detection," in *Proceedings of the European Conference on Computer Vision*, 2002, pp. 113–127.

[7] S. Ullman, M. Vidal-Naquet, and E. Sali, "Visual features of intermediate complexity and their use in classification," *Nature Neuroscience*, vol. 5, no. 7, pp. 682–687, 2002.

[8] D. I. Perrett, P. A. J. Smith, D. D. Potter, A. J. Mistlin, A. S. Head, A. D. Milner, and M. A. Jeeves, "Visual cells in the temporal cortex sensitive to face view and gaze direction," *Proceedings of the Royal Society of London. Series B: Biological Sciences*, vol. 223, pp. 293–317, 1985.

[9] M. J. Black and P. Anandan, "A framework for the robust estimation of optical flow," in *Proceedings of the International Conference on Computer Vision*, 1993, pp. 231–236.

[10] Y. Adini, Y. Moses, and S. Ullman, "Face recognition: the problem of compensating for changes in illumination direction," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, pp. 721–732, 1997.

[11] C. Wallraven and H. H. Bülthoff, "Automatic acquisition of exemplar-based representations for recognition from image sequences," in *CVPR 2001 - Workshop on Models vs. Exemplars*, 2001.

[12] S.-H. Lai, "Robust image matching under partial occlusion and spatially varying illumination change," *Computer Vision and Image Understanding*, vol. 78, pp. 84–98, 2000.

[13] G. D. Hager and P. N. Belhumeur, "Efficient region tracking with parametric models of geometry and illumination," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 10, pp. 1025–1039, 1998.

[14] M. J. Black and A. D. Jepson, "EigenTracking: Robust matching and tracking of articulated objects using a view-based representation," in *Proceedings of the European Conference on Computer Vision*, 1996, pp. 329–342.

[15] D. Tell and S. Carlsson, "Combining appearance and topology for wide baseline matching," in *Proceedings of the European Conference on Computer Vision*, 2002, pp. 68–81.

[16] V. Ferrari, T. Tuytelaars, and L. V. Gool, "Simultaneous object recognition and segmentation by image exploration," in *Proceedings of the European Conference on Computer Vision*, 2004.

[17] D. G. Lowe, "Three-dimensional object recognition from single two-dimensional images," *Artificial Intelligence*, vol. 31, no. 3, pp. 355–395, 1987.

[18] P. Viola and W. Wells, "Alignment by maximization of mutual information," *International Journal of Computer Vision*, vol. 24, no. 2, pp. 137–154, 1997.

[19] E. Bart, E. Byvatov, and S. Ullman, "View-invariant recognition using corresponding object fragments," in *Proceedings of the European Conference on Computer Vision*, 2004, pp. 152–165.

[20] A. Chowdhury, R. Chellappa, and T. Keaton, "Wide baseline image registration with application to 3D face modeling," accepted to IEEE Trans. Multimedia.

[21] E. Sali and S. Ullman, "Combining class-specific fragments for object recognition," in *British Machine Vision Conference*, 1999, pp. 203–213.

[22] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[23] H. Abdi, "Kendall rank correlation," in *Encyclopedia of Measurement and Statistics*, N. J. Salkind, Ed. Thousand Oaks (CA): Sage, 2007.

[24] J.-M. Laferte, P. Perez, and F. Heitz, "Discrete markov image modeling and inference on the quadtree," *IEEE Transactions on Image Processing*, vol. 9, no. 3, pp. 390–404, 2000.

[25] F. R. Kschischang, B. J. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Transactions on Information Theory*, vol. 47, no. 2, pp. 498–519, 2001.

[26] http://vision.stanford.edu/~birch/klt/.

[27] http://www.cs.brown.edu/people/black/ignc.html.

[28] http://www.inrialpes.fr/lear/people/Mikolajczyk/.

[29] T. Sim, S. Baker, and M. Bsat, "The CMU pose, illumination, and expression (PIE) database of human faces," The Robotics Institute, Carnegie Mellon University, Tech. Rep. CMU-RI-TR-01-02, January 2001.

[30] P. J. Phillips, H. Wechsler, J. Huang, and P. Rauss, "The FERET database and evaluation procedure for face recognition algorithms," *Image and Vision Computing*, vol. 16, no. 5, pp. 295–306, 1998.

[31] O. Chum and J. Matas, "Randomized RANSAC with $t_{d,d}$ test," in *British Machine Vision Conference*, 2002.

[32] "Oxford colleges," http://www.robots.ox.ac.uk/~vgg/data2.html.

[33] "Weizmann institute toy car database," http://www.wisdom.weizmann.ac.il/~cars.

[34] S. Ullman and E. Bart, "Recognition invariance obtained by extended and invariant features," *Neural Networks*, vol. 17, pp. 833–848, 2004.

[35] E. Bart and S. Ullman, "Class-based matching of object parts," in *Proceedings of CVPR Workshop on Image and Video Registration*, 2004.