ILLUMINATION-ROBUST PATTERN MATCHING USING DISTORTED HISTOGRAMS

Georg Thimm   Juergen Luettin

Illumination-robust Pattern Matching using Distorted Histograms

Georg Thimm   Juergen Luettin

Abstract. It is argued that global illumination should be modeled separately from other incidents that change the appearance of objects. The effects of intensity variations of the global illumination are discussed and constraints deduced that restrict the shape of a function that maps the histogram of a template to the histogram of an image location. This approach is illustrated for simple pattern matching and for a combination with a PCA (*Eigenface*) model of the grey-level appearance.

Keywords: illumination variation, histogram, pattern matching, appearance, Eigenface model

1 Introduction

The appearance of objects (*i.e.* its image) depends on several variables, as for example the scene illumination. As changing one or more of these variables changes the appearance of an object, a recognition system has to be robust against them. In this context, the most important difficulties are caused by:

- the *scene illumination* (the position of light sources and their occlusion (shadows); specular reflections),

- *occlusions* by other objects,

- the *reflectance properties* of the object, respectively its transparence (X-ray images),

- the *viewing angle*, and

- the *shape* of the object if it is flexible.

Although all these points are worth to be investigated, we restrict us for obvious reasons to illumination related problems (see [3][8][10] for other topics). Highly sophisticated approaches use for example an approximate 3-dimensional representation of the scene and try to estimate shape, reflection coefficient, and illumination [7][10], a combined PCA model of shape and intensity on landmark points [5], respectively *active shapes* [6][9], a model for the object under multiple illumination situations [15](*Eigenfaces*) [2], a model of the illumination variation and specular reflections [3], or 3-dimensional models and neural networks to estimate the position of the light sources [4].

Global illumination changes and total shadowing, however, are not well modeled by these approaches. To the knowledge of the authors, global illumination changes were only considered in combination with other image analysis methods, for example in the context of *change detection* (see [14] for more references), or *optical flow* [12]. We assume that it is inefficient to model the illumination of an object as a whole and to neglect its different causes. A better approach will model the illumination that is global to the object separate from other appearance changes.
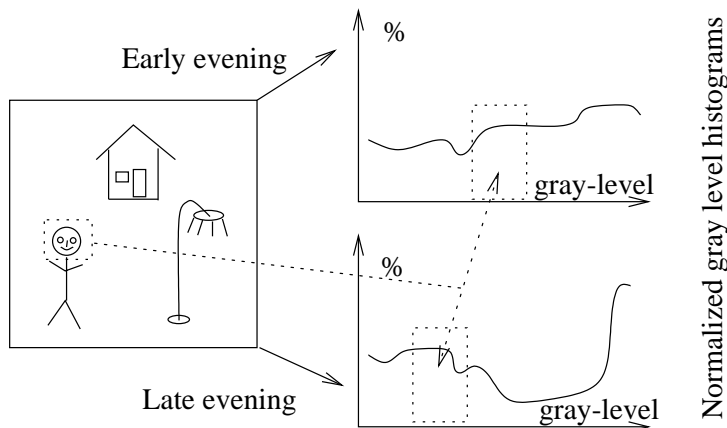

Figure 1: Illumination changes distort the grey-level histogram.

Figure 1 illustrates a possible situation considered in this publication. Suppose that faces have to be recognized in an outdoor scene (neglecting appearance changes other than global illumination changes). Depending on the daytime, the illumination of the scene changes and at the same time the relative brightness of objects. Consequently, a normalized grey-level histogram[1] of the scene is also subject to alternations. Along with these alternations, the average brightness of objects (such as the face in figure 1) will change. In other words, a certain object will contribute to other parts of the histogram. Approximately, the grey-level histogram of a single object is projected onto another, or *distorted*.

---

[1] A normalized grey-level histogram is scaled and translated to fit the whole range of grey-levels, in this publication the integer values form 0 to 255.

2 The approach

In order to compensate for the distortion of the grey-level histograms, the histogram of the template has to be modified prior to a comparison with some image location. The function projecting the histogram of the template to the histogram of the image can be regarded as a model of the illumination variation. Therefore the shape of this function is constrained according to three assumptions (which are not necessarily always true for real scenes):

1. As the image is normalized, the lowest and highest intensities in the grey-level histogram will be mapped onto themselves.

2. Contrasts diminish or augment smoothly when the global illumination changes. Therefore, modifications of grey-levels must vary smoothly within neighboring intensity values.

3. The relative brightness of arbitrary objects must remain unchanged: if a certain spot in the image is brighter than another spot, it will remain brighter or, in the limit, assume the same intensity.

A simple pattern algorithm using such a histogram mapping function matching can be formulated in the following way: let $p = (p_1, p_2, \ldots, p_N)$ be a feature vector of grey values, representing the template and $\hat{p}(x, y)$ a vector extracted from some image to be compared with p. Then the most likely position (x^*, y^*) for the object represented by the template can be defined as

$$(x^*, y^*) = \underset{x,y}{\operatorname{argmin}} \sum_i \|\hat{p}_i(x, y) - f_{\boldsymbol{\alpha}}(p_i)\|.$$

In this formula, the function $f_{\boldsymbol{\alpha}}$ models the distortion of the grey-level histogram. $f_{\boldsymbol{\alpha}}$ is parameterized by a vector $\boldsymbol{\alpha}$ corresponding to the deviation of the illumination of the image as compared to the illumination of the template. Since $\boldsymbol{\alpha}$ is usually unknown, it must be included into the error minimization:

$$(x^*, y^*) = \underset{x,y}{\operatorname{argmin}} \ \underset{\boldsymbol{\alpha}}{\min} \sum_i \|\hat{p}_i(x, y) - f_{\boldsymbol{\alpha}}(p_i)\|.$$

As discussed earlier, $f_{\boldsymbol{\alpha}}$ has to fulfill some conditions in order to avoid a too flexible mapping which would result in low scores for illicit image locations.

1. The invariability of the lowest and highest light intensity can be directly formulated as a condition on $f_{\boldsymbol{\alpha}}$. Supposed that p and $\hat{p}(x, y)$ are extracted from normalized images and black is coded as $c_{\min}$ and white as $c_{\max}$ (usually 0 and 255). Then $f_{\boldsymbol{\alpha}}$ has to fulfill:

$$f_{\boldsymbol{\alpha}}(c_{\min}) = c_{\min}, \quad \text{and} \quad f_{\boldsymbol{\alpha}}(c_{\max}) = c_{\max}$$

2. The similarity constrain on the variation of close grey-levels can be fulfilled by demanding that $f_{\boldsymbol{\alpha}}$ possesses a smooth first derivative.

3. That grey-levels are not interchangeable implies, that the mapping function is non-decreasing for the range of valid grey-levels. As $f_{\boldsymbol{\alpha}}$ possesses a first derivative:

$$f'_{\boldsymbol{\alpha}}(x) \geq 0 \quad \text{for } c_{\min} \leq x \leq c_{\max}.$$

Considering these constraints, $f_{\boldsymbol{\alpha}}$ was chosen to be a second order polynomial, although other functions which fulfill these conditions exist. It follows from the constraints above that $f_{\boldsymbol{\alpha}}$ has the form

$$f_{\boldsymbol{\alpha}}(p) = p + \alpha (p - c_{\min})(p - c_{\max}),$$

with a free variable α restricted to the interval

$$\alpha \in \left[\frac{1}{c_{\min} - c_{\max}}, \frac{1}{c_{\max} - c_{\min}} \right]. \tag{1}$$

Figure 2 shows f_α for $c_{\min} \equiv 0$, $c_{\max} \equiv 255$, and $\alpha \in \{\frac{-1}{255}, 0, \frac{1}{255}\}$.
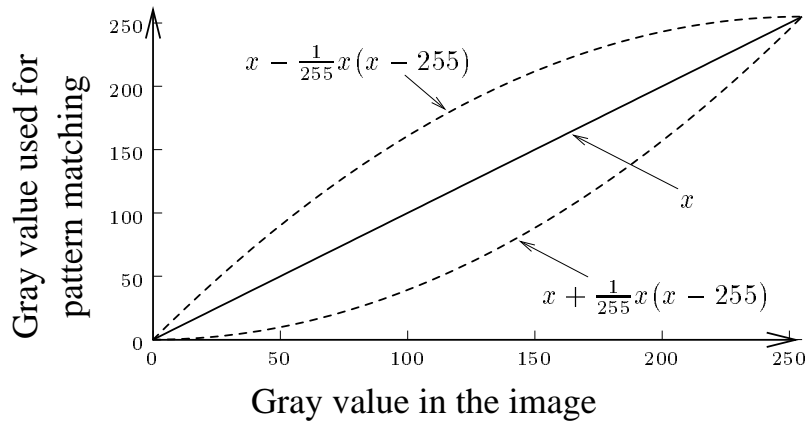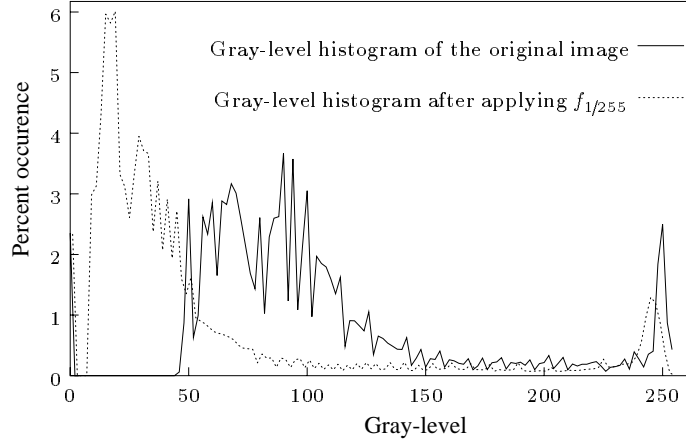
Figure 2: Possible instances of the mapping function f_α.

Figure 3: The histogram of an image before and after applying $f_{1/255}$.

This function has the property, that, depending on the sign of α, either the contrasts in the brighter parts, respectively the darker parts, of the image are augmented. At the same time, the contrasts in the darker parts, respectively the brighter parts, are lowered. Figure 3 shows the effect of $f_{1/255}$ on the histogram of an image shown in figure 4. It can be seen that gray-levels in the interval $[50, 120]$ are "moved" into the previously empty space of $[10, 40]$. At the same time, the steepness of the histogram, respectively the contrast, in the lower part is increased, whereas the contrast in the higher part is reduced.

The form of f_α has the advantage that an explicit solution (not respecting equation (1)) for α exists if $|| \cdot ||$ is the mean square norm:

$$\alpha = \frac{\sum_i \hat{p}_i (p_i - c_{\min})(p_i - c_{\max})}{\sum_i (p_i - c_{\min})^2 (p_i - c_{\max})^2} \tag{2}$$

In the way the distorted pattern matching approach is described above, it can not sensibly be applied when appearance changes are evoked by global illumination changes **and** other incidents. This deficiency may in principle be overcome by combining it with other techniques, for example Eigenface models [6]. Let be given an eigenvector matrix $\mathbf{P}$, a mean appearance $\mathbf{m}$, and an appearance vector $\mathbf{b}$ that describe the appearance of an object under constant global illumination. Then, for this method of object modelization, the most likely position (x^*, y^*) of some feature can be redefined as

$$(x^*, y^*) = \operatorname*{argmin}_{x,y} \ \min_{\boldsymbol{\alpha}, \mathbf{b}} \sum_i || \hat{p}_i(x, y) - f_{\boldsymbol{\alpha}}(\mathbf{P}_i \mathbf{b} + \mathbf{m}_i) ||, \tag{3}$$

The minimization in equation (3) for $\mathbf{b}$ can be done, for example, by using the simplex algorithm, whereas α can still be determined by applying equation (2) to the vector $(\mathbf{Pb} + \mathbf{m})$ with $\mathbf{b}$ proposed by the simplex algorithm.

3   Experimental Evaluation

The proposed method was tested on 4,000 X-ray images of the vocal tract of talking persons [11]. In these tests, fillings in the upper and lower teeth, as well as the tips of the front teeth of the sequence Laval 43 were tracked. These tracking tasks are difficult as the illumination of the images is not stable, and a lot of noise is present. Examples of this sequence are shown in figure 4. The results are compared in experiments with a standard pattern matching algorithm (that is equivalent to $\alpha \equiv 0$) and the Eigenface method using different numbers of eigenvectors.
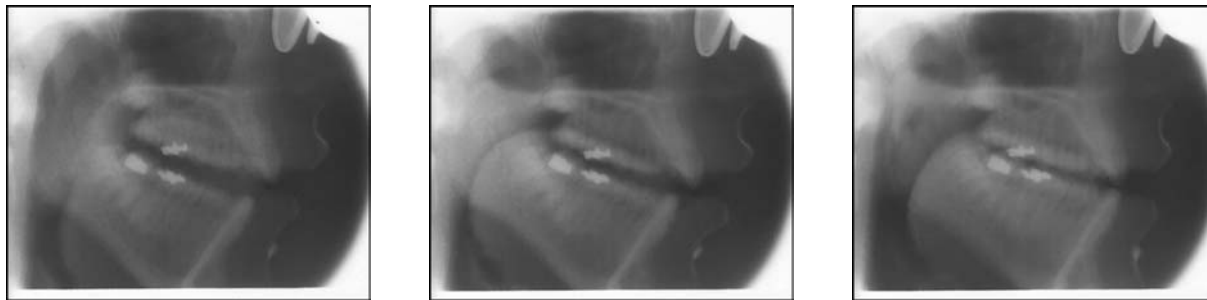

Figure 4: Examples from the sequence Laval 43 of the X-ray database.

It is very laborious or even infeasible to evaluate quantitatively the performance of pattern matching algorithms, as this would require a huge amount of data being labeled by hand. In order to circumvent this difficulty, but still comparing objectively pattern matching algorithms, the following was done: first, using all algorithms the same feature was located in a large number of images by searching the neighborhood of the previously detected feature position. Then, the locations are inspected visually by cutting out a small region around the found locations and visualizing them together. This permitted to correct situations where the tracking failed grossly[2]. For the hundred locations, where the distances between the detected locations are biggest, were then visually inspected and whenever one of the algorithms performed better, its "did it better" score is increased by one.

Table 1 shows the results of these experiments. Each entry in this table is a comparison of the distorted gray-level algorithm with the algorithm at the column top and the feature at the left of the row. The numbers reflect the performance in the following way: the first number is the number of cases where the proposed algorithm performs better, the second where the algorithm specified at the top of the column performs better. The missing cases are those where both algorithm performed equally good. It can be seen that the proposed algorithm outperforms the normal pattern matching algorithm, which failed entirely to track the lower front teeth. As compared to Eigenfaces, the algorithm performed much better for the upper and lower teeth fillings, and comparably well for the front teeth. The results for the latter are affected by a considerable "measurement error" as the precise contour of the front are often impossible to see and the distances for the detected feature location is with the exception of 3 to 16 images, depending on the experiment, smaller than 4 pixel.

Further experiments were performed with 239 randomly generated images of faces under various illuminations [1] (implemented by H. Rowley [13]). Examples are shown in figure 1. However, the basic approach using one pattern of an eye, respectively the mouth, that is matched with the images does not perform well. The illumination variations include an important amount of shadowing as the virtual light source can be located at many different places. This violates the assumption that the illumination change is global for the object to be recognized.

---

[2]In a few cases the feature moved faster than 15 pixel per frame. The tracking was then restarted with approximately correct coordinates.

	pattern match-ing	Eigenfaces: number of used shape vectors			
		15	5	2	1
Upper teeth filling	100:0	100:0	100:0	100:0	100:0
Lower teeth filling	100:0	93:6	96:3	95:2	94:4
Lower front teeth*	—	34:34	33:44	37:31	45:43

Table 1: The performance of illumination corrected pattern matching with normal pattern matching and Eigenface models with four different numbers of used shape elements.

Figure 5: Example images from the artificial face database.

Therefore, the distorted gray-level approach was combined with the Eigenface model (compare equation (3)). In the experiments using these images, the combined method showed some improvement for the tasks "locate the mouth" and "locate the left eye" over distorted grey-level histograms with Eigenfaces and the original Eigenface models. Here, as the precise location of the mouth and the left eye is unchanged for all training test images, absolute errors can be calculated. In a first experiment, the training examples for the detection of an eye were selected to cover global and local illumination changes. This leads to an Eigenface model that performs better than the combined approach (for 5 shape vectors both performed best with 1.6, respectively 0.6, pixel mean error).

However, if the training samples are selected to have approximately the same average grey-level, it turns out that the novel approach performs better than the basic Eigenface model trained from the same data and from the training data covering global and local illumination changes (see figure 6A). Similarly for the mouth, the illumination corrected Eigenface model performed better than the basic Eigenface model, as shown in figure 6B.

Conclusion

We proposed a simple to use, but still efficient, method for the modelization of global illumination using distorted grey-level histograms. A quantitative comparison in experiments with standard pattern matching and Eigenface models shows that the proposed algorithm outperforms both.

In applications where only global illumination changes occur, pattern matching with distorted his-
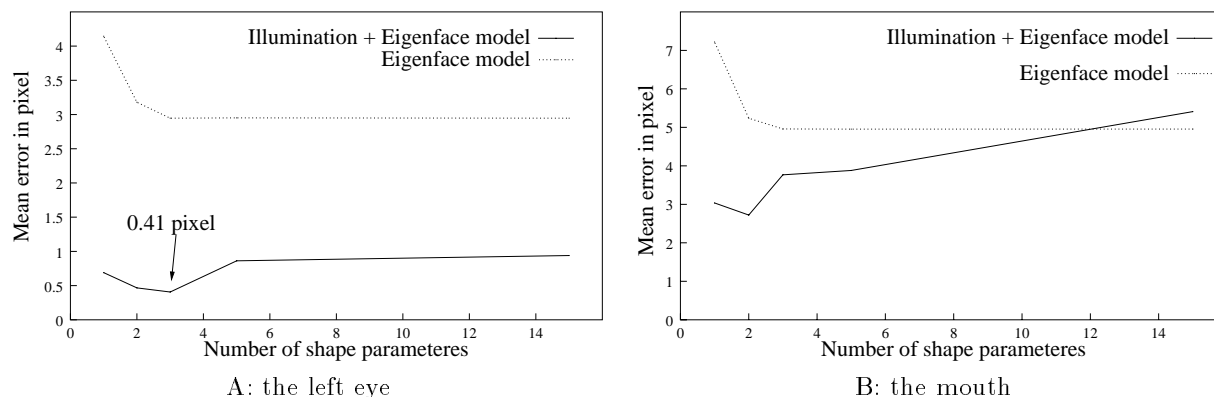
A: the left eye

B: the mouth

Figure 6: Mean error for the location of the left eye and the mouth committed by the basic Eigenface model and the illumination corrected Eigenface models if the training samples cover only local illumination changes.

tograms has a complexity close to standard pattern matching. This gives a further advantage over the Eigenface algorithm, which has a higher computational complexity and is somewhat more difficult to use and implement.

If local and global illumination changes are observable at the object, a combination of the illumination correction and the Eigenface approach outperforms the basic Eigenface modelization, even if they are trained on data including global and local illumination variations. Note, that it seems to be important that the training data model should not include global illumination changes, as this apparently degrades the performance of the combined approach.

References

[1] P. N. Belhumeur and D. J. Kriegman. What is the set of images of an object under all possible lighting conditions? *International Journal of Computer Vision*, 28(3):245–260, 1998.

[2] Martin Bichsel. Analyzing a scene's picture set under varying lighting. *Computer Vision and Image Understanding*, 71(3):271–280, September 1998.

[3] Michael J. Black, David J. Fleet, and Yaser Yacoob. A framework for modeling appearance change in image sequences. In *Proc. of the Sixth International Conference on Computer Vision (ICCV98)*. IEEE, January 1998.

[4] R. Brunelli. Estimation of pose and illuminant direction for face processing. *Image and Vision Computing*, 10(15):741–748, 1997.

[5] T.F. Cootes and C.J. Taylor. Modelling object appearance using the grey-level surface. In *Proceedings of the 5th British Machine Vision Conference*, pages 479–488, York, 1994.

[6] T.F. Cootes and C.J. Taylor. Using grey-level models to improve active shape model search. In *Proceedings - International Conference on Pattern Recognition*, volume 1, pages 63–67. IEEE, Piscataway, NJ, USA, 1994.

[7] A.S. Georghiades, D.J. Kriegman, and P.N. Belhumeur. Illumination cones for recognition under variable lighting: Faces. In *IEEE Conf. on Computer Vision and Pattern Recognition*, 1998.

[8] Berthold Klaus Paul Horn. *Robot Vision*. The MIT Press, 1996.

[9] A. Lanitis, C.J. Taylor, and T.F. Cootes. Recognising human faces using shape and grey-level information. In *Proceedings of the 3rd International Conference on Automation, Robotics and Computer Vision*, volume 2, pages 1153–1157, Singapore, 1994.

[10] N. Mukawa. Estimation of shape, reflection coefficients, and illuminant direction from image sequences. In *International Conference on Computer Vision (ICCV90)*, pages 507–512, 1990.

[11] K.G. Munhall, E. Vatikiotis-Bateson, and Y. Tokhura. X-ray film database for speech research. *Journal of the Acoustical Society of America*, 98(2):1222–1224, 1995.

[12] S. Negahdaripour and C.H. Yu. A generalized brightness change model for computing optical flow. In *International Conference on Computer Vision (ICCV93)*, pages 2–11, 1993.

[13] Henry Rowley. Home page. http://www.cs.cmu.edu/afs/cs.cmu.edu/user/har/Web/home.html, 1998.

[14] K.D. Skifstad and R.C. Jain. Illumination independent change detection for real world image sequences. *Computer Vision Graphics and Image Processing (CVGIP)*, 46(3):387–399, June 1989.

[15] M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1):71–96, 1991. anonymous ftp: `whitechapel.media.mit.edu/pub/images/`.